

**An Intelligent Automation Platform for Bioprocess  
Development**

**A thesis submitted to the University College London  
for the degree of  
DOCTOR OF PHILOSOPHY**

**Tianyi Wu**

**2015**

**The Advanced Centre for Biochemical Engineering  
Department of Biochemical Engineering  
University College London  
London UK**

## **Declaration**

I, Tianyi Wu confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Sign:

Date:

To my loving parents

## **Acknowledgements**

I would like to thank my supervisor Dr. Yuhong Zhou and my advisor Dr. Martina Micheletti for their support and guidance throughout my PhD studies. Thanks to all colleagues in the department of Biochemical Engineering for supports and helps.

The financial support provided by UCL Faculty of Engineering Science and China Scholarship Council are gratefully acknowledged.

To my parents, thank you very much for unwavering support and love. I would like to give a special thanks to my wife Liangliang Song for encouraging me to take on this endeavour.

## **Abstract**

Bioprocess development is very labour intensive, requiring many experiments to characterize each unit operation in a process sequence to achieve product safety and process efficiency. Recent advances in microscale biochemical engineering have led to automated high throughput experimentations. The activities for bioprocess development are implemented sequentially in which 1) liquid handling system performs the wet lab experiments; 2) standalone analytical devices detect the data; and 3) specific software is used for data analysis and experiment design. The experiment design, data analysis and process understanding require substantial engineer's time. It becomes one of the time consuming bottlenecks in bioprocess development particularly when a high number of experiments are needed to explore a large design space in order to discover high process performance.

This thesis addresses this challenge by reducing the development time yet achieving high process efficiency. A closed-loop learning approach has been adopted in bioprocess design in which the design objectives are driven iteratively by intelligent data acquisition. A framework that brings all of the elements performed manually into an automated fashion has been established to deliver the closed-loop learning approach. Based on the framework, a novel prototype of Intelligent Automation Platform for Bioprocess Development (IAPBD) has been built using multi-agent architecture. A rational database and four agents including Coordinate Agent, Experiment Design Agent, Execution Agent and Assay Agent have been designed to perform individual tasks and worked as a team to deliver the bioprocess design objectives. The multi-agent architecture used a blackboard mechanism to connect the four agents so that they are able to communicate with each other during operations. Starting with a set of initial experiments from the user or database, IAPBD is able to drive the robotic arm to perform defined initial experiments, and drive the analytical devices to detect the data after the experiments are completed. Then it will pick up the data from the analytical devices, and carry out data processing and process evaluation

based on the optimization objective to achieve the design solution. After evaluation of the design solution, it will decide to stop or continue to design the next round experiments for optimization.

The prototype has been evaluated first by a lysozyme precipitation process design, which involves typical microscale experimental procedures such as mixing, shaking, sample preparation and high throughput data detection instruments. All of the devices used have programmable interface so agents can control them directly. An optimal design solution that maximizes the yield of lysozyme and maximizes the ammonium sulphate concentration was found within a few of iterations using simplex search algorithm or Artificial Neural Network (ANN) and the whole tasks were completed automatically without human intervention. The success of this case study proved the concept of IAPBD.

The second case study was designed to further evaluate the prototype by the precipitation for monoclonal antibody purification. A new “watcher” algorithm in the Assay Agent has been further developed to communicate with instruments that do not have programmable interface e.g. HPLC. IAPBD carried out two sets of precipitation experiments automatically. The “watcher” algorithm has been proved robust and efficient to retrieve the data from HPLC.

The third case study is designed to use the sequential Design of Experiment method to optimize the production of a whole cell biocatalyst in fermentation. An optimal solution for medium composition and operating conditions was found successfully. It was then confirmed by large scale fermentation experiment that the optimal solution has increased the biomass/product more than 5 folds.

The benefit of this novel IAPBD is its automation that reduces the bioprocess design time significantly and frees engineer’s time for other intellectual tasks. Prove-of-concept of IAPBD has been achieved with limited real experimental evaluation. With further development and evaluation, its potential to significantly reduce the time and

cost of the whole bioprocess development may be realized.

## Table of Contents

<b>Abstract.....</b>	<b>5</b>
<b>List of Figures.....</b>	<b>12</b>
<b>List of Tables.....</b>	<b>18</b>
<b>Chapter 1 Introduction.....</b>	<b>19</b>
1.1 Perspective and motivation.....	19
1.2 High throughput laboratory automation devices .....	22
1.2.1 Microwell plate .....	22
1.2.2 Liquid handling system (LHS).....	23
1.2.3 Separation equipment for microwell plate .....	25
1.2.4 Signal detection.....	27
1.2.5 Integrated robotic arm and workstation .....	27
1.2.6 Application of high throughput techniques in bioprocess development .....	29
1.3 Data analysis methods for bioprocess development.....	37
1.3.1 Data storage and management .....	38
1.3.2 Experiment design and selection methods .....	39
1.3.3 Experiment data analysis and modelling methods .....	43
1.4 Quality assurance for laboratory automation systems .....	50
1.5 Multi-agent based architecture for system integration .....	53
1.5 Conclusions .....	55
<b>Chapter 2 Intelligent Automation Platform for Bioprocess Development (IAPBD)</b> <b>.....</b>	<b>58</b>
2.1 Introduction .....	58
2.2 Conceptual design of the intelligent automation platform for bioprocess development (IABPD).....	59
2.2.1 A closed-loop learning framework .....	59
2.2.2 Background and terminologies in IAPBD .....	62

2.3 Detailed IAPBD design .....	67
2.3.1 Multi-agent architecture for the intelligent automation platform .....	67
2.3.2 Database and data structure design .....	69
2.3.3 Agent design.....	75
2.4 Quality assurance for IAPBD .....	85
2.5 Conclusions .....	85
<b>Chapter 3 IAPBD verification by a lysozyme precipitation process design .....</b>	<b>87</b>
3.1 Introduction .....	87
3.2 Material & Methods.....	88
3.3 IAPBD implementation for protein precipitation experimentation.....	90
3.3.1 Multi-agent based architecture for precipitation process design.....	92
3.3.2 Experiment Design Agent: simplex method .....	95
3.3.3 Experiment Design Agent: ANN.....	99
3.4 The lysozyme precipitation process development.....	103
3.4.1 Precipitation optimization using the simplex algorithm .....	103
3.4.2 Precipitation optimization using the ANN algorithm.....	106
3.5 Conclusions .....	107
<b>Chapter 4 IAPBD implementation in a monoclonal antibody precipitation process design .....</b>	<b>109</b>
4.1 Introduction .....	109
4.2 Materials and Methods .....	114
4.2.1 Materials.....	114
4.2.2 Methods.....	115
4.3 IAPBD implementation for mAb precipitation experimentation .....	119
4.3.1 Assay Agent.....	119
4.3.2 Execution Agent .....	123
4.3.3 Experiment Design Agent .....	123
4.3.4 Coordinate Agent .....	124
4.4 The mAb precipitation process development .....	124

4.4.1 1 <sup>st</sup> experiment: mAb precipitation optimization on initial feedstock.....	125
4.3.2 2 <sup>nd</sup> experiment: mAb precipitation optimization on elute from AEX.....	129
4.3.3. Comparison study of a two-step mAb purification process.....	131
4.5 Conclusions .....	132
<b>Chapter 5 IAPBD implementation for fermentation process optimization.....</b>	<b>134</b>
5.1 Introduction .....	134
5.2 Materials and Methods .....	137
5.2.1 Reagent and cells.....	137
5.2.2 Shake flask MAO Fermentation.....	137
5.2.3 Microscale automated fermentation.....	137
5.3 MAO-N-D5 expressing <i>E coli</i> fermentation optimization .....	140
5.4 IAPBD implementation for fermentation optimization.....	144
5.4.1 Coordinate Agent .....	144
5.4.2 Assay Agent.....	144
5.4.3 Execution Agent.....	145
5.4.4 Experiment Design Agent .....	146
5.5 IAPBD demonstration through fermentation optimization .....	147
5.6 Comparison between tradition implementation and IAPBD.....	155
5.7 Effect of dissolved oxygen in fermentation.....	156
5.8 Pilot scale verification .....	157
5.9 Conclusions .....	158
5.9.1 Evaluation of IAPBD .....	158
5.9.2 Fermentation process optimization .....	159
<b>Chapter 6 Conclusions and future work.....</b>	<b>161</b>
6.1 Conclusions .....	161
6.2 Future work.....	166
6.2.1 Applications of microscale chromatography optimization on IAPBD .....	166
6.2.2 Optimized planning and scheduling functionalities during experiment execution .....	167

6.2.3 Integration with knowledge engineering techniques.....	168
6.2.4 Whole process design and optimization on IAPBD.....	169
<b>References.....</b>	<b>170</b>
<b>Appendix I Raw data for <i>E. coli</i> fermentatio .....</b>	<b>184</b>
<b>Appendix II Paper by the Author.....</b>	<b>188</b>

## List of Figures

Figure 1.1 An automated filtration setup. ....	26
Figure 1.2 A fully integrated workstation served by a robotic arm. ....	28
Figure 1.3 A customized 48 microwell bioreactor with impeller to increase dissolved oxygen level. (a) First prototype with free floating, magnetically driven impellers. (b) Second prototype with hollow shaft (Puskeiler et al. 2005). ....	31
Figure 1.4 A common procedures for sample preparation from compound collection for HTS. ....	33
Figure 1.5 Illustration of how the tips are operated with a robotic system.....	35
Figure 1.6 Schematic drawing of LHS compatible micro column (Wiendahl et al. 2008). ....	36
Figure 1.7 A classification algorithm illustrating active learning concept.....	41
Figure 1.8 Hypothesis-driven closed-loop learning of prototype robot scientist "Adam" (King et al. 2004).....	42
Figure 1.9 (1) Sequential approaches of experiments design for direct process optimization, (2) Sequential approaches for model-based design of experiments (Franceschini & Macchietto 2008) .....	44
Figure 1.10 An example of simplex search algorithm running on 2-dimension space .....	45
Figure 1.11 Structure of the applied optimization tool and data exchange between the software elements for optimization of chromatographic phase systems (Susanto et al. 2009) .....	47

Figure 1.12 an example of three-layer ANN with 5 input factors and one output.....	50
Figure 1.13 An intelligent agent with learning function. ....	54
Figure 2.1 Framework of intelligent automation platform to realize the closed-loop learning .....	61
Figure 2.2 A. A class diagram containing class name, attributes and operations (methods); B. An example of the class diagram representing a class of “bank account” .....	63
Figure 2.3 A ER diagram representing the database design for experiment storage in database.....	65
Figure 2.4 Multi-agent architecture for the intelligent automation platform.....	68
Figure 2.5 Entity-relationship model design for the database .....	70
Figure 2.6 Classes related to experiment. ....	72
Figure 2.7 Tables in IAPBD’s database that related to experiment .....	72
Figure 2.8 An example of the layout structure with a carrier carrying three disposable tip racks and an <i>Infinite 200</i> plate reader connected.....	74
Figure 2.9 Classes related to tree-structure layout.....	74
Figure 2.10 Tables and relations for layout in the database.....	75
Figure 2.11 Architecture of the Execution Agent .....	76
Figure 2.12 Class diagram for controlling Tecan LHS. ....	77
Figure 2.13 A simple example illustrating how translator works together with Execution Agent to pass information into LHS during operation.....	78

Figure 2.14 Interface design for experiment design algorithms. ....	79
Figure 2.15 Architecture of Assay Agent.....	80
Figure 2.16 An example of messages on blackboard showing an experiment published by the Experiment Design Agent. ....	82
Figure 2.17 Tables containing buffer information in the database.....	84
Figure 3.1 Calibration curve of the absorption at 280 nm and lysozyme concentration. .....	89
Figure 3.2 Hardware configuration of intelligent automation platform for precipitation. (1) disposable tips rack; (2) tube carrier; (3) 96-well microplates carrier; (4) vacuum filtration; (5) UV plate reader .....	90
Figure 3.3 Multi-agent architecture used in IAPBD for precipitation process design .....	91
Figure 3.4 Multi-agent architecture for lysozyme precipitation process development .....	94
Figure 3.5 The experiment design agent to facilitate the simplex search algorithm .	98
Figure 3.6 Schematic of a single hidden layer neural network.....	101
Figure 3.7 The Experiment Design Agent to facilitate the simplex search algorithm .....	102
Figure 3.8 The objective function increases with the number of iterations in the simplex search algorithm .....	104
Figure 3.9 The estimated surface of the objective function based on all 18 experimental points.....	105

Figure 3.10 The predicted lysozyme concentration response surface at various ammonium sulfate concentrations and pH values by the ANN method .....	106
Figure 3.11 The predicted objective function surface from the ANN method.....	107
Figure 4.1 Two downstream purification processes: (a) protein A chromatography platform; (b) precipitation based platform. (Ji 2012).....	111
Figure 4.2 Two different approach for process optimization that involves more than one unit operations.....	113
Figure 4.3 Calibration curve of total protein concentration using Bradford assay ..	117
Figure 4.4 Calibration curve of mAb concentration by peak area using HPLC .....	118
Figure 4.5 Calibration curve of impurity concentration by peak area using HPLC	119
Figure 4.6 Structure of assay agent in the case study .....	120
Figure 4.7 HPLC retention time distribution among 21 experiment runs .....	121
Figure 4.8 Illustration of simplex search algorithm on the objective function's surface. The designed experiment points are shown as the black dots. The number near the black dot is the iteration number. The final optimal point is shown in small red circle. ....	126
Figure 4.9 mAb and impurity concentration surface under different pH and ammonium sulphate concentration in 1 <sup>st</sup> run. ....	128
Figure 4.10 Objective function surface under different pH and ammonium sulphate concentration in 1 <sup>st</sup> run.....	128
Figure 4.11 mAb and impurity concentration surface under different pH and ammonium sulphate concentration in 2 <sup>nd</sup> run .....	130

Figure 4.12 objective function surfaces with different pH and ammonium sulphate concentration in 2 <sup>nd</sup> run.....	130
Figure 4.13 (A) mAb purities from the conditions selected by 1 <sup>st</sup> run and 2 <sup>nd</sup> run. (B) mAb yield from the conditions selected by 1 <sup>st</sup> run and 2 <sup>nd</sup> run .....	132
Figure 5.1 MAO-N-D5 mediated oxidation of secondary amine 3-azabicyclo octane (1) to the corresponding imine (2).....	136
Figure 5.2 Detail of the microwell automated system. (a) Photograph of the Tecan robot deck and ancillary equipment. (b) Deck layout as represented in the Tecan EvoWare software: 1) tip washing station, 2) 1000 and 200 µl sterile tip racks, 3) microplate racks for biomass sampling, 4) 20 µl sterile tip racks, 5) waste disposal unit, 6) microplate rack for colorimetric assay preparation, 7-9) microwell fermentation stations, 10) fermentation media containers, 11) microplate centrifuge, 12) microplate microfiltration and low pressure unit, 13) HPLC automated rack container, 14) microplate spectrophotometer.....	139
Figure 5.3 Schematic representation of the automated platform which performs microscale fermentation and biomass concentration measurement.....	140
Figure 5.4 (A) Two-level full factorial design of three factors in which all possible combinations among high and low levels of three factors are required to be examined. (B) Two-level fractional design in which some combinations are ignored to reduce the number of experiment runs and some three-way interactions terms are confound with two-way interactions. (C) Box-benhnken response surface design for three factorsin which the convexity of response surface is studied by introducing central points in experiment runs.....	143
Figure 5.5 Interactions among Assay Agent, Execution Agent and experiment devices. ....	145

Figure 5.6 Flowchart of Experiment Design Agent .....	146
Figure 5.7 Main factor plot of biomass concentration. A gray background represents a factor considered insignificant in the model. ....	150
Figure 5.8 Response surface plot of biomass concentration on four factors: glycerol (A), NH <sub>4</sub> Cl (B), shaking speed (C), fill volume (D).....	154
Figure 5.9 Time consumption in the traditional approach and IAPBD .....	156
Figure 5.10 Microscale fermentation following (•) DOT (%) and (▲) biomass (g <sub>DCW</sub> l <sup>-1</sup> ) using the selected fermentation conditions shown in where A stands for the concentration of glycerol, B for the concentration of NH <sub>4</sub> Cl, C for shaking speed, and D for fill volume. ....	157
Figure 5.11 The biomass growth in the 5L fermentation using (▲) selected media from Table 10 and (■) LB. (◆) is the DOT for the selected media .....	158

## List of Tables

Table 1.1 Comparisons of main liquid handling systems (Lye et al. 2003).....	24
Table 4.1 Conditions selected for two-step purification process .....	131
Table 5.1 The ranges of operating factors and medium chemicals .....	141
Table 5.2 Matlab toolboxes and functions used in Experiment Design Agent .....	147
Table 5.3 Fractional factorial design for factor screening. -1: low level, 1:high level. .....	148
Table 5.4 ANOVA of fraction factorial design.....	149
Table 5.5 Response surface design for selected four factors .....	151
Table 5.6 Factors included in the response surface design .....	152
Table 5.7 ANOVA for response surface design on four factors: glycerol (A), NH <sub>4</sub> Cl (B), shaking speed (C), fill volume (D). .....	153
Table 5.8 Optimal media and operating conditions for microscale fermentation ....	155

## **Chapter 1 Introduction**

### **1.1 Perspective and motivation**

Bioprocess development is to identify a robust operation space for a sequence of unit operations to produce, recover and purify a specific bioproduct and to achieve a desired yield and purity. It requires experiments to understand the impact and interaction of the process parameters of the whole bioprocess. Usually, a considerable number of experiments are needed for screening the feasible design space, which would consume lots of processing materials and time. The two resources are limited in bioprocess development, especially at the early stage. Hence, scientists have developed high throughput approaches of implementing experiments, which only requires small volume of processing materials to produce the results fast. Initially developed for drug discovery research where thousands of compounds are tested to identify drug candidates, the concept of high throughput approach has been adapted in the last decade to reduce the time and cost in bioprocess development. A series of unit operation ultra scale down (USD) technologies have been developed in which a microscale device is used to mimic the key features of a large scale operation (Tustian et al. 2007; Neal et al. 2003; Ma et al. 2009). It allows various combinations of parameters to be studied quickly at same time. In an example of investigating the best combination of buffer concentrations and pH value to capture polyclonal antibody by multi-mode resin (Chhatre & Titchener-Hooker 2009; Chhatre et al. 2010), 80 three-parameter combinations were studied in parallel. It demonstrates that the high throughput approach can implement experiments in parallel to generate process relevant experimental data quickly for bioprocess design and development.

Microscale systems have been most widely studied for bioprocess development based on microwell plate and automated liquid handling systems (LHS). An automated LHS is programmable to perform required experimental procedures. The robotic arm on

LHS is software-driven to perform addition and removal of liquids among different wells and to transport microwell plates between the attached automated devices. Such devices include shakers, vacuum separators, centrifuge, incubator etc. The automated system is then capable of doing most of the microscale experiments automatically for both up-stream such as *E. coli* growth kinetics and recombinant enzyme expression and downstream bioprocesses such as precipitation and chromatography with pre-packed micro column attached to pipetting tips on LHS (Micheletti & Lye 2006; Ferreira-Torres et al. 2005a; Knevelman et al. 2010; Titchener-Hooker et al. 2008). Several examples have been published to indicate that micro scale systems have shortened the bioprocess development time because of high throughput experimentation on microwell plate (Titchener-Hooker et al. 2008; Susanto et al. 2009).

The advances in automated system have also raised new challenges in bioprocess development. First, the development of high throughput assays does not match the rapid pace of experiment execution. Currently, the high throughput experimentation by liquid handling system is carried out and then the data is analysed to understand the process. Often 96 experiments are run in parallel on a 96-well plate, using statistic experiment design method (DoE) to screen a wide design space. Several replications of the experiments may be required to ensure the accuracy of experiments at micro scale. Therefore, hundreds of experiments are designed and executed simultaneously on such automated platforms. However, the development of high throughput assays does not match the rapid pace of experiment execution. For example, in monoclonal antibody (mAb) precipitation experiments, it took 30 minutes to run 96 experiments in a microwell plate but required more than 30 hours to detect mAb concentrations and total protein concentrations of the samples from 96 experiments. Although advanced HPLC systems with the ultra pressure or low volume have reduced the assay time, there is still a mismatch between time on experiment execution and time on assay in most bioprocess studies. Besides mAb and total protein concentrations, a large number of parameters for product and impurities such as nucleic acid and host cell

proteins need to be measured to characterize complex biological drugs. However, most biological assays need sample preparation and are time consuming.

The second challenge is data analysis and experiment design. The LHS is capable of carrying out 96 or 384 experiments in parallel by benefiting from microscale experimentation technology. After experimentation and assay, data analysis is the key step in bioprocess design. However, the software applications in automated instruments include data analysis modules for data treatment but are not designed to guide further experiment design or process understanding. For instance, the algorithm in ChemStation (Agilent Technologies, California, USA), which supports the Agilent HPLC instrument, can intelligently recognize the peaks and calculate the areas. More sophisticated data analysis software applications capable for design and analysis experiments such as Design-Expert from Stat-Ease Software (Minneapolis, MN), however, are stand-alone software and not integrated with automated systems. Therefore when the dataset is large, researchers often take substantial time for data analysis, process evaluation and process understanding. The new experiment design requires the user's inputs which come from the process understanding, e.g. which area should be explored, which variables should be studied, and what limits of variables are. Therefore, data analysis step becomes a potential bottleneck in bioprocess development. Hence, data analysis as well as experiment design need to be automated to further reduce the time and cost and improve the efficiency of the current practice.

The time needed for experiments and assays is dependent on the nature of the experiments and analytical instruments. Little space is available for time reduction. However different strategies for experiment design can lead to very different outcomes. In addition, automation of data handling and process evaluation can help us to eliminate the time wasted in data handling and software handling. Therefore, the current state of the art on high throughput technology, data analysis methods and automation technology will be reviewed.

## 1.2 High throughput laboratory automation devices

The laboratory automation has been developed for last few decades. Typically, a laboratory automation system consists of robots, conveyor systems, machine vision, and computer hardware and software (Markin 1992). For bioprocess development, the common devices adapted in automated high throughput systems are for liquid handling, separation equipment, and signal detection. A review of current state of the art on the high throughput laboratory automation is given below.

### 1.2.1 Microwell plate

In 1951, the first microplate originated as a micro-machined 8\*12 array in acrylic. Later the first molded 96-well plate was produced. Since then, microplates entered common usage. Later in 1996, the Society for Biomolecular Screening proposed standardization of micro-plate dimensions with a center-to-center well spacing of 9 mm in 96 wells (Mannhold et al. 2006). The standard microplates are  $86 \times 128 \times 14$  mm<sup>3</sup>. Typically, the plates contain 6, 12, 24, or 96-wells of variable geometry and volume, but 384, 1536 and now 3456-well plates are available for ultra high throughput screening (UTHS) applications. The volume of each well ranges from tens of nanolitres to several millilitres. Wells can be either circular or square. The standardization of microwell plates from late 1990s allows interoperability between plates, instrumentation and equipment from different suppliers This standardization is particularly important and provides the base for laboratory automation.

Microwell plates are being used for every major application in life science researches, which involves filtration, separation, optical detection, storage, reaction, mixing and cell culture. For example, UV transparent microwell plates enable automated plate reader to detect UV absorption of all wells in one run. Developed in early 1980s, microwell plates with an embedded layer of filter material make them possible to

perform microscale filtration in parallel on automated systems. The introduction of “tissue culture treated microplates” further helps studies of live cells that grow inside the microwells. The surface of these plates is modified using an oxygen plasma discharge to make their surface more hydrophilic so that it becomes easier for adherent cells to grow on the surface which would otherwise be strongly hydrophobic.

### 1.2.2 Liquid handling system (LHS)

Liquid handling system is widely used in high throughput technologies. Usually it is a robotic arm that dispenses a selected quantity of reagent, samples or other liquid to a designated container. The simple function is to dispense an allotted volume of liquid from a motorized pipette or syringe. However, complicated functions can also be implemented by the liquid handling system, e.g. plate handling.

Many commercial robotic arms are commercially available. Their specific properties are listed in Table 1.1. The number and dispense volumes represents the throughput which LHS can achieve. The function of liquid level sensing can increase the reliability of the systems. Other special features listed in Table 1.1 demonstrate the ability of the system to integrate with other devices, e.g. washer, shaker, vacuum pump.

Table 1.1 Comparisons of main liquid handling systems (Lye et al. 2003)

Manufacturer and main product	Pipettes	Dispense volume	Plate positions	Liquid level sensing	Plate handling	Special features
Beckman Coulter Biomeck FX workstation	8, 96, 384	1-250 $\mu$ L	12	Yes	Yes	Barcode reader, heating /cooling shaking. Two pipette heads operate at the same time. Gripper can assemble/ disassemble vacuum filtration manifold.
Cartesian Technologies SQ series workstations	1, 4, 8, 96	20 nL to 250 $\mu$ L	Up to 19	No	Optional	Very high speed, as dispenses 'on the fly' as the head passes over the wells. Optional plate loader can increase the capacity of the workstation
Gilson Cyberlab C-400 workstation	1, 8, 96	1-100 $\mu$ L	36	No	Yes	Gripper tool, vortex mixing of multiple plates, heater/chiller
Hamilton Company MICROLAB 4000 series	4, 6, 8, 12, 48, 96	1 $\mu$ L to 25mL	38	Yes	Optional	Gridding head for transfer from microplates to membranes, agar filters and inoculation of daughter plates. Solid phase extraction set up.
Robbins Scientific Tango series workstations	1, 8, 12, 96, 384	<1-200 $\mu$ L	12	No	No	Computer controlled plate positioning stage. Automated washing system. Disposable heads.
Tecan Genesis Workstation	4, 8	10 nL to 5 mL	20	Yes	Yes	Six slot shaker incubator, and integrated washing. Up to three heads working at the same time.
TOMTEC Quadra series workstations	4, 8, 96, 384	300 nl to 450 nl	8	No	Yes	Up to two heads operate at the same time. Shutting system means plate capacity is well above the deck capacity.
Zymark Scitone workstation	8, 96, 384	500 nl to 25 $\mu$ L	20	No	Optional	Disposable heads. Deck easily accessible from all four sides

The main concerns of the liquid handling systems currently are reliability, displacement techniques and integration with other automated devices (Felton 2003; Kong et al. 2012). Among these topics, the integration has attracted most attention. Integration is to link LHS with other relevant automated devices for biochemical experimentation, e.g. shaker, incubator and centrifuge. The advantage of the integration is that the automated system can perform more complicated experiment procedures other than liquid handling, so that human errors can be avoided. For laboratory automation, instead of purchasing a whole customized complex system, nowadays the trend is to acquire a number of different devices and components and integrate them together. The reason is that few manufacturers are capable of providing these integrated customized workstations, and these workstations are not cost effective in most cases (Kong et al. 2012).

Unfortunately, the components and devices for integration are often from different suppliers and no consideration is given to the integration, which makes integration a challenge. The automated devices, e.g. LHS, usually come with software that can control and monitor the devices. Very few of suppliers actually publish interfaces for third party program to interact with either their software or their device. It is difficult to work with a customized system that aims to integrate more than one device together. Some suppliers do provide options for the integration. For example, the software called EVOware from Tecan can simultaneously control several devices including LHS, vacuum separator and plate reader. However, such integration functionalities are largely limited within the devices from the same suppliers. For instance, EVOware can only control devices from Tecan. Thus, it is still a challenge to integrate the lab devices from different suppliers in most cases. Without the integration laboratory automation can not be realized.

### 1.2.3 Separation equipment for microwell plate

Centrifugation and filtration are common primary recovery procedures in bioprocess downstream process. Both separation operations are used in current high throughput

screening systems.

The centrifuges used in the high-throughput screening platform are similar to the common centrifuges where the processing materials are separated by gravity forces. These centrifuges integrated to the automated system are capable to operate with microwell plates, i.e. the centrifugation tasks can be programmed and the liquid handling system can deliver such tasks.

The filtration equipment consists a filter plate and a vacuum separation instruments. Due to the introduction of micro filter plate, different from the tradition harvester, the contents of microwell plate can be filtered individually through filter plate by a vacuum pump. The vacuum separation modules (Te-VacS) manufactured by Tecan is an example of such techniques. The cross-section of a single vacuum filtration manifold showing a 96-well filter plate and individual permeate collection tubes is given in Figure 1.1(Jackson et al. 2006). The modules are designed to automate cartridge-based and solid phase extraction processes. The devices can also be integrated with liquid handling system.

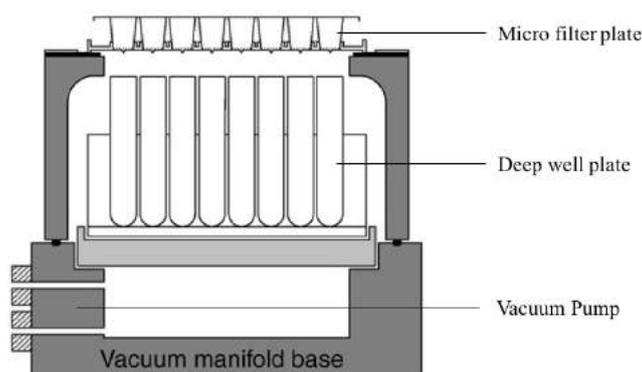


Figure 1.1 An automated filtration setup.

#### 1.2.4 Signal detection

Owing to the increasing number of types of assay format, the devices involved in signal detection are the most diverse part of automated high throughput systems. But for high throughput purpose, majority of assays rely on radiometric, fluorimetric, colorimetric or lumimetric detection systems. Several instrument manufacturers have developed very effective microwell plate based devices covering colorimetric, flourimetric and luminometric detection.

The optical plate reading devices are playing major role in signal detection. With UV transparent microwell plate, the plate reader is capable of measuring optical signals of 96 or 384 samples in one run. At the same time, a lot of assays based on optical signals have been developed for bioprocess development, e.g. bovine serum albumin (BSA) assay to measure total protein concentration, K-ACET assay from Megazyme (Wicklow, Ireland) to measure acetic acid concentration. High throughput and large number of applications make optical plate reading device an attractive assay solution for integrated automation systems.

#### 1.2.5 Integrated robotic arm and workstation

In 1961, the first industrial robot, called Unimate, developed and patented by Unimation Inc. was used for die-casting (Kumar 1998). It initiated a technological revolution in robotics. Subsequent improvements in computer control and gradual size reductions by automation industry and the U.S. space program development during the 1960s and 1970s gave rise to robotic devices that were attractive for clinical laboratory applications (Boyd 2002). Then in early 1980s, clinical laboratories actually began to use robotics and other automation technology to deal with thousands of samples a day. The applications were largely related to automated liquid handling robot for sample preparation.

After decades' development, the lab automation is becoming an integrated system, which is capable of managing sequential complex tasks rather than a one-job workstation. Thus,

more and more researchers begin to focus on the structure of the system. The concept of total laboratory automation is the use of automated devices and robots to perform all phases of clinical laboratory testing. A fully integrated solution called "Total Laboratory Automation" (also referred as fully integrated robotic system) was first introduced from US and later rapidly developed in Japan (Sasaki et al. 1998).

A fully integrated workstation is a collection of diverse peripheral instruments, which are served by an articulated robotic arm or conveyor to move plates between devices (Sasaki & Sugiura 2000). Such systems are controlled by scheduling software and capable of performing many different procedures such as assays. Earlier systems were room sized and were configured to perform a relatively limited range of high throughput screening assay methods. Figure 1.2 shows an example of such a fully integrated workstation served by a robotic arm. The independent automated devices like specimen manager, centrifuge and analyzer are linked by an articulating robotic arm. The robotic arm can transfer the microplate among different devices to perform different tasks.

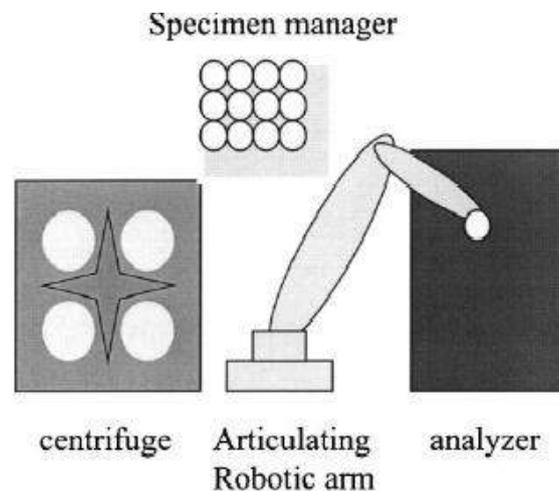


Figure 1.2 A fully integrated workstation served by a robotic arm.

Another type of workstation is a 'turnkey workstation' which is designed for performing an

entire automated process without human intervention (Hüser et al. 2006). The system includes pre-validated protocols for reagent/assay applications and the software for the specific applications. PAMPA Evolution invented by pION Inn is a good example of such a turnkey workstation (Galini-Luciani et al. 2007). It is a turnkey workstation for sample preparation and reading, data analysis, and presentation in a parallel artificial membrane permeability assay.

In comparison, the fully integrated workstation is less intelligent than the turnkey workstation. The 'turnkey workstation' has created the closed learning and data analysis capacities to some level. However, it lacks flexibilities since the turnkey workstation is designed for pre-validated applications. It is good for specialise applications.

#### 1.2.6 Application of high throughput techniques in bioprocess development

High throughput approaches have been applied to the study of many unit operations in a whole bioprocess sequence (Micheletti & Lye 2006). But most studies are still focusing on the upstream process development. A large number of researches focus on the high throughput screening (HTS) which is an essential part of drug discovery and development (Gibbon & Andreas 2005; Mishra et al. 2008). Many studies have contributed to microbial fermentation (Kostov et al. 2001; Rios-Solis et al. 2013). However, by taking the advantage of recently developed liquid handling systems, several examples are published to demonstrate that high throughput technology may shorten the development time of downstream process such as sample preparation (Saitoh & Yoshimori 2008), precipitation (Knevelman et al. 2010; Titchener-Hooker et al. 2008) and chromatography (Chhatre et al. 2010; Wiendahl et al. 2008).

### 1.2.6.1 Fermentation

The automated experimentation on fermentation studies has centred on either quantification of *E. coli* growth kinetics (Rios-Solis et al. 2013; Ferreira-Torres et al. 2005b) or medium optimization (Micheletti & Lye 2006).

The experimental setup usually involves LHS, deep-well microplates, shaker, and incubators. After media preparation step is completed by LHS, fermentations can be run in parallel in incubators. The shaking speed can control the oxygen to some extent.

However, in standard deep-well microplate, dissolved oxygen limits the growth of cells, which is not quite same as the larger scale environment. In order to overcome the limitation, several customized microwell plates have been introduced to better mimic the performance in larger bioreactor (Szita et al. 2005; Puskeiler et al. 2005).

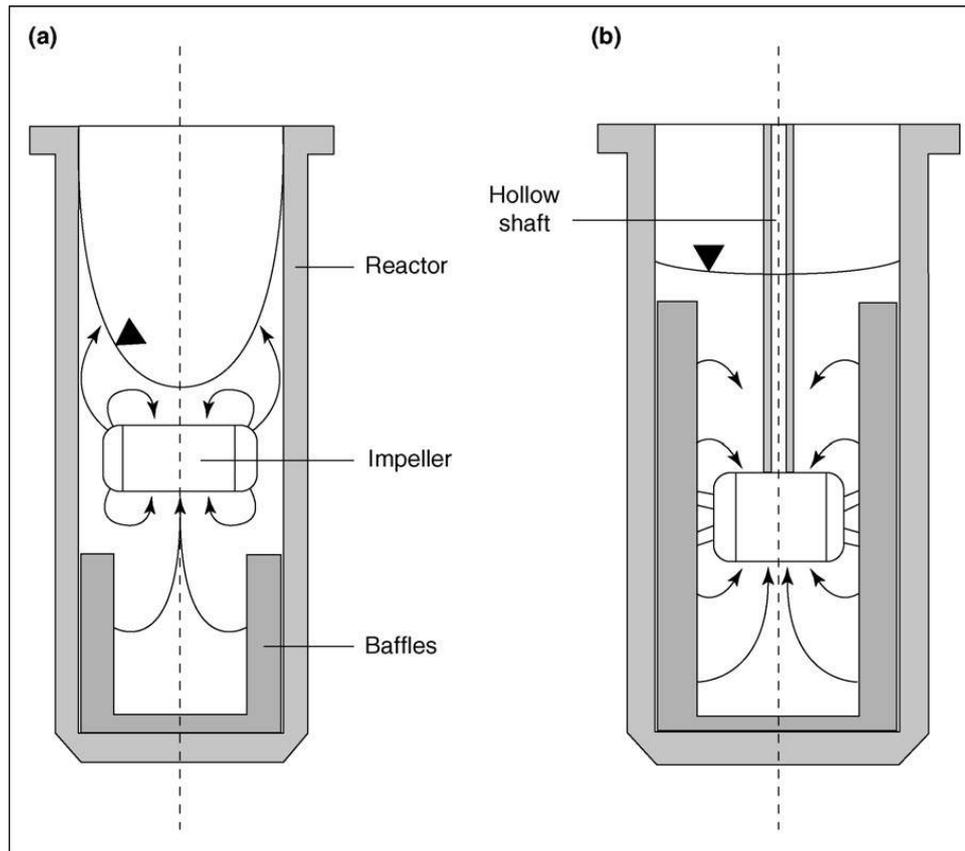


Figure 1.3 A customized 48 microwell bioreactor with impeller to increase dissolved oxygen level. (a) First prototype with free floating, magnetically driven impellers. (b) Second prototype with hollow shaft (Puskeiler et al. 2005).

The studies have shown that the micro scale fermentation performed on microwell plate can mimic the large scale process, and the results gained from micro scale experimentation provide useful guidance on large scale production optimization. Researchers are starting to overcome the limitation of dissolved oxygen level in microwell plate in order to further explore the potential of micro scale fermentation. An example is a customized 48 microwell bioreactor with impeller to increase dissolved oxygen level shown in Figure 1.3.

The uses of microwell plates with standard size make them very convenient to perform on automated system with LHS. However, the increasing volume of data generated by those

high throughput experimentation techniques makes data analysis a potential bottleneck to slow down the micro scale fermentation process development. Thus, an integrated and automated data analysis module is needed to match the increasing throughput.

#### 1.2.6.2 Sample preparation

The automated sample preparation process has been demonstrated in several prototypes of total laboratory automation (Sasaki & Sugiura 2000). Generally, the concept of sample preparation refers to the ways in which a sample is treated prior to its analysis, and it usually includes the following steps:

- Weighing
- Dissolving
- Dilution
- Sample distribution to microwell plates
- Sample Replication

Many large pharmaceutical companies possess compound stores, which comprise 100,000 to 500,000 structurally distinct samples (Devlin 1997). These compounds stored as dry material in individual bottles need to be formulated in a proper manner. Figure 1.4 shows a typical preparation methodology for HTS. Firstly, the compounds are retrieved from compound store. Then, those compounds are solubilized in ‘mother’ microwell plates. Before transport into screening devices, the mother plate sets are replicated into several ‘daughter’ plate sets. The preparation of multiple daughter copies permits compounds to be submitted to multiple screens and allows conservation of compound collection.

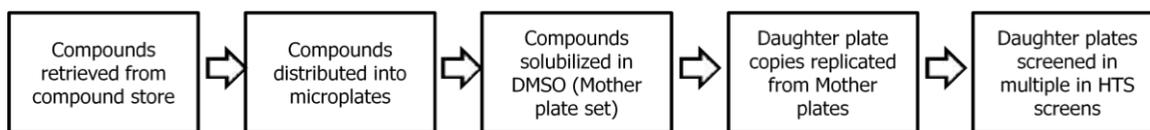


Figure 1.4 A common procedure for sample preparation from compound collection for HTS.

In 2008, a prototype of fully automated laboratory robotic system proposed by Saitoh & Yoshimori consists of liquid handling system, parameter control of temperature and volume etc., and analytical HPLC (Saitoh & Yoshimori 2008). Unlike the other automated systems, the software modular of sample management and scheduling is added into the system. In this way, the time of whole experimental process is significantly reduced by running different tasks in parallel.

#### 1.2.6.3 Precipitation

Precipitation has been developed as a method for protein purification and recovery. Precipitation has the advantages of being relatively fast and low cost. It also has simplicity of operation and proven track record in large-scale blood product fractionation (Curling 1980).

The experiment setup for USD precipitation usually involves LHS, microwell plate, shaker, incubator, microwell plate reader for UV absorption, and HPLC. After sample preparation step is completed by mixing feedstock solution with buffer solution at various ion strengths and pH values, the concentrations of product and impurities are then measured by either plate reader or HPLC.

Several challenges were risen up when using precipitation for industrial scale antibody purification. Currently, the development methods of precipitation are highly empirical (Niktari et al. 1990; Habib et al. 2000). Several parameters such as precipitant type, temperature, pH, and ionic strength can significantly impact precipitation outcomes. Therefore, large amount of experiments need to be executed in order to establish an empirical mathematical model (Knevelman et al. 2010). High throughput automated system is well suited for precipitation studies. The combination of microwell plates and liquid handling system was proved to largely improve the efficiency of research process. For example, the optimal condition of antibody precipitation using PEG on Tecan liquid handling system was reported (Knevelman et al. 2010).

#### 1.2.6.4 Chromatography

Chromatographic processes are widely used in industrial purification process in the biotechnological sector. The performance of a chromatographic separation is naturally influenced by a various parameters e.g. salt concentration, pH and matrix etc. (Titchener-Hooker et al. 2008). The screening studies of high throughput chromatography mainly lie on ligand screening, resin screening, and column screening.

In order to test a large amount of experimental settings by microwell plate in parallel, several approaches have been developed: (1) the microlitter batch incubation approach; (2) micropipette chromatography tips; and (3) miniature packed columns (Chhatre & Titchener-Hooker 2009). Common procedures in these approaches are summarized below:

- (1) applying samples to micro device;
- (2) determining the parameters through micro scale work (resin type, binding capacity, flow rate range, etc.);
- (3) executing further experiments to test the optimal conditions;

(4) evaluating the result generated from laboratory studies by pilot scale studies.

The microlitter batch incubation approach is to use microwell filter plates where a well-defined volume of matrix is loaded into each well in the form of plaques or slurries of resin (Bergander et al. 2008). Buffer conditions for binding and elution between proteins and resins can be added followed by incubation, mixing, sample preparation and data analysis. This microwell filter plates have been proven to be successful in investigation of many buffers, pH values, concentrations and matrix types simultaneously.

The second approach uses packed chromatography pipette tips to form micro columns. The detailed design of the tips is shown in Figure 1.5. The resin shown in grey at the base of tip is constrained between two screens, whilst the adaptor on a robotic pipetting arm is connected to the tip and thus enables the pump to draw liquid through the end of the tip before dispensing it back into the well (Chhatre & Titchener-Hooker 2009). The tips enable the liquid handling system to investigate the whole chromatography processes. Several studies used micropipette chromatography tips to model the performance of column (Coffman et al. 2008; Chhatre et al. 2009; Chhatre et al. 2010).

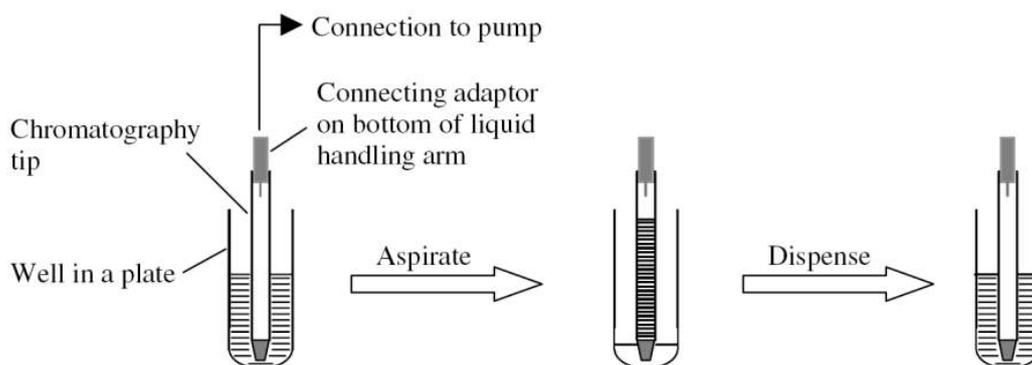


Figure 1.5 Illustration of how the tips are operated with a robotic system.

The third approach is to apply miniature columns. Compared with the tips or filter plates, the miniature column is more geometrically similar to the large scale columns. The miniature

columns can be used by either hand or robot (Wiendahl et al. 2008). As shown in Figure 1.6, a micro column is capable of directly attached to pipetting tip of robotic LHS. The flow rate can be concisely controlled by the LHS. It is also reported it is very efficient when studying breakthrough curve and other performance parameters of chromatography.

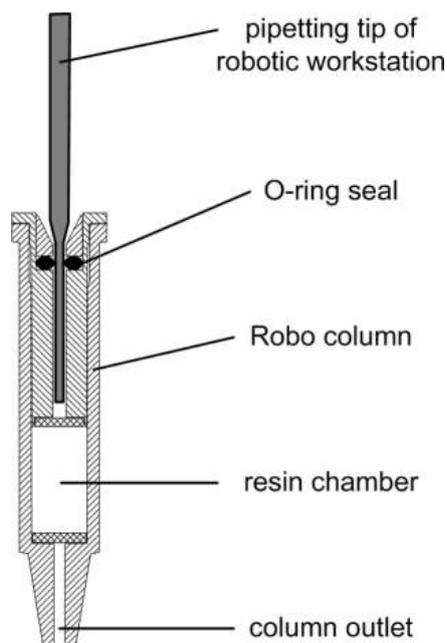


Figure 1.6 Schematic drawing of LHS compatible micro column (Wiendahl et al. 2008).

In conclusion, the automated laboratory equipment has been developed for decades. The applications of automated approaches in bioprocess development have been reported for many unit operations in bioprocessing. The obvious advantage of these automated approaches is the high throughput experimentation capability. However issues in the automated systems still exist. The prediction model generated from microscale experiments may not be able to fully mimic the outcome of large scale situation (Micheletti & Lye 2006). For example, in micro scale fermentation studies, the growth curve behaves differently comparing with large-scale performance due to oxygen limitation.

Furthermore, new challenges come with such an approach. Since the liquid handling system is capable of carrying out 96 or 384 experiments in parallel by benefiting from microscale experimentation technology, the burden on sample data analysis has increased hugely which requires long time and large quantity of reagents for biological assays.

### **1.3 Data analysis methods for bioprocess development**

The applications of automated systems for experiment implementation have demonstrated that these devices are able to perform the experimental procedures and assays automatically for bioprocess development. Then data analysis is the key step in bioprocess design after experimentation and assay, because a large amount of experimental data can be generated from the automated experimentation. Data analysis includes: 1) to convert raw data from the measurement instruments into biological parameters; 2) to understand the process performance based on the data and, more important, 3) to design the new experiments for process development. The methods involved in these tasks will be reviewed next.

The automated and high throughput approaches are not able to release full potential without intelligent experiment design and data analysis methods. In fact, the data storage and processing functionalities are playing an even more important role and considered as advances on laboratory automation (King et al. 2004). There are three types of data analysis tasks:

1. To convert raw data from the measurement instruments into biological parameters. For example, in protein quantification using UV spectrophotometry the spectrophotometry data need to be translated into protein concentrations using calibration models. The software applications in most instruments include data analysis modules for data treatment. For instance, the algorithms in ChemStation (Agilent Technologies, California, USA), which supports the Agile HPLC instrument, can intelligently recognize the peaks and calculate the areas.

2. To store laboratory and experiment related data into the information management system. A system that stores and manages experiment related data is required for management of users, samples, and automated devices etc. For instance, the database stores all historical experiment design and outcome data in order to accumulate data and knowledge to guide future experiments on bioprocess development.

3. To understand the process performance based on the data and, more importantly, to design the new experiments for process development. Here we use experiment design to refer to determining the experimental conditions, not the experimental procedure. For instance, there are sophisticated data analysis software applications capable of the design and analysis of experiments such as Design-Expert from Stat-Ease Software (Minneapolis, MN).

As mentioned before, first type of tasks has been solved by software from automated device suppliers. Therefore, this sector will focus on the other two types of tasks.

### 1.3.1 Data storage and management

At present, most software for laboratory data storage and management comes under the heading of 'laboratory information management system' (LIMS). A LIMS is a software system used in laboratories for integrating all laboratory software, instruments, and the management of samples, laboratory users, standards and other laboratory functions such as sample planning, plate management, and workflow automation. In short, LIMS provide a solution to track, analyse, and store the data and metadata produced through the whole lab experimentation. Some LIMSs even provide direct integration of devices and the use of web-based systems in order to minimize installation and maintenance costs (Thurow et al. 2004). However, the LIMS is usually designed for data storage and operations other than for direct interrelation with physical devices.

In a laboratory, stand-alone devices are often used where a whole process study is separated

into individual steps such as preparation, reaction, assay, etc. This resulted in nonhomogeneous system landscapes with more or less autonomous automation cells. LIMS tries to tackle the problem by consolidate all the data into one system. However, in most cases, it is relatively difficult to automatically retrieve the data from different laboratory devices. Usually, manufacturers of experiment devices provide specific software managing one or more involved devices. Since no standardized interface has been defined for different hardware and software so far, it will apparently increase the installation cost when trying to integrate devices from different manufacturers to implement LIMS. Besides, since no sophisticated data analysis functionality coming with LIMS, the bottleneck still exists in processing experimental data especially when integrating with high throughput approaches.

### 1.3.2 Experiment design and selection methods

Beside data storage, the very first task of data analysis is to design experiments, i.e. to select different experiment conditions for characterizing or modelling a process. The simplest method is the brute-force method if neither material cost nor time consuming is considered as constrains. The method here is simply to test every possible experimental condition. While the brute-force method is simple to implement and will always find a solution if it exists, its cost has positive linear relation with the number of possible experiment conditions, which in many practical cases tends to grow very quickly when the number of input parameters increases. However, at early stage of bioprocess development, biological material is not readily available for large number of experimental investigation. Even with micro scale approaches, information rich experiments design is still necessary not merely to reduce experimental cost but to generate more accurate and reliable prediction model for bioprocess development. The common experiment selection methods are random selection, factorial selection and methods in active learning.

Random design or random selection means to randomly generate a given number of experiments within the constraints of experimental inputs. For example, in some search

algorithms, experiment results generated from a given number of randomly selected initial experiments can often provide some directional information for future investigation. However, due to the nature of search algorithms, random selection of initial experiments may result in a local optimal rather than a global optimal point. Besides, it is relatively hard to gain any information about the relation among inputs and outputs parameters, i.e. response surface profile.

The factorial selection is suitable for the experiment whose design consists of two or more factors, each with discrete possible values or “levels”. The factors, however, can be either continuous or discrete variables. The factorial selection is to select the experiments on all possible combinations of these levels across all such factors. Such selected experiments allow studying the effect of each factor on the response variables, as well as the effect of interactions between factors on the response variables. The method is commonly used in DoE (Design of Experiments).

Another type of experiment design is ‘active learning’. Unlike traditional machine learning methods, active learner is allowed to choose the data from which it learns so that it performs better with less training (Settles 2010). The incentive behind active learning is to reduce the number of experiments because, sometime, it is difficult, time-consuming, or expensive to obtain labelled instances. For instance, in bioprocess development, the experiment results (label) for a set of condition variables (instances) is not easy to obtain due to high material cost and long processing time. Thus, it may be a promising experiment selecting tool for rapid bioprocess design, which is to locate the optimal condition settings by requiring a small number of experiments.

One of the active learning methods is the membership query synthesis. The algorithm selects the experiments based on a given model. The model is then refined according to the results of these experiments (Angluin 1988). Different from other sampling methods, this approach generates instances (experiment design) by itself rather than using real world unlabelled instances (historical experimental results). The heuristic of active learning is as follows (Figure 1.7):

1. Start with a pool of unlabeled data (Figure 1.7(1));
2. Pick a few points at random and get their labels (Figure 1.7(2));
3. Repeat the following:
  - a. Fit a classifier to the labels seen so far (Figure 1.7(3), Figure 1.7(5))
  - b. Pick the best unlabeled point to get a label for closest to the boundary (Figure 1.7(4), Figure 1.7(6))

In this case, the accurate classifier has been trained without labelling all points in data set.

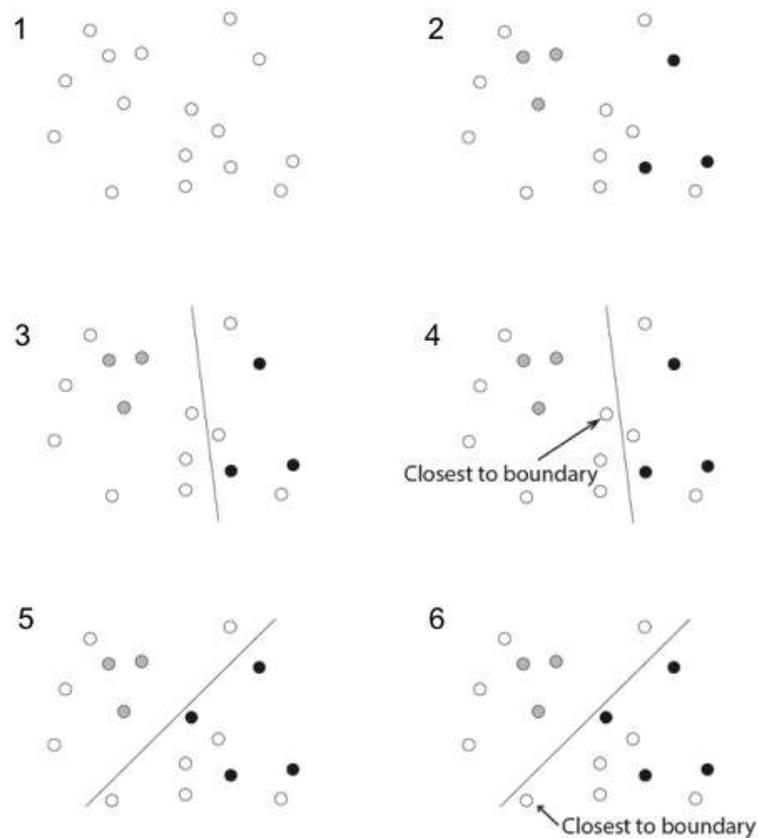


Figure 1.7 A classification algorithm illustrating active learning concept

The first prototype of robot scientist (Whelan et al. 2004) actually implemented the concept of membership query synthesis. The system can execute a series of autonomous biological



cost-sensitive active learning in settings where annotation costs are variable and not known. The cost here is a function of time it may take to perform a query. Another successful example of experimental design to minimize the cost is the robot scientist Adam (King et al. 2004). It will estimate the probability of each hypothesis being correct and to calculate the time and cost it may take to verify each hypothesis. A mathematical model is then established to deal with the trade off between the cost and the time to generate the optimal design. The method results in a three-fold decrease in the cost of experiment materials compared to naively running the least expensive experiment, and a 100-fold decrease in cost compared to randomly generated experiments.

### 1.3.3 Experiment data analysis and modelling methods

Experimental data analysis and modeling methods have been established to investigate the appropriate operating conditions for complex bioprocess understanding. Most methods in the literature are either model-based or direct process optimization (Susanto et al. 2009). These methods usually have sequential or cyclical nature.

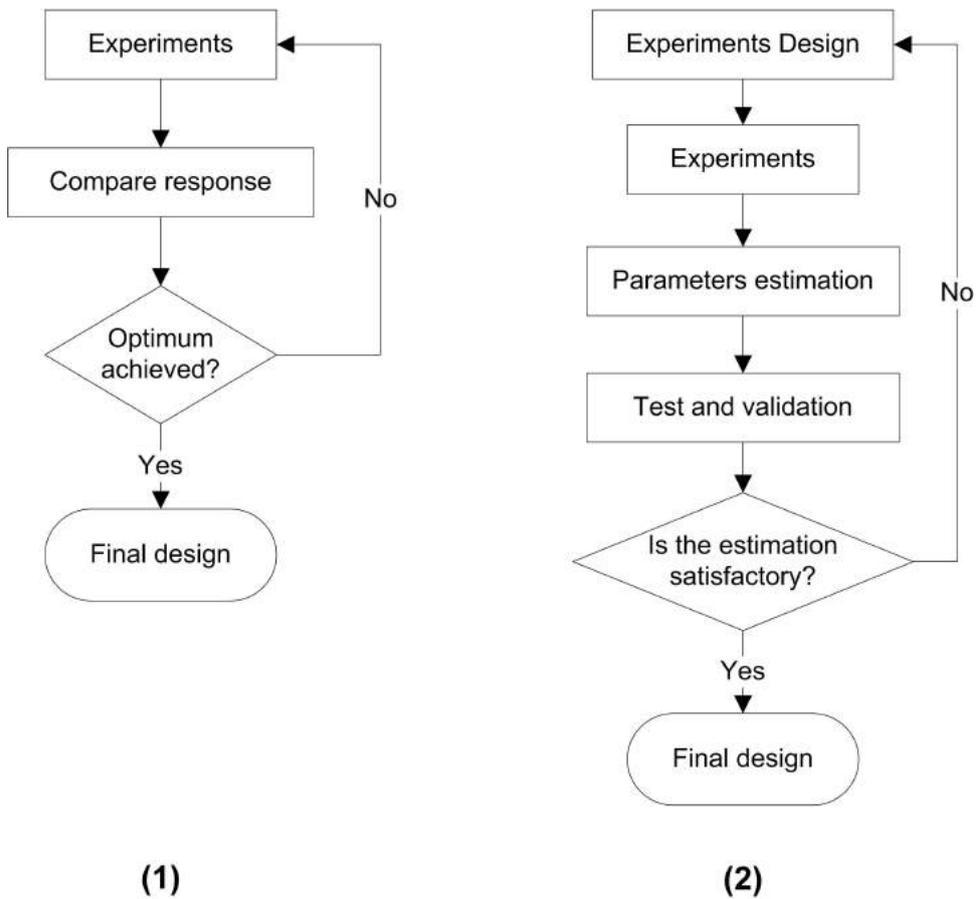


Figure 1.9 (1) Sequential approaches of experiments design for direct process optimization, (2) Sequential approaches for model-based design of experiments (Franceschini & Macchietto 2008)

### 1.3.3.1 Direct process optimization

Direct process optimization is a group of methods that uses search algorithms to find an optimal operating condition with specified properties. The search algorithms fall into three classes according to the collections to be searched: virtual search spaces, sub-structures search of a given structure, and quantum computers (Knuth 1973). Virtual spaces search is to find a variable assignment that will maximize or minimize a cost function of those variables as well as satisfying specific mathematical equality and inequality constraints. The sub structures search of a given structure is to find a specific sub-structure of a given discrete

structure, e.g. list, tree, graph, etc. There are also search algorithms that are designed for quantum computers. Currently, only the virtual spaces search algorithms are adapted in the optimization problems of bioprocess development. When it is not necessary to investigate the profile of whole experimental space, search algorithms are often preferable to other learning techniques.

Simplex search algorithm is a simple optimization algorithm seeking the vector of parameters corresponding to the global extreme (maximum or minimum) of any  $n$ -dimensional function searching through the parameter space (Olsson & Nelson 1975). The idea of simplex algorithm is to iteratively discover the  $n$ -dimensional response surface by changing the initial simplex, which is generated from  $n+1$  points on the surface. Figure 1.10 gives a simple example of simplex search algorithm on 2-dimension space. Start with initial 'simplex' ( $\Delta BWG$ ), the algorithm will iteratively construct new simplex using the points from Reflection (R), Contraction (C), Expansion (E), or Shrinkage (S, M). The algorithm will be explained in Chapter 3.

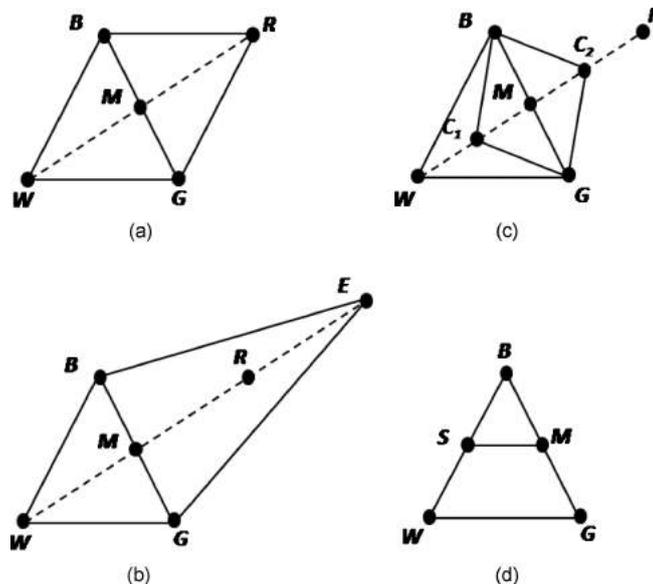


Figure 1.10 An example of simplex search algorithm running on 2-dimension space

The early application that implemented simplex algorithm into bioprocess development is to use the algorithm to discover the optimal chromatography conditions to separate different components inside the HPLC system (Ghodbane & Guiochon 1988). This study actually combined both factorial design and simplex search optimization. Simplex search algorithm was firstly used to locate the optimal condition in small number of iterations. Then a full factorial design was conducted to understand the response surface around the optimal condition. However, there are some certain disadvantages of simplex algorithm (Horstkotte et al. 2010). Firstly, the performance of the algorithm largely depends on the selection of initial points. The algorithm might fail to find the global optimal solution when more than one extreme point are presented in the response surface, which means that the iteration might result in local optimal rather than global one.

Another type of algorithms that are widely used is metaheuristic methods. The widely known algorithms within the class are simulated annealing, tabu search and EA (evolutionary algorithm). While normal heuristics are problem-dependent techniques, metaheuristics are problem independent, which means they do not take advantage of any specificity of the problem and, therefore, can be used as black boxes (Duran et al. 2009). These techniques usually have very good performance dealing with NP-hard (Non-deterministic Polynomial-time hard) problems, However, similar with traditional heuristic methods, metaheuristics do not guarantee that a globally optimal solution can be found on some class of problems (Olafsson 2006). And these methods usually implement some form of stochastic optimization, so that random variables generated can be crucial for final results for some cases (Olafsson 2006).

The applications of these algorithms in the field of bioprocess development majorly focus on process scheduling and planning because these methods are powerful when dealing with discrete optimization problem. For example, the algorithms has been used to search optimal solutions for various biomanufacturing process economics model (Allmendinger et al. 2014). In addition, some applications on chromatography optimization are reported. Among all bioprocess unit operations, chromatography requires large initial fixed capital investment

and is lack of flexibility. Many studies used the algorithms to find optimal sizing strategies for the sequence of chromatography steps employed to purify biopharmaceuticals (Allmendinger et al. 2014).

An application of EA in chromatographic phase systems was to optimize the buffer composition in the capture step for a binary protein mixture of lysozym and cytochrome and the multi-linear gradient elution for the separation of a ternary mixture of ribonuclease, cytochrome, and lysozyme (Susanto et al. 2009). The research proposed a closed-loop optimization strategy that integrates data storage in Excel, data analysis by EA in Matlab, and automated devices of Tecan Evoware as shown in Figure 1.11. The research provides an approach to enhance efficiency of laboratory automation devices by advanced data analysis algorithm. However, since the structure of the whole system is designed especially for chromatographic phase optimization, the architecture is not capable for general use. It is also lack of integration with large data storage and other devices apart from Tecan Evoware.

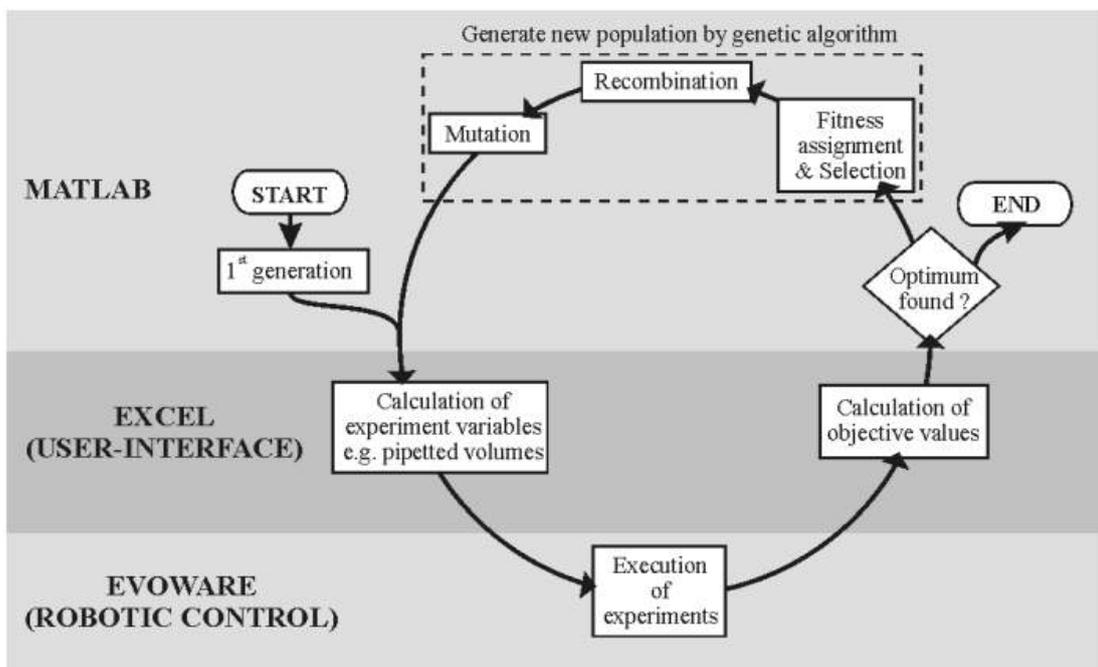


Figure 1.11 Structure of the applied optimization tool and data exchange between the software elements for optimization of chromatographic phase systems (Susanto et al. 2009)

### 1.3.3.2 Modelling methods

Comparing with direct process optimization, the model-based methods are aiming to fetch more information about mathematical relation among all input and output variables. The methods can start with either a well-defined model or a polynomial equation, to estimate the best set of parameters that fits the experiment data.

The most traditional statistical methods fall into the category of model-based methods. The aim of statistical methods is to investigate causality and to draw a conclusion on the effect of changes in the values of predictors on dependent variables. These methods are efficient especially when the relations between responses and dependent variables are known, such as regression methods on coefficient estimate. DoE method falls into this category.

The technique of DoE is often used to plan experiments. The idea of DoE is to use statistics model to determine the significant factors as well as to estimate corresponding coefficients. ANOVA (analysis of variance) is widely used in DoE. ANOVA is a particular form of statistical hypothesis test of whether or not the means of several groups are equal, and therefore generalizes the t-test to more than two groups. At same time, ANOVA can provide a linear model with coefficient estimates and standard errors.

The fundamental technique is a partitioning of the total sum of squares ( $SS$ ) into components related to the effects used in the model. For example, the model for a simplified ANOVA with one type of treatment at different levels (one-way ANOVA).

$$SS_{Total} = SS_{Error} + SS_{Treatments}$$

And F-test statistic can be calculated for comparing the factors of the total deviation.

$$F = \frac{MS_{Treatments}}{MS_{Error}} = \frac{SS_{Treatment}/(I - 1)}{SS_{Error}/(n_T - I)}$$

where  $MS$  is mean square,  $I$  is number of treatments,  $n_T$  is total number of cases. And the null hypothesis for F-test is:

$$H_0: \mu_0 = \mu_1 = \dots = \mu_I$$

where  $\mu_i$  is mean of  $i$ th treatment.

If  $H_0$  is rejected from F-test, the treatment is proved to be significant factor under chosen confidence level.

However, the experimental design is done to ensure that a particular series of experiments will be statistically capable of answering the given question, rather than to minimize the actual cost of the experiments. However, the biochemical material could be quite expensive, which results in the experimental cost being a relatively significant factor that must be taken into consideration during experimental design. Besides, the polynomial regression model generated by DoE may result in unsatisfactory fitting on complex non-linear system.

Artificial neural network (ANN) is becoming popular in many research fields including bioprocess development. Instead of giving an exact mathematical expression, the alternative approach is to establish an artificial neural network inspired by the biological neural network. The central idea is to extract linear combinations of the inputs as derived features, then model the target as a nonlinear function of these features. The network consists of several linked layers, which are input layer, output layer and several hidden layers (Lek et al. 1996). Each layer has a certain number of neurons (node). Each node in hidden layer has an activation function, which can be either sigmoid function or tanh function. It takes output values from linked nodes in previous layer as inputs, and generates output value based on activation function. The activation function of output node can be either sigmoid function or linear function depending on whether it is a classification or regression problem. In this case, the model generated from ANN is no longer linear combination of input factors. The complex multi-layer structure enables ANN to learn rather complicated non-linear relation between input factors and output value.

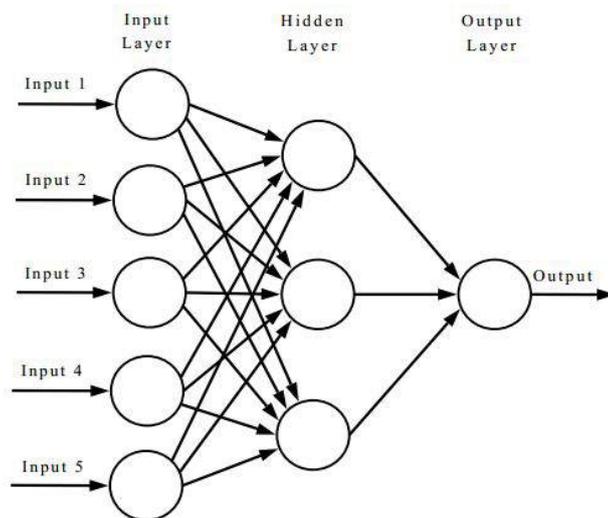


Figure 1.12 an example of three-layer ANN with 5 input factors and one output.

There are a few of successful ANN applications for modelling the biochemical systems, especially the solubility studies (Eslamimanesh et al. 2011; Huuskonen et al. 1997; Zhang et al. 2004). The most attractiveness of this method is that it can capture the patterns of data precisely without knowing the mathematical relations previously. However, the method may be too sensitive to conclude even the errors into the final model. So the parameters including starting values, number of layers and learning rate should be carefully chosen before execution of the method.

#### **1.4 Quality assurance for laboratory automation systems**

The fault detection and quality assurance are a very important aspect in fully automated solutions, especially for laboratory automation. Because majority of USD experiments are currently carried out in microwell plate on LHS, fault detection and quality assurance on LHS have drawn most attention in laboratory automation (Felton 2003).

A few techniques have been developed to build a reliable fault detection system on LHS.

These are some common technologies used in LHS for fault detection.

- Confirmation of tip mounting: A feedback sensor on each liquid handling channel to monitor the presence or absence of a disposable tip mounted on the cone.
- Plate grip sensing: A feedback sensor system on robotic arm to confirm that microwell plate has been gripped.
- Liquid level detection: A sensor system to detect liquid level in micro well based on liquid conductivity and/or pressure .
- Safe pathways: A route planning algorithm for liquid handler to make sure tips never travel over a sample or other piece of labware in order to reduce the risk of cross contamination.

Although fault detection is a built in function for most modern LHS, quality assurance is still needed to reduce inaccuracy and imprecision of experiment results. Broadly speaking, the quality of biochemical experiment results depends on two factors: 1) the assay variability, and 2) the product concentration (Hentz & Knaide 2014). In turn, the sample concentration in the assay depends on many factors, such as initial concentration of the stock solution, sample storage, handling throughout the process, solubility and stability of the sample in medium buffer, and pipetting accuracy and dilution scheme (Popa-Burke et al. 2009). Therefore, assessment of LHS performance has become important in order to minimize experimental error occurred in automated high throughput experimentation (Taylor et al. 2002). One of most important indicator in evaluating the performance of LHS is accuracy, which is commonly reported as %bias (Rose 1999).

$$\%bias = 100 \times \left( \frac{V_m - V_t}{V_t} \right)$$

where  $V_m$  is the measured volume and  $V_t$  is the desired volume. %bias represents the deviation from the desired volume, and 0% means no deviation.

Reproducibility is measured by precision, which is calculated from the mean and standard deviation (SD) of a set of measurements. It is usually reported as percent coefficient of

variation (%CV).

$$\%CV = 100 \times \frac{SD}{mean}$$

There are several ways for volume verification. Because a modern LHS is usually equipped with capacity of liquid level detection, liquid volume can be measured by calculation based on microwell plate dimensions and liquid level in microwell. However, gravimetric measurements have been used extensively to calibrate the liquid dispensers in LHS (Bergsdorf et al. 2006; Rhode et al. 2004). It utilizes the mass and the density of solution to determine the volume. Usually, the liquid is dispensed into a pre-weighed microwell plate, which is weighed immediately after dispensing to prevent evaporation (Taylor et al. 2002). Some studies applied photometric methods, such as absorbance and fluorescence, for volume measurement as well (Chai et al. 2013). A study has shown that the performances of absorbance and fluorescence have no significant difference on determining the precision of liquid handlers (Petersen & Nguyen 2005).

It is commonly accepted that a LHS is required to have a %bias value below 5% and a %CV below 10% (Chai et al. 2013).

The outlier detection is another important factor to be considered in a fully automation solution. Because outlier detection in bioprocess experiments normally only involves one-dimensional data (e.g. protein concentrations), the simple way to identify potential outliers is to mark any points that are more than two standard deviation from mean value. However, the presence of outliers is likely to have a strong effect on the mean and the standard deviation, which makes this approach unreliable. MAD (median absolute deviation) is considered a better method for outlier detection (Popa-Burke et al. 2009). MAD is a measure of statistical dispersion, and it is more resilient to outliers in a data set than the standard deviation. To calculate MAD of  $n$  observations,  $\{x_1, x_2, \dots, x_n\}$ :

1. Calculate  $D_i = |x_i - med|$ , where  $med$  is median of  $\{x_1, x_2, \dots, x_n\}$ .
2.  $MAD = median(D_1, D_2, \dots, D_n)$

Normally, observation farther than 4.5MAD from the median are marked to be outlier (Popa-Burke et al. 2009). Since the standard deviation of a normal distribution is 1.5 times the MAD (Pereira & Williams 2007), the cutoff is three standard deviations from the mean, where only 0.3% population will lie here for a normal distribution. The observations this far away can be reliably called outliers.

### **1.5 Multi-agent based architecture for system integration**

Due to the complexity of the dynamic automation system and the demanding of a closed-loop learning performance, a sophisticated architecture is required. Agent-based techniques are considered to be suitable to accommodate the integration of the current independent components (Sycara 1998). Within the agent-based architecture, an agent is defined as an autonomous unit capable of executing actions in this environment to meet specific objectives (Wooldridge & Jennings 1995). An intelligent agent (IA) is an autonomous entity, which observes and acts upon an environment and directs its activity towards achieving goals. Intelligent agents may also learn or use knowledge to achieve their goals. They are able to cooperate to accomplish a complex task in a distributed way by the capability of interaction and communication. The diagram shown in Figure 1.13 illustrates the basic principle of an intelligent agent. There is a closed learning loop inside each agent. The learning loop consists of two key elements: learning element and performance element. Agents can obtain information from the outside environment. The learning loop gains new knowledge according to the experiments executed by performance element. According to the knowledge base inside the agent, it will take actions according to the environmental information.

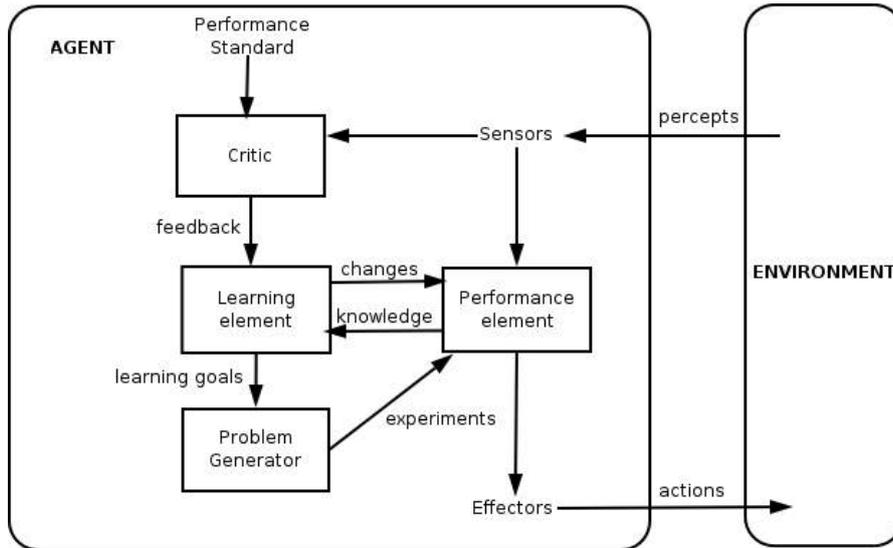


Figure 1.13 An intelligent agent with learning function.

The agent-based architecture has shown its advantages in several diverse engineering areas including problem solving, simulation, collective robot etc. (Ferber & Drogoul 1992). Compared with traditional integration approaches, an agent-based architecture has a number of advantages satisfying the requirements of the intelligent automation systems. Using agent-based techniques, the modular approach can be adopted for biopharmaceutical process design because it is easier and less costly for system software modification when adding or deleting components (Gao et al. 2010; Genesereth & Ketchpel 1994). An agent-based architecture can divide complex tasks into small and manageable sub-problems (Sycara 1998). The agents can collaborate to solve the presented problem by communication and interaction with each other so the simultaneous facilitation of information exchange is feasible. Based on these features, the agent-based architecture provides an ideal approach to integrate all the components involved in the required closed-loop learning. For example, the design algorithm, experiment execution and assay may each be regarded as an autonomous agent to perform defined tasks. The flexibility of the architecture makes it easy to add new algorithms and devices into the platform without significant modification of architecture software.

Although there are few publications on applications of multi-agent system for bioprocess design and optimization (Gao et al. 2010), the architecture has been successfully implemented in the mechanical manufacturing optimization (Chen & Chen 2010; Maturana & Norrie 1996). Different from traditional manufacture system for mass production, the flexible manufacture is becoming preferable to survive the vicissitude of the market. So the production process and schedule need to be re-designed frequently. Due to the large scale of the production process, multi-agent architecture is regarded as the best solution: each agent represents a specific machine or procedure and the best design can thus be achieved through the auction mechanism of agents. Traditionally in bioprocess development, each unit operation is studied separately, which makes it difficult to capture interactions among unit operations. However, multi-agent based approach can overcome this problem to potentially provide a global optimal solution (Gao et al. 2010). A framework of multi-agent system may be developed to handle the optimization of bioprocess, in which each agent consists of knowledge base and learning module for a specific bioprocess operation, and agents cooperate to find the optimal design according to the design requirement (purity, yield etc.).

## **1.5 Conclusions**

The wide implementation of microscale devices and equipment enables a new way of doing experiments with extremely small volumes. In this way, the cost of materials would be significantly decreased and at the same time, large number of experiments can be executed in parallel.

The automated instruments have been introduced into the bioprocess development for many years. The liquid handling systems are liberating researchers from simple and repeatable liquid transfer tasks. In addition, the errors that generated by researchers during performing these tasks can be avoided. The total automated laboratory system enables even more experimental tasks to be done automatically.

The software applications, which came along with the automated instruments, have shown

their power in many bioprocess development cases. The software developed by Tecan such as EVOware and FACTS are capable of controlling several different devices within a single software environment. These two software applications make it possible to add scheduling features while executing experiments so as to better utilize all the instruments. The laboratory information management system can store and analysis the data generated from different instruments.

All these advanced technologies are currently used in development of bioprocess. However, researchers tend to use them separately so as an experiment usually involves several instruments and software. Therefore, the following disadvantages would appear.

- Simple and repeatable manual work is still required such as liquid handling, simple calculation of pH, volume, instrument configuration etc.
- Experimental data sometimes is difficult for researchers to organize and analysis because those can be generated from different experiments.
- Instruments may have conflicts on scheduling, e.g. one instrument can be extremely busy at one time period while the others may be idle when there are more than one experiments being executed.
- Laboratory logistics can be unnecessarily increased due to frequent transfer of experimental materials among different instruments, which are usually located in different locations.

Although there is no such a prototype that can perform entire closed-loop learning tasks on bioprocess development, each part of the learning loop such as automated instruments, laboratory information management system, and data analysis and modelling methods has been well developed for several decades. Therefore, a software that integrates both laboratory automation device and modern data analysis techniques may be feasible and can potentially speed up bioprocess development significantly.

Thus, an efficient and cost-effective system for automated bioprocess development is

extremely important and desirable. The system should bring all the elements operated manually into an automated platform. This includes the physical devices that carry out experiment preparation, experiment execution and assays. In other words, the system should consist of five types of software:

- 1) Software to drive each device timely;
- 2) Software to consolidate the data for each device due to their different format;
- 3) Software of sequential design algorithms that realize the process design objectives and capable of process evaluation, prediction and optimization;
- 4) A database used for storing the experimental data and information of instrument configurations;
- 5) Software to control and manage above software.

My research project aims to develop and evaluate such an intelligent automation platform for bioprocess development (IAPBD). Firstly, a feasible architecture is required for integration of both physical devices and data analysis. The architecture should be capable of controlling liquid handling system, plate shaker, plate reader and other assay devices. Besides, the effective communication mechanism will be presented to integrate automated hardware devices with data storage and analysis modules. Secondly, a group of agents will be designed to deliver tasks in experiment design, experiment execution, assay, data analysis and etc. Thirdly, a series of representative bioprocess development case studies will be performed using the platform. The case studies will cover both upstream and downstream bioprocess unit operations. The purposes of the case studies is to examine whether a intelligent bioprocess design platform including experiment design, execution and sequential analysis algorithms can be automated; to evaluate the capability of the platform in solving practical bioprocess development problems.

## **Chapter 2 Intelligent Automation Platform for Bioprocess Development (IAPBD)**

### **2.1 Introduction**

The applications of microscale bioprocessing studies have covered all major unit operations ranging from upstream and downstream. The wide implementation of these microscale bioprocessing techniques offers the potential to speed up bioprocess design, reduce research and development cost and shorten delivery time of new drugs to the market. The application of robotic automated devices enables even faster bioprocess development by significantly reducing the experimentation time and eliminating errors from manual operations. These automated approaches are reported to be capable of speeding up the acquisition of process information (Micheletti & Lye 2006; Jackson et al. 2006).

However, it is also reported that biopharmaceutical development is becoming more time consuming and expensive due to the difficulties in bioprocess design and optimization. The drug molecules are getting more complex and requiring more experiments to characterize their properties. With the automated platform, a much larger design space could be explored to discover much better solutions, which leads to the increase of the experiments (Karlsson et al. 2004). The automated platform can deal with the substantial increase in the experiments well via parallel or high throughput experimentation at micro scale. However the data analysis burden can be handed by some of the high throughput analytical instruments but a large portion of biological analysis is not high throughput compatible. On the other hand, the volume of the data generated has also substantially increased. To treat these raw data and make sense out of the data efficiently require large amount of time. The volume, complexity and diversity of the data sets have made data analysis the new bottleneck for the process development. A few techniques have been implemented in bioprocess design and optimization such as statistical modelling, search algorithms and

metaheuristic methods. However these methods and their software are isolated from automated experimentation devices so the data processing and analysis are labour intensive hence slow down the process design.

In this chapter, an intelligent automation platform for bioprocess development (IAPBD) has been proposed to address the data processing and analysis problems in micro scale experimentation. IAPBD aims to automatically perform the experiments and data analysis without the involvement of manual operations. This novel approach of laboratory automation integrates the currently independent parts, the experiment design, experiment execution and assay analysis together to allow bioprocess solutions to be investigated rapidly within the automated platform.

In this chapter, I have created software that realizes IAPBD's functions. The concept of IAPBD design is presented in section 2.2.1. The detailed design for each element in IAPBD is given in Section 2.3 after the introduction of relevant background and terminologies in Section 2.2.2.

## **2.2 Conceptual design of the intelligent automation platform for bioprocess development (IABPD)**

### **2.2.1 A closed-loop learning framework**

When exploring a large design space, the sheer volume of experiments needed is substantial. In a fermentation process, more than 20 variables in medium design and process operations need to be characterized. In chromatography process, to characterize several resins compound with their buffer conditions for load, elution and equilibrium require a huge amount of experiments. If whole bioprocess design approach is considered, the interactions between steps also need to be investigated, then the demand on experimentation will be very high.

As reviewed in Chapter 1 the search methods are preferred as such approaches start with a

small set of experiments to get knowledge of the biological material properties and key process performance parameters. Based on the feedback from the gained knowledge new experiments are then designed iteratively to achieve their process development goal. Such approaches, if used manually, are practical, manageable but often time consuming.

The concept of the closed-loop learning has been adopted in machine learning research for many years (Settles 2010). The incentive behind the concept is that the current optimal solution can be iteratively improved by taking latest feedback into consideration. The implementation of the closed-loop learning in bioprocess development involves the control of experimentation devices and operations, data processing and decision-making on new experiment design. Conceptually, starting with a set of initial experiments from the user or database, the intelligent automation platform for bioprocess development based on the closed-loop learning approach intends to drive the experimentation devices to perform defined initial experiments and drive the analytical devices to detect the data after the experiments are completed. Then it will pick up the data from the analytical devices and perform data analysis that turns the raw data into variable values such as protein concentrations, and carry out process evaluation based on the design objectives. After evaluation of the design solutions, it will decide to stop or continue to design the next round experiments for further optimization. As shown in Figure 2.1, the framework of IAPBD consists of five different components: 1) data & knowledge base; 2) sequential design method; 3) execution of experiments; 4) assay execution; 5) data treatment.

*Data & Knowledge base* is a database designed to store and accumulate the data of previous experiment results. It will have the capability of providing supporting data for experiment design such as operating area, input variables and output variables.

*Sequential Design Method* is a software element that has the experimental design algorithms established, capable to design a batch of experiments compatible with micro scale devices, then translate the experiments into detailed simple tasks, finally write the command scripts for these tasks that can be recognized by experiment devices. It also performs evaluation of returned experiment data based on the pre-defined research objectives, makes decisions on

whether it is necessary to plan next batch of experiments to get further results. If the expected objectives have been achieved, i.e. new knowledge has been successfully discovered, the new knowledge will be added into knowledge base. If not, further validation experiments will be send to Execution of Experiments & Assay element to execute.

*Execution of Experiments & Assays* are hardware associated with the experimentation to deliver the pre-defined tasks. IAPBD should be designed to drive this hardware at right time through communication.

*Data Treatment* is another software element that processes raw data, translates raw data into the format that has biochemical meanings. For example, it translates uv absorption data into protein concentration.

A key element in the closed-loop learning is the objectives of the activity, which needs to be defined beforehand. To deliver a process design activity, the objectives must be defined and formulated mathematically by the users. Typical design objectives include purity, yield, cost, productivity or combination of them. This will help the biochemical engineers focus on their goals.

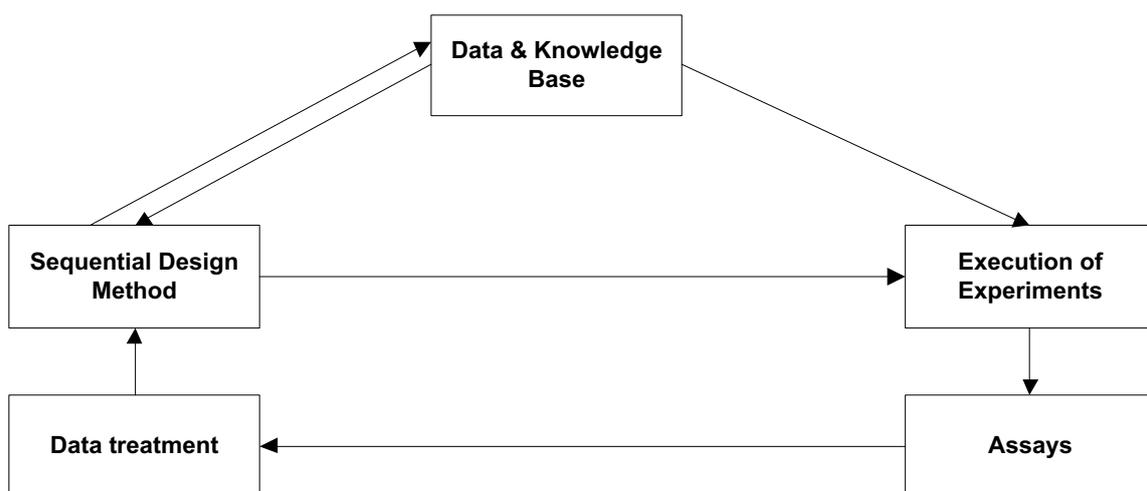


Figure 2.1 Framework of intelligent automation platform to realize the closed-loop learning

## 2.2.2 Background and terminologies in IAPBD

Utilized artificial intelligence (AI) and objects-oriented programming (OOP) IAPBD has been designed to drive the framework to deliver automated bioprocess design tasks. The development of the intelligent automation platform involves a few of techniques in software engineering. Before presenting the detailed design, some of the background and terminologies will be introduced.

### 2.2.2.1 Object oriented programming

Object oriented programming (OOP) is a programming paradigm based on the concept of “objects”, which are data structures that contain data field (attributes) and methods (functions). An important feature of OOP is that an object’s methods can access and often modify the attributes of the object with which they are associated.

The ideas behind OOP are back from a programming language called SIMULA (Nygaard & Dahl 1978). The language was designed for doing simulations. Moreover, it firstly introduced objects, classes, inheritance and subclasses, and discrete event simulation. Later the concept of objected-oriented has been applied in systems programming, graphics, and AI programming (Stefik & Bobrow 1985). Comparing with other program paradigms such as Procedural Programming and Functional Programming, OOP has certain advantages in code reusability, reliability and flexibility, and maintenance (Nierstrasz 1989). Therefore, OOP is chosen for development of IAPBD.

To increase code reusability and extensibility, the notion of “class” is introduced. “Class” defines the data format (attributes) and available procedures (methods) for a given type of objects (Stefik & Bobrow 1985). “Object” inherits the definition from “class”, whose structure contains both attributes and methods. In other words, each object is said to be an instance of a particular class. For example, an object with its name attributes set to be “computer” might be an instance of class “Product”.

Another important concept in OOP is “interface” (Stefik & Bobrow 1985). An interface is a common protocol for unrelated objects to communicate with each other. It is a set of definitions of methods and attributes, which the classes agree upon in order to cooperate.

The design of a class is illustrated in a box containing three parts as shown in Figure 2.2A. The top part contains the name of the class. The middle part contains the attributes of the class. The bottom part contains the methods, which the class can execute. An example of a class diagram is shown in Figure 2.2B. A class of “bank account” has two attributes: owner and balance. The class has two methods: deposit and withdraw.

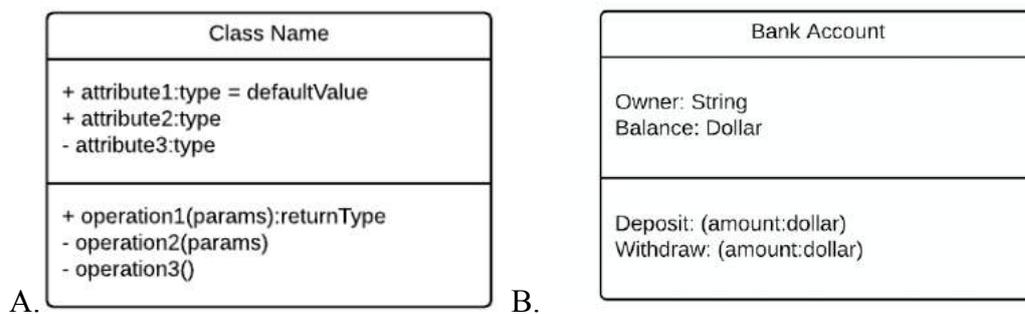


Figure 2.2 A. A class diagram containing class name, attributes and operations (methods);  
B. An example of the class diagram representing a class of “bank account”

OOP has several great advantages. 1) It increases code reusability. Objects created for OOP can easily be reused in other programs. 2) An object has ability to hide certain parts of themselves from other objects, and defines how it can be interacted. This feature prevents various kinds of errors. For example, with OOP, the pH value in an experiment cannot be set beyond 14. 3) OOP is easier to design and program comparing with other programming styles. The incentive behind OOP is experience in real world, which makes it easy to write and maintain

### 2.2.2.2 Entity relation model (ERM)

An ERM is a model that uses a systematic process to describe and define a subject area of data. ERM was first introduced in late 1970s for unification of different views of data: the network model, the relational model, and the entity set model (Chen 1976). In ERM, the data are represented as entities that are linked with each other by relationships. The relationships express the dependencies and requirements between entities. Similar to OOP, the entities can have various attributes. ERM now is widely implemented in database design due to several advantages: 1) straightforward relation representation; 2) easy conversion to other data model; 3) graphical representation for better understanding (Chen 1976). The database that follows the concept of ERM is called a rational database. In this case, each table in the database is one instance of an entity, while each row in the table is the attributes.

The diagram that represents ERM is called entity relationship diagram. There are a few different ways to draw ER diagram. A commonly used diagram is very similar to the class diagram as mentioned earlier in this chapter. Same as the class diagram, the entities in ER diagram are represented by boxes with three different parts. The top part is the name of an entity. The middle part is the primary attributes (primary keys), which work as indexes in data tables. Thus, the value of primary key (PK) must be unique in each row. The lower part is other normal attributes. The linkages among different entities represent the relationships. There are three types of those relationships: 1) one to one; 2) one to many; 3) many to many. The attributes who have one-to-one relationship with other table's PK are called foreign keys (FK). Figure 2.3 shows an ER diagram representing the database design for experiment data storage in database. For an experiment, the experiment ID, the input data and output data need to be stored. Therefore the diagram has three entities: experiment, experiment design and experiment result. Because each experiment has a unique ID, three entities are linked by two one-to-one relationships.

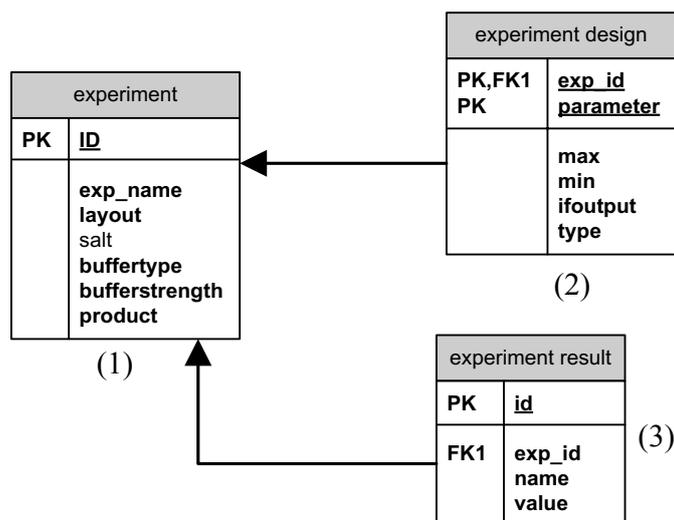


Figure 2.3 A ER diagram representing the database design for experiment storage in database.

There are two major advantages using the rational database based on ERM in IAPBD. Firstly, it reduces redundancy in database. Because relationships among different data entries are well defined in ERM, same sets of information is not necessarily stored in multiple data tables. Secondly, it is simple and easy to understand. The structure is described in business user's language, and even non-technical user can understand easily.

### 2.2.2.3 Application Programming Interface (API)

An API is a set of commands as well as the format of those commands which can be sent by one program to another so that programs can communicate and use each other's functions directly. An API defines functionalities regardless of their respective implementations, in which case the definitions and implementations are not necessary to be same.

In laboratory automation, an API is provided by an automated device's supplier, so that the third party program can use its commands to communicate with either control software or

the device directly. For example, EVOware is the software designed to control Tecan robotic LHS, and its API provides commands that allow a third party program to communicate with EVOware. Therefore the LHS can be controlled by a third party program.

#### 2.2.2.4 Extensible Markup Language (XML)

XML is a markup language that defines a set of rules for encoding documents in a format, which is readable for both human and machine. XML was developed by an XML Working Group formed under the auspices of the World Wide Web Consortium (W3C) in 1996 (Bray et al. 1997). The XML is designed to transport and store data. According to design goals of XML, it has some advantages: 1) it is straightforwardly usable; 2) it supports a wide variety of applications; 3) it is easy to write programs which process XML documents; and 4) it is human-legible and clear.

The basic building block of an XML document is an element, defined by tags. An element has a beginning and an ending tag. The XML document forms a tree structure that starts at the “root” element and branches to the “leaves”. This structure enables XML to support hierarchical data structures or “nested” elements. An example of using XML to describe a precipitation experiment is given below:

```
<?xml version="1.0" standalone="yes"?>
<experiment>
  <input>
    <salt_concentration unit="mg/mL">10</salt_concentration>
    <pH>7</pH>
  </input>
  <output>
    <yield>80</yield>
    <purity>90</purity>
  </output>
  <objective_function>0.86</objective_function>
```

</experiment>

In this example, the experiment has two input variables, which are salt concentration and pH value, two output variables, which are yield and purity, and the value of an objective function. All of the information has been represented in the XML format as shown in the example.

### **2.3 Detailed IAPBD design**

In order to deliver the closed-loop learning in bioprocess design activities, I have designed the data structure and a rational database, four agents, and a multi-agent architecture to form the IAPBD. In the following sections, multi-agent architecture, database design as well as data structures will be described first. Then each of four agents will be presented respectively.

#### **2.3.1 Multi-agent architecture for the intelligent automation platform**

The multi-agent architecture has been proposed for my intelligent automation platform for bioprocess development which will be referred as the platform. The platform has four individual yet connected agents: Coordinate Agent, Execution Agent, Experiment Design Agent, and Assay Agent as shown in Figure 2.4.

In order to control all of the software agents to accomplish a defined task, an efficient communication protocol is needed to link all the agents together. An interaction protocol that is able to support several activities such as the communication of global goals, sharing of knowledge, and avoidance of conflicts is essential in multi agent architecture design (Davidsson & Wernstedt 2002). The most widely applied interaction protocols are contract

net (Davis & Smith 1983), market mechanism (Wellman 1995), and blackboard (Nii 1986). The first two are designed to simulate real market activities on the assumption that a global optimum exists for the equilibrium of the market such that the “seller” wants to maximize profit while the “buyer” wants to maximize utility. The agents in these protocols therefore have similar functionalities such as “buy” and “sell.” When agents are capable of different functions, the architecture based on the “blackboard” mechanism is attractive. The blackboard is a database, which all relevant agents can access to. The agents perform tasks based on the information on the blackboard. The agents publish the results on the blackboard at real time and continuously until the goal has been achieved.

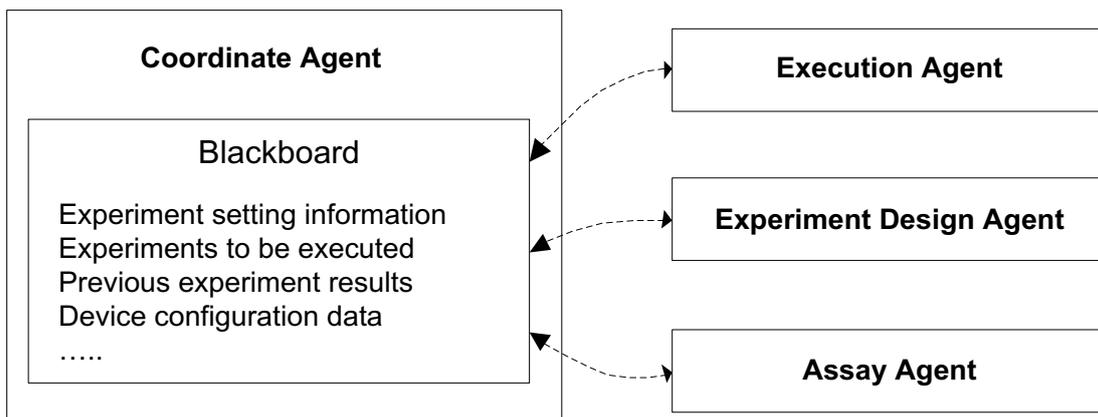


Figure 2.4 Multi-agent architecture for the intelligent automation platform.

*The Coordinate Agent* retrieves the historical experiment data from an ER database and writes the new executed experiment results into the database. It also translates the information in the database in the message format to blackboard. The blackboard is the place to publish two types of information. One is the configuration data such as platform layout, database connection parameters etc. The other is the experimental data. The data of each experiment will be published in the format similar to that stored in ER database. The message format in an experiment has the tree-structure stored in the XML file where the

parameters will be defined. It includes experimental number, attributes of experimental setting (e.g. parameters and assays) and results. An example of message on the blackboard for lysozyme precipitation experiment is shown in Figure 2.5. The agent selects all relevant experiment data both from database and other agents and displays them on blackboard. Other agents can access to update specific fields, e.g. experiment result, when they are available.

*The Execution Agent* controls the liquid handling system, e.g. Tecan Freedom 200, to add the different reagent solutions into the microwell plates, the plate shaker to perform mixing related tasks, and the vacuum pump to perform separation tasks.

*The Assay Agent* controls assay devices, e.g. the plate reader and HPLC. It communicates with the Coordinate Agent to carry out the experiments for calibration if there is insufficient data from Coordinate Agent. Based on the calibration results, the agent reads data from the assay devices and calculates the desired readings of each well.

*The Experiment Design Agent* comprises machine learning algorithms. Based on these algorithms, the agent publishes the newly designed experiments on blackboard for execution. It also carries out data analysis and process evaluation.

## 2.3.2 Database and data structure design

### 2.3.2.1 ER database design

The data and knowledgebase is designed to store three different types of information: 1) historical experiment data including experiment conditions, input parameters, and experiment result; 2) device configuration data such as layout information about different devices on LHS, connection parameters for the plate reader; 3) general knowledge, e.g. pH buffer mixture settings. The information is stored in database following ERM. A typical experiment involves chemical reagents, labware and automated platform configuration, experiment design and results. The entity-relationship model designed for the experiment in

the database is shown in Figure 2.5.

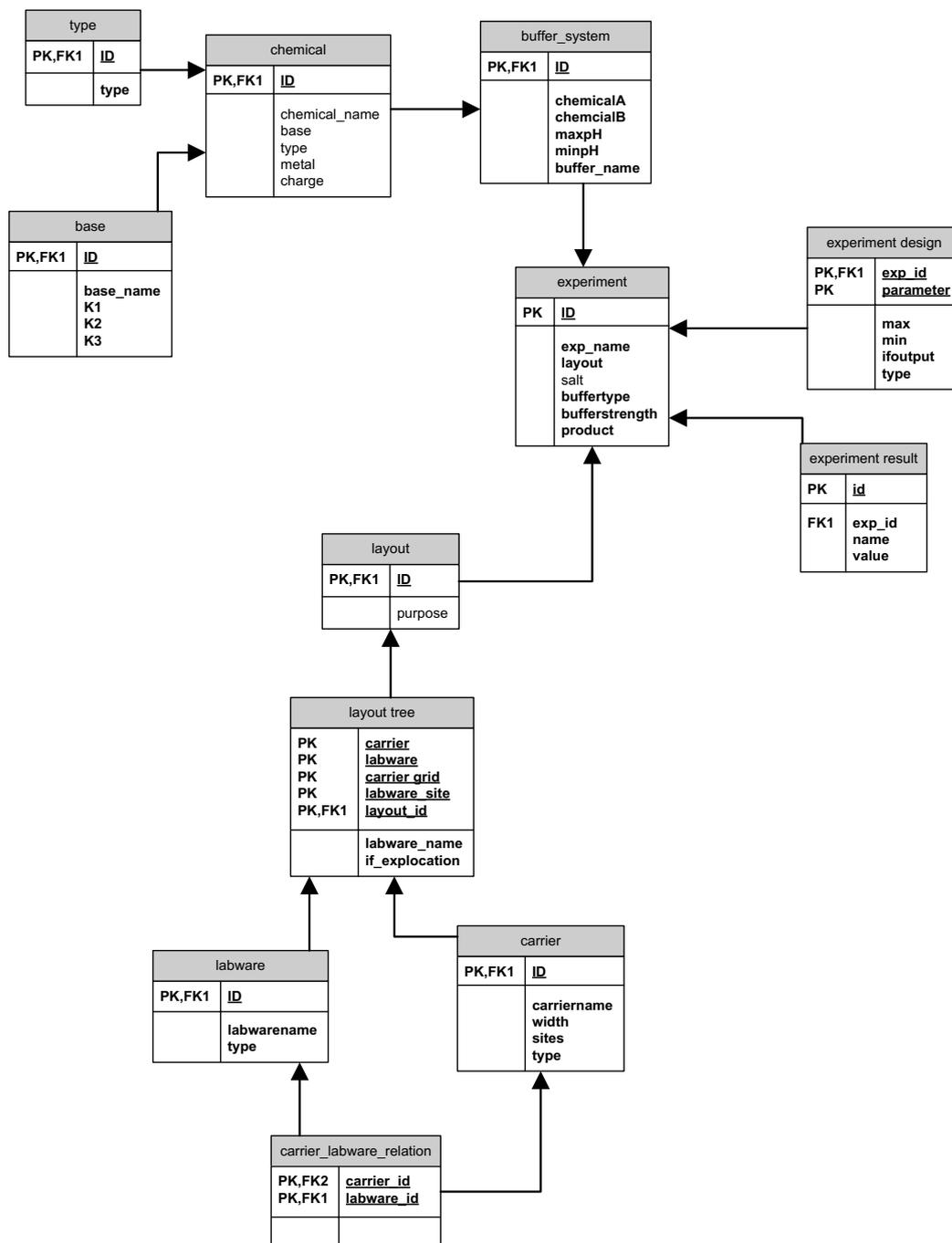


Figure 2.5 Entity-relationship model design for the database

### 2.3.2.2 Data structure of experiment

Experiment data include inputs and outputs as well as other information such as ID. The data structure must be defined to allow agents to recognise them and dispose data correctly. Four classes are designed as shown in Figure 2.6.

Figure 2.6(1) defines the format of experiment output variables where “name” means the parameter’s name and “value” means the measurement of corresponding output parameter. Default value is -1 means the output value has not been obtained.

Figure 2.6(2) defines the format of input parameters where “type” means the name of input. The operating range of the corresponding input parameter is given by “min” and “max” value to define design space.

Figure 2.6(3) defines data structure for processing experiment design information. The binary variable of “complete” indicates whether the experiment is finished. The parameters involved in the experiments are defined in attributes “inputs” and “outputs”, which are the lists of instances of “input” class and “output” class. The method of “read\_record” is a function to check from database if the experiment has already been done previously. The method of “record” is to store the experiment result into database after the experiment has been completed.

Figure 2.6(4) defines the other experiment parameters like buffer type, buffer strength, and platform layout.

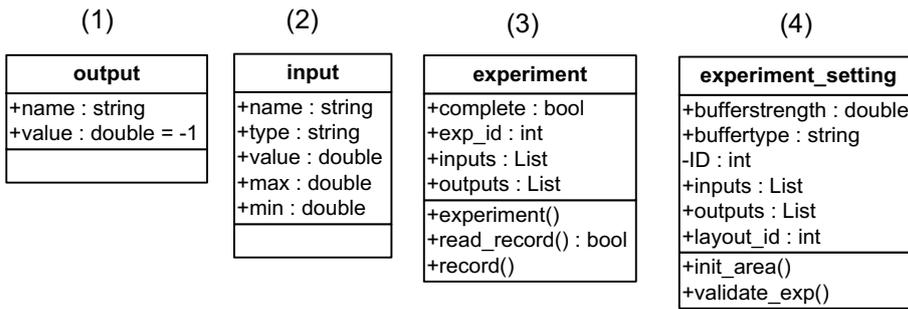


Figure 2.6 Classes related to experiment.

Corresponding to the class design, the ER model of the experiment in the database is extracted from Figure 2.5 and shown in Figure 2.7. The table of “experiment” records the basic configuration information. The table of “experiment design” consists of all the parameters involved in the experiment. The historical experiment results are stored in the table of “experiment result”. Each experiment is assigned with a unique globally unique identifier (GUID) code.

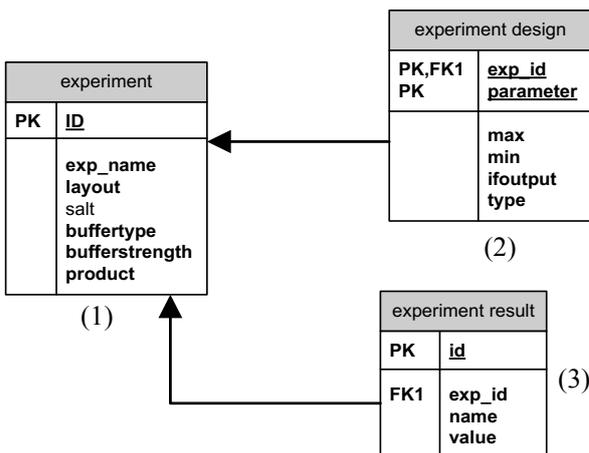


Figure 2.7 Tables in IAPBD’s database that related to experiment

### 2.3.2.3 Data structure of device information: LHS layout

The robotic arm moves within a defined area in the platform to deliver tasks. A set of coordinates on platform has been defined to allow the robotic arm deliver tasks accurately, e.g. to add reagents to defined locations. There are two types of instruments and devices in the platform. One is “labware” like disposable tips racks, microwell plates, filter plate etc. The other type is called “carrier” where the labware can be located. The locations of both carriers and labware is represented by tree structures. As the example shown in Figure 2.8, each layout of these carriers and labwares has a unique ID in the database, which acts as the highest level of the tree structure. Carriers are in the middle level, while labware is at the lowest level.

The coordinates of the platform consist two variables, *grid* and *site*. Each grid can only be occupied by one carrier in a layout, however each carrier can occupy more than one grid dependent on the width of the carrier. The number of sites depends on the locations on a carrier where labware can be placed on.

These classes in the database are then designed according to the layout structure as shown in Figure 2.9. Figure 2.9(1) defines the configuration parameters of the carrier: grid, width, site number in this carrier, and a list of labware located on the carrier. Figure 2.9(2) defines the configuration parameters of labware: site, type and name. Figure 2.9(3) is a list of carriers within the layout. The method “load\_layout” is to load a layout from database by a given layout ID. The method of “layout\_script” is to translate the layout into Tecan Script format, which can be recognized by Tecan LHS.

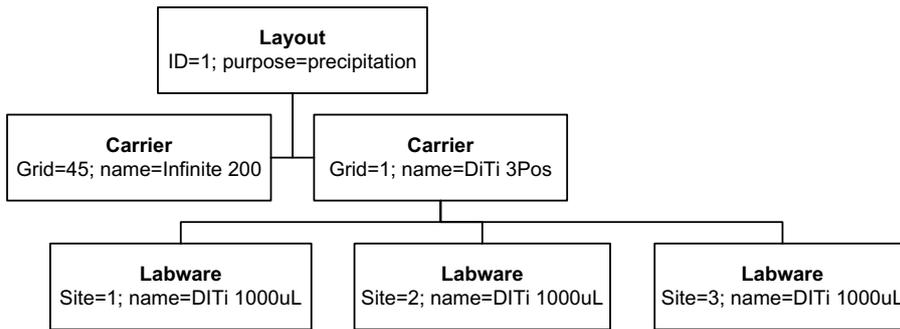


Figure 2.8 An example of the layout structure with a carrier carrying three disposable tip racks and an *Infinite 200* plate reader connected

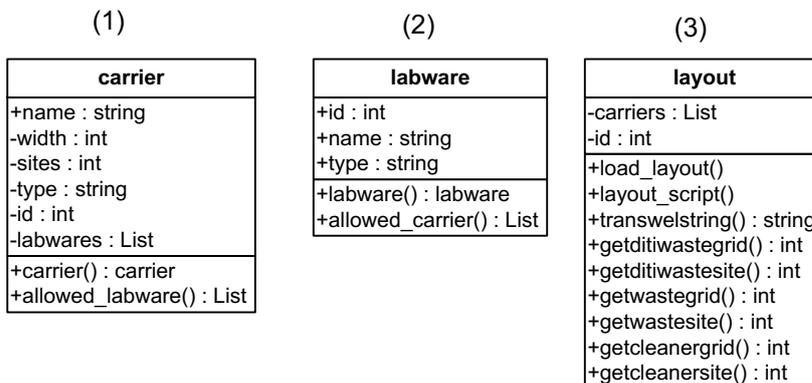


Figure 2.9 Classes related to tree-structure layout.

The ER model between the carrier and labware in the database is shown in Figure 2.10. The tables of labware and carrier have the information of every devices and equipment in the system. The table of “carrier\_labware\_relation” stores which labware can be located on each carrier.

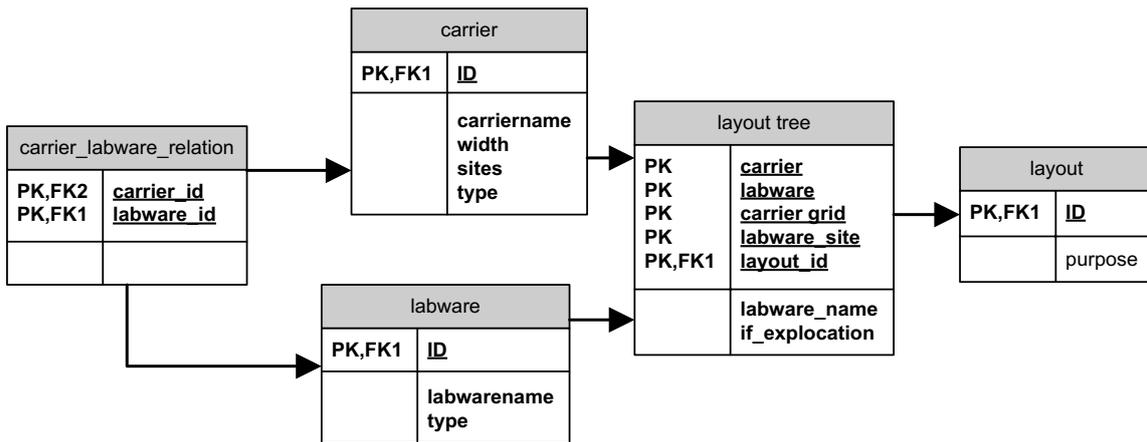


Figure 2.10 Tables and relations for layout in the database.

The data structure designed provides a way for different agents to exchange information efficiently so that each agent is able to translate information from blackboard into the formats that can be understood.

### 2.3.3 Agent design

#### 2.3.3.1 Execution agent

The Execution Agent is responsible for carrying out the experiments. Its major task is to manage connection between platform and laboratory devices. As discussed in previous chapter, API can be used for program communication. The functions can be shared by each other. For automated devices with API, the Execution Agent can control them directly. Although manufacturers provide software to control their automated devices, the APIs are often not included. Then it is not possible for the agent to get direct control of those devices. However, some software has functionalities to call a third party executable file and to take return data into its program during operation. This makes it possible to insert data into their software. Therefore, four solutions have been developed in the architecture of Execution Agent, shown in Figure 2.11, for the agent to connect with experiment execution devices.

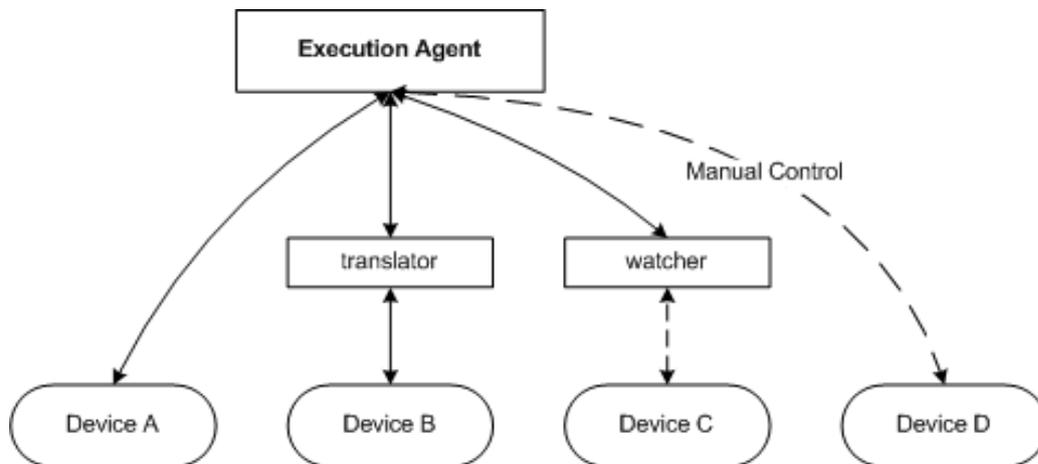


Figure 2.11 Architecture of the Execution Agent

1) The agent has full access to the devices (Device A in Figure 2.11). In this case, our program can directly control the devices without any problem. For example, Tecan provides APIs that allow the third party program to send commands directly to the Tecan LHS. Based on the APIs, I designed a class to compose and send operating scripts, which can be understood by Tecan LHS. Figure 2.12 shows the class diagram for controlling Tecan LHS. The commands include all the basic operations such as aspirate, dispense, drop tips, get tips etc.

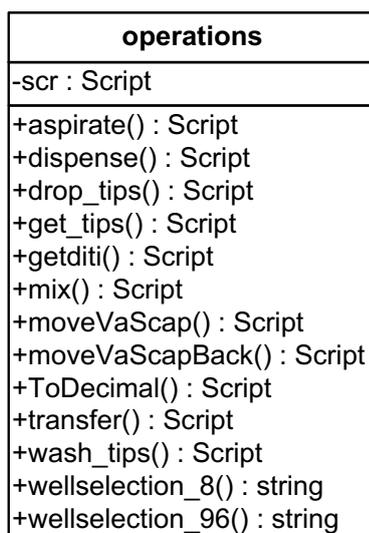


Figure 2.12 Class diagram for controlling Tecan LHS.

2) The agent has no access to the devices (Device B in Figure 2.11), but the original controlling software allows calling the third party program during the operation. For example, older model of Tecan LHS controlled by software called Gemini. In this case, since the agent cannot command the device, a ‘translator’ is developed to pass values of parameters from original software to the platform. The ‘translator’ is a small program that will be called by the original software during the operation. As shown in Figure 2.13, although the agent cannot control LHS directly, a translator will be called from LHS and return back the information needed.

3) The agent has no access to the device (Device C in Figure 2.11), and the original software has no functionalities to call third party program during operation, but the device software will generate some log files after operation. In this case, a program called ‘watcher’ is designed to monitor a pre-defined system path. Whenever there are new data generated in the path, the ‘watcher’ will automatically report status to the agent.

4) The agent has no access to the device, and there is no log file generated after operation. In this case, the status parameters will be report to the agent manually.

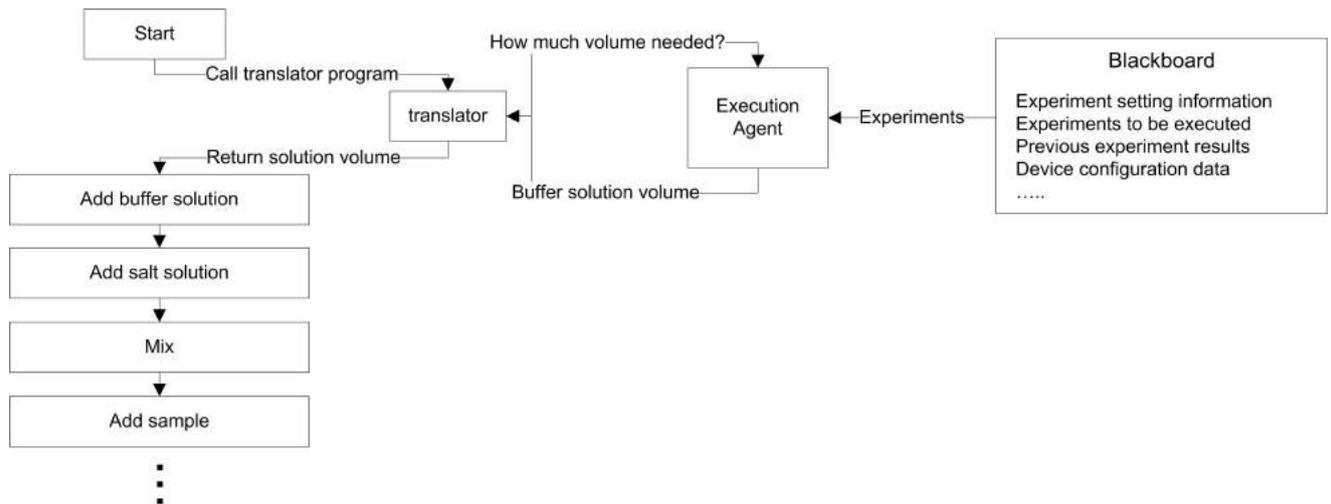


Figure 2.13 A simple example illustrating how translator works together with Execution Agent to pass information into LHS during operation

### 2.3.3.2 Experiment Design Agent

A predefined interface has been designed for algorithms to follow as shown in Figure 2.14. The interface makes it very convenient to integrate IAPBD with other algorithms since no additional modification is needed in source code. The interface has six variables and one method: 1) name: this is the name of algorithm, which will be stored in database; 2) sample experiments: this contains experiment setting data e.g. input and output parameters, to guide experiment design; 3) working experiments: this shows the list of experiments currently in memory and being analysed in which the unfinished experiments will be published on blackboard for execution; 4) finish: this parameter indicates if the iteration has been finished; 5) method: this shows the method that the algorithm employs to process and design experiments. This general interface is designed to allow any new algorithm to be added with minimum effort as long as the format is followed.

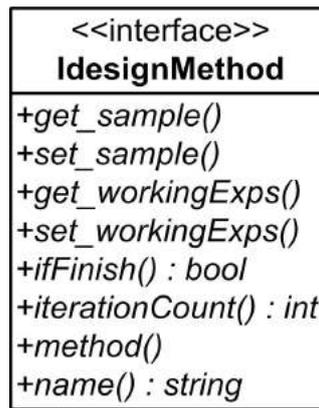


Figure 2.14 Interface design for experiment design algorithms.

The Experiment Design Agent will communicate with other agents through publishing data in the blackboard. When a new set of experiments has been designed, the experiment conditions with data structure of experiment will be sent to blackboard. The Execution Agent can pick up the information to plan ahead. Once all the working experiments have been labelled as completed, the agent will start data analysis and process evaluation based on selected design method. And the agent will decide whether optimal solution has been achieved. If not, it will publish new set of experiments to be executed for further evaluation.

### 2.3.3.3 Assay agent

The Assay Agent is similar to the Execution Agent. It is responsible for managing different assay devices, and one of the key tasks is to retrieve the data from the devices. The architecture of the Assay Agent is similar to the one of the Execution Agent as shown in Figure 2.15. Because not all manufacturers provide API access to their analytical devices, the “translator” and “watcher” techniques have been used to fetch experiment results.

Additional information may be needed to convert raw data collected (e.g. UV absorption) to parameter values (e.g. protein concentration). The agent is capable of publishing new experiments on blackboard to establish calibration of the analytical devices. After the

experiments have been executed on the platform, the raw data collected from these experiments is used to generate calibration curve for those experiments results, which are from Experiment Design Agent. In the case that it is not feasible to carry out calibration experiments on automated platform, the experiments will be executed manually before writing relevant calibration data into database.

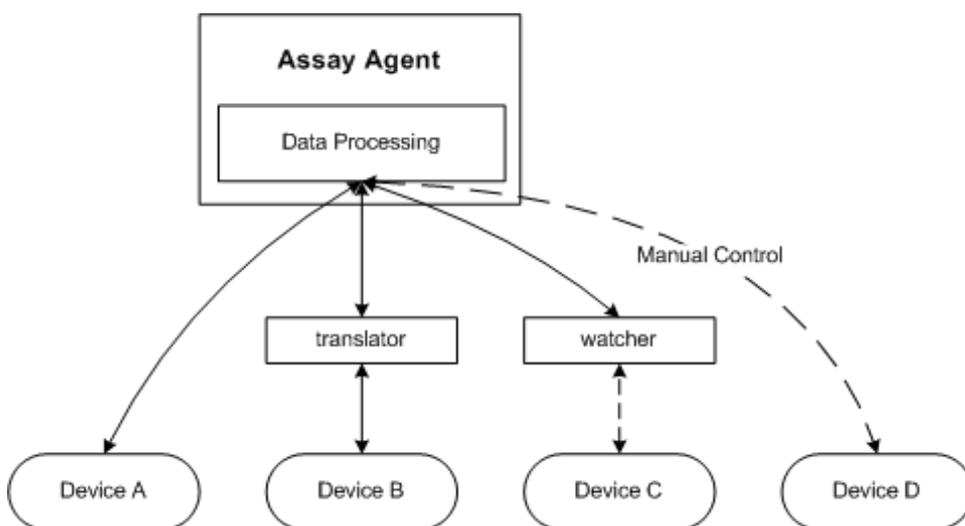


Figure 2.15 Architecture of Assay Agent

### 2.3.3.4 Coordinate Agent

The Coordinate Agent manages the database by retrieving historical experiment data and configuration data, writing new executed experiment information from blackboard into the database. Besides database management, the Coordinate Agent manages the information published on blackboard. This agent is also responsible for some calculations based on general background knowledge that stored in the database, e.g. calculating desired volumes for each solution when preparing buffer solution.

The blackboard contains the information of experiment settings, previous results and automated devices configuration data. It is an information-sharing place and other agents are

allowed to access it, obtain new tasks and publish results on it. As illustrated in Figure 2.4, Execution Agent, Experiment Design Agent and Assay Agent are all communicating with the Blackboard. For example, when the Experiment Design Agent completes an experiment design, it will publish the experiment design information on the blackboard. The message is formatted in XML.

A message showing an experiment published on blackboard by Experiment Design Agent is illustrated in Figure 2.16. From left to right, the objective function value is 0.5. The experiment's objective function is defined by  $0.4*a+0.6*b$ , where 'a' is the salt concentration and 'b' is the lysozyme concentration. The parameters investigated in the experiment are salt concentration and lysozyme concentration. The parameter consists of design space information (maximum & minimum values and interval) and the actual value. The protein involved in the experiment is lysozyme and its peak appears at 15 min when using HPLC to measure the concentration.

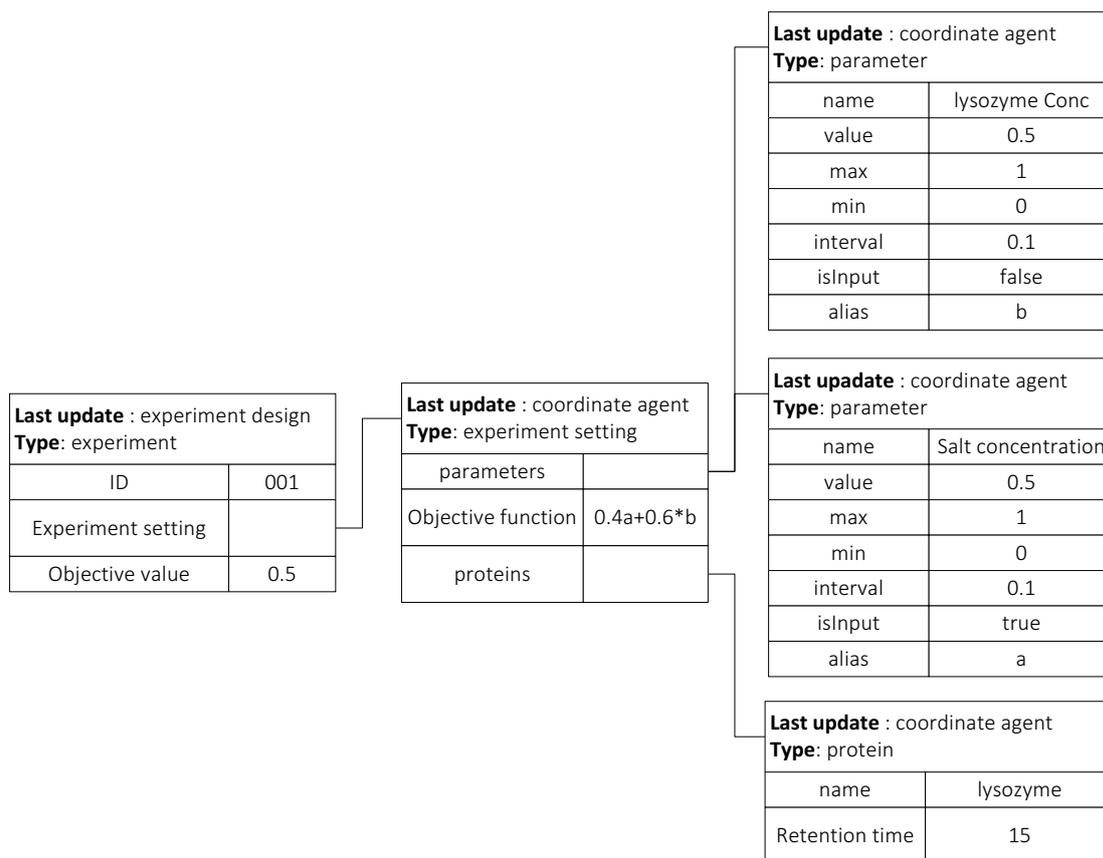


Figure 2.16 An example of messages on blackboard showing an experiment published by the Experiment Design Agent.

Buffer preparation is a one of the common tasks in biological experiments. Traditionally when buffer solutions for different pH value is prepared before experimentation manually. For example, when preparing phosphate buffer solutions with various pH values, researchers have to check the volumes of monosodium phosphate and disodium phosphate solution needed from a predefined table, and then prepare each buffer solution separately by mixing two reagent solutions manually. However, it is not feasible in closed-loop learning approach. Each experiment is design dynamically by the algorithms, which means the pH value needed in the experiment is unknown before the experiment and changing at each iteration. Thus the buffer is prepared by mixing two solutions of specific volumes calculated in real time based on the knowledge in the database. The Coordinate Agent is designed to be capable of calculating desired volumes of each reagent solutions for

preparing buffer solution. The calculation performed by the Coordinate Agent is shown below.

If a buffer is composed of  $H_nA^{N-n}$  ( $n = 0 \dots N$ ), where  $H$  is hydrion,  $n$  equals the number of hydrion,  $A$  is the acid radical, and  $N$  equals the total positive charge of  $A$ , its dissociation constant  $K_n$  is:

$$K_n = \frac{[H^+][H_{n-1}A]}{[H_nA]} \quad n = 1 \dots N$$

$$K_{n+1} = \frac{[H^+][H_nA]}{[H_{n+1}A]} \quad n = 1 \dots N$$

...

$$K_N = \frac{[H^+][H_{N-1}A]}{[H_NA]} \quad n = 1 \dots N$$

At a given value of pH, these expressions give the relative ratios of each component:

$$\frac{[H_nA]}{[H_NA]} = \frac{K_{n+1}}{[H^+]} * \frac{K_{n+2}}{[H^+]} * \dots * \frac{K_N}{[H^+]}$$

$$\frac{[H_nA]}{[H_NA]} = \frac{K_{n+1}K_{n+2} \dots K_N}{([H^+])^{N-n}}$$

Assume  $B$  is the buffer strength:

$$B = \sum_{n=0}^N [H_nA]$$

$$B = [H_NA] \sum_{n=0}^N \frac{[H_nA]}{[H_NA]} = [H_NA] \sum \frac{K_{n+1}K_{n+2} \dots K_N}{([H^+])^{N-n}}$$

Thus, the actual concentrations can be obtained by following equations:

$$[H_nA] = \frac{K_{n+1}K_{n+2} \dots K_N}{([H^+])^{N-n}} [H_NA] = \frac{K_{n+1}K_{n+2} \dots K_N}{([H^+])^{N-n}} * \frac{B}{\sum \frac{K_{n+1}K_{n+2} \dots K_N}{([H^+])^{N-n}}}$$

$$[H_nA] = \frac{B}{([H^+])^{N-n} \sum \frac{1}{([H^+])^{N-n}}}$$

For buffer solution make of  $M_{N-b}H_bA$  and  $M_{N-a}H_aA$ , we have:

$$[H^+] + x * (N - b) * B + (1 - x) * (N - a) * B = [OH^-] + \sum_{n=0}^N (N - n)[H_nA^{n-N}]$$

where  $x = \frac{\text{volume of } M_{N-b}H_bA \text{ solution}}{\text{total volume of buffer solution}}$

Thus, the require volume of  $M_{N-b}H_bA$  and  $M_{N-a}H_aA$  can be calculated by solving the equation above. The information of buffer system (i.e.  $K_n$ ) is stored in the database. The design of the tables and relations is shown in Figure 2.17. The system will firstly search the preferable buffer system in the corresponding experiment setting. If there is no preference, the system will give all the available buffer systems within the feasible range of pH value.

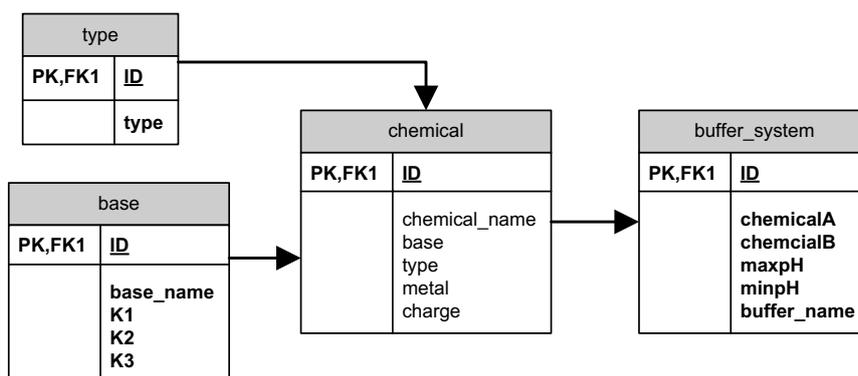


Figure 2.17 Tables containing buffer information in the database

## 2.4 Quality assurance for IAPBD

As a fully automated solution, quality assurance is an essential aspect for IAPBD. A series of measures have been introduced to improve the quality of bioprocess designs generated from IAPBD.

1. The agents were tested on simulated virtual devices before linking them with real instruments. Using the simulation packages provided by Tecan EVOware, experiments were firstly carried out on simulated LHS with 3D view, and each step was monitored and verified manually. The design algorithms in Experiment Design Agent were tested by simulation data generated from Assay Agent.
2. A routine check will be performed on LHS each time before running real experiments. The check is to perform series pipetting tasks on a microwell plate. Then, %bias and %CV will be measured to evaluate performance of LHS as reviewed in Chapter 1.
3. During execution of experiments, the liquid level in each microwell will be measured after each pipetting task. The measured volume will be compared with theoretical value to make sure that %bias is within defined range (i.e.  $\leq 5\%$ ). IAPBD will be suspended if abnormal %bias was observed.
4. MAD of replicates' result from Assay Agent will be calculated for outlier detection as reviewed in Chapter 1. The design algorithm will ignore data from outliers.
5. IAPBD will be suspended when hardware failure has been detected. Since automated devices (e.g. LHS, HPLC, plate reader) usually have failure detection functions built in, IAPBD continuously monitors hardware status and suspends all operation in the case of hardware failure.

## 2.5 Conclusions

An intelligent automation platform with multi-agent architecture is established to realise the

closed-loop learning approach in bioprocess development. An ERM based database has been designed to store the data of historical experiment results, system configuration and other general knowledge. Five agents in the platform communicate with each other via the blackboard mechanism, where everyone can read and write new messages for experimentation progress. The platform integrates the current independent parts including automated devices, analytical instruments and intelligent experiment design algorithms together to perform the bioprocess design tasks. The platform learns from each batch of the experiment results, and tries to achieve better bioprocess design based newly gained knowledge iteratively. There are several significant advantages of this approach:

- Researchers are completely free from the simple and repeatable tasks such as liquid handling and plate transfer.
- Experimental data can be well organized and analysed.
- Integrating the microscale instruments into the platform would reduce the material cost.
- Different algorithms or methods related to experiment design could be implemented quickly and straightforward.

In next three chapters, the agent development will be discussed further through several case studies and the performance of IAPBD will be evaluated using real examples.

## Chapter 3 IAPBD verification by a lysozyme precipitation process design

### 3.1 Introduction

The intelligent automation platform established in Chapter 2 based on the closed-loop framework and the multi-agent architecture is designed to be capable of driving all relevant automated devices, analytical instruments and experiment design and data analysis algorithms. A case study of a precipitation process has been carried out on the platform for the purpose of verification and demonstration. This case study is deliberately chosen for its simplicity so as to verify the functions of the IAPBD. Although the precipitation experiment is very simple, it has several common operations such as mixing, incubation, and filtration in sample preparation that involved in the most of experiments. A plate reader has realized the measurement of protein concentrations.

The experiment is to investigate the solubility of lysozyme so as to identify the optimal ion strength and pH in the precipitation process. Ammonium sulphate and phosphate buffer are used to achieve the ion strength and pH values. The goal of the experiments was to find an optimal set of conditions that maximizes the yield of lysozyme in supernatant and maximizes the ammonium sulphate concentration. Here maximizing the ammonium sulphate concentration reflects maximizing the purity for a crude industrial feed. The objective function,  $J$ , is shown in equation (3.1).

$$J=\alpha(c/c_0)+\beta(s/s_0)$$

where  $\alpha$  and  $\beta$  are weighting factors,  $c_0$  the initial lysozyme concentration,  $c$  the final lysozyme concentration in the supernatant,  $s_0$  the maximum ammonium sulphate

concentration, and  $s$  the ammonium sulphate concentration used in the specific experiment. The objective function can also represent the scenario where both yield and purity are maximised in complex multicomponent precipitation processes.

In this chapter, the material and methods for the precipitation will be detailed in section 3.2, which describes the procedures for the microscale precipitation experiment. The agent designs to realise these procedures automatically and IAPBD implementation will be given in section 3.3. In particular, the design of the simplex search algorithm and the ANN algorithm in Experiment Design Agent to facilitate the closed-loop learning will be described in details. The integration of agents and the automated devices and instruments is straight forward as their APIs are available for use. The experimental results achieved by IAPBD will be shown in section 3.4 followed with the conclusions.

### **3.2 Material & Methods**

For our experiments the ammonium sulphate, monosodium phosphate, and disodium phosphate were obtained from Sigma-Adrich. The ammonium sulphate solution was prepared at a concentration of 4 mol/L. The phosphate buffer was chosen to maintain the desired pH value. The monosodium phosphate and disodium phosphate solution are firstly prepared at the concentration of 300 mM then diluted. The lysozyme was obtained from Sigma-Adrich and prepared at 20 mg/mL. The 96-well Multiscreen filter plates were sourced from Millipore (Hertfordshire, U.K.) and the 96-well flat bottom Costar UV Microwell plates from Corning (Leicestershire, U.K.).

The experiment involves adding 100  $\mu$ L of lysozyme solution and 50  $\mu$ L of buffer solution, a mixture of monosodium phosphate and disodium phosphate, into the microwell. Thus, the initial lysozyme concentration in each well is 6.67 mg/mL. The ammonium sulphate concentration ranges from 0 to 1 mol/L. The range of pH is between 6 to 8. The interval of ammonium sulphate concentration was 0.1 mol/L and the interval of pH is 0.1. The incubation time was 2 hours.

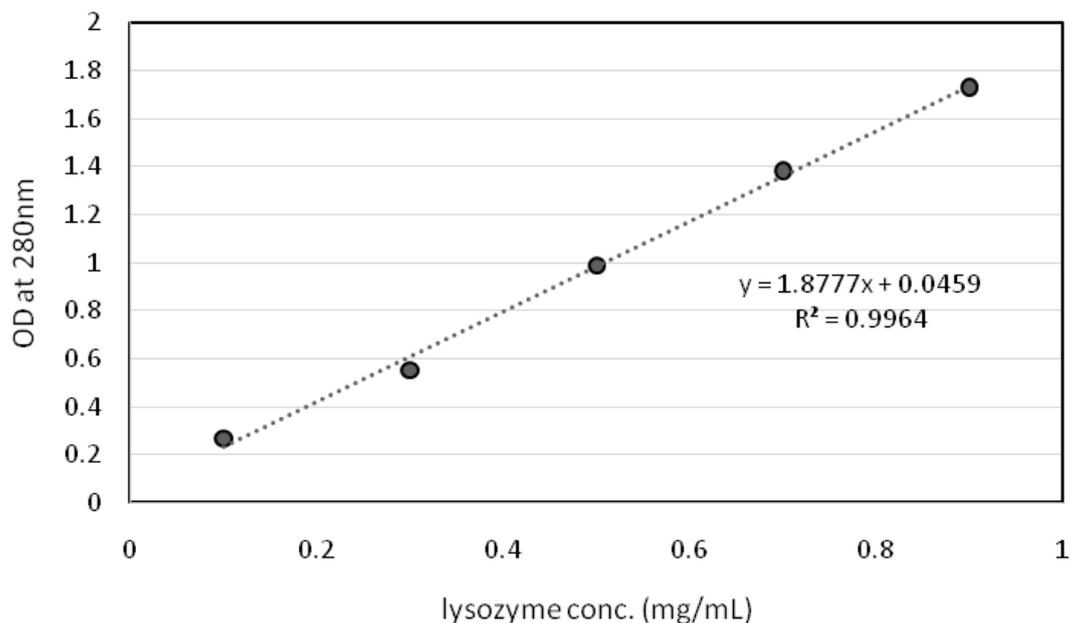


Figure 3.1 Calibration curve of the absorption at 280 nm and lysozyme concentration.

A Tecan Freedom 200 LHS executes all of the liquid handling tasks. The hardware configuration of the intelligent automation platform for precipitation is shown in Figure 3.2. A Tecan Infinite 200 plate reader is used to detect the lysozyme concentration at absorption of 280 nm. The calibration curve shown in Figure 3.1 confirmed that the absorption at 280 nm was good for lysozyme measurement. A vacuum separator from TecanVacS series is performed for the filtration operation.

The simplex search method and artificial neural network (ANN) model-based method are chosen as the sequential experiment design methods in this study.

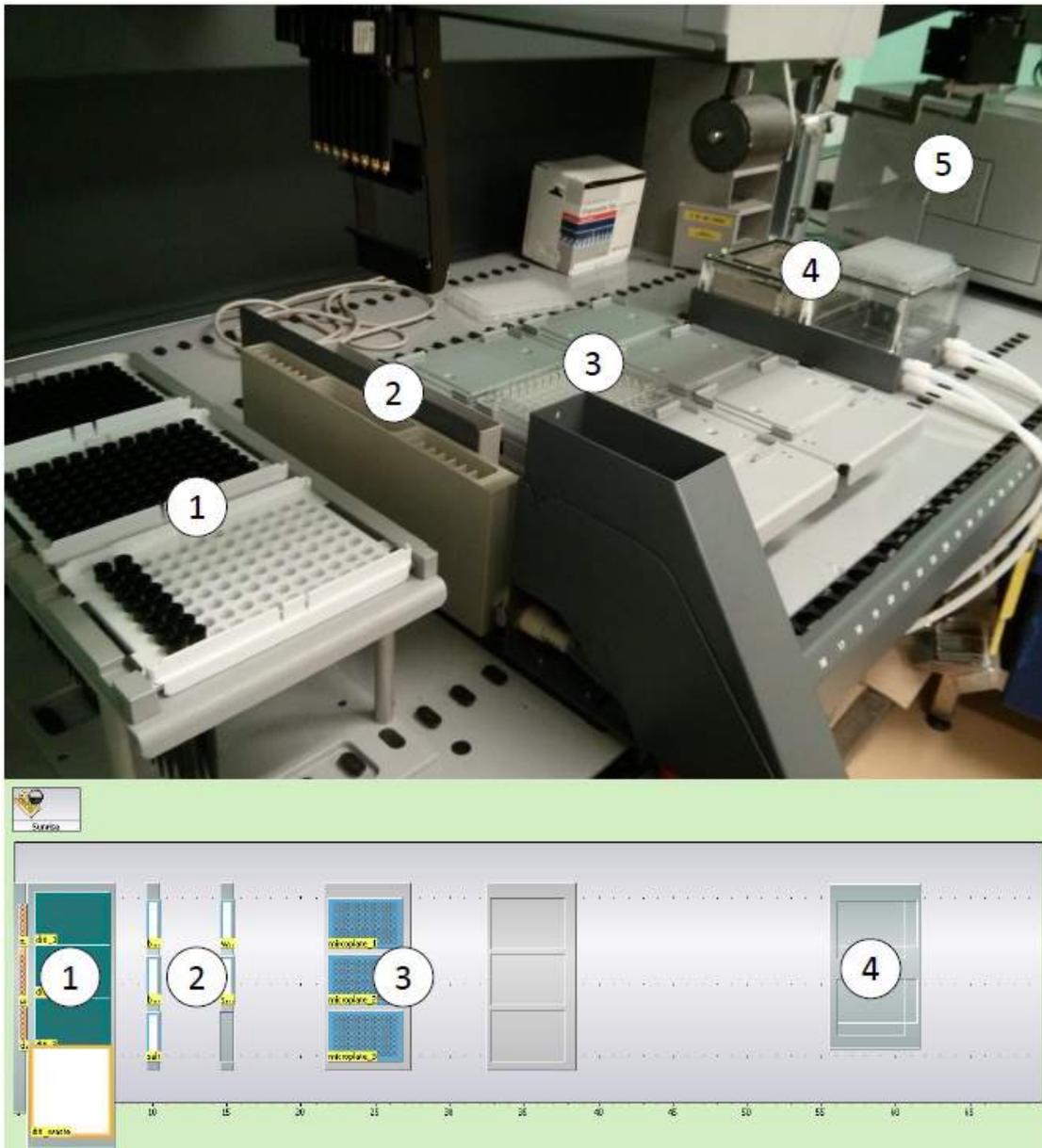


Figure 3.2 Hardware configuration of intelligent automation platform for precipitation. (1) disposable tips rack; (2) tube carrier; (3) 96-well microplates carrier; (4) vacuum filtration; (5) UV plate reader

### 3.3 IAPBD implementation for protein precipitation experimentation

Based on the framework from Chapter 2, the prototype for lysozyme precipitation process

development has been established and is illustrated in Figure 3.3. The data and knowledge base initially stores the system configuration information of the devices, e.g. Tecan configuration. The configuration information in the database consists of both connection parameters to realize communication among devices and the dimension and layout information to guide the movement of the robotic arms during the execution of experiments. Any experiment designed later can be translated automatically into executable programs to drive the devices. The experimental results obtained later will also be stored in the database. The database uses a ERM in Microsoft Access.

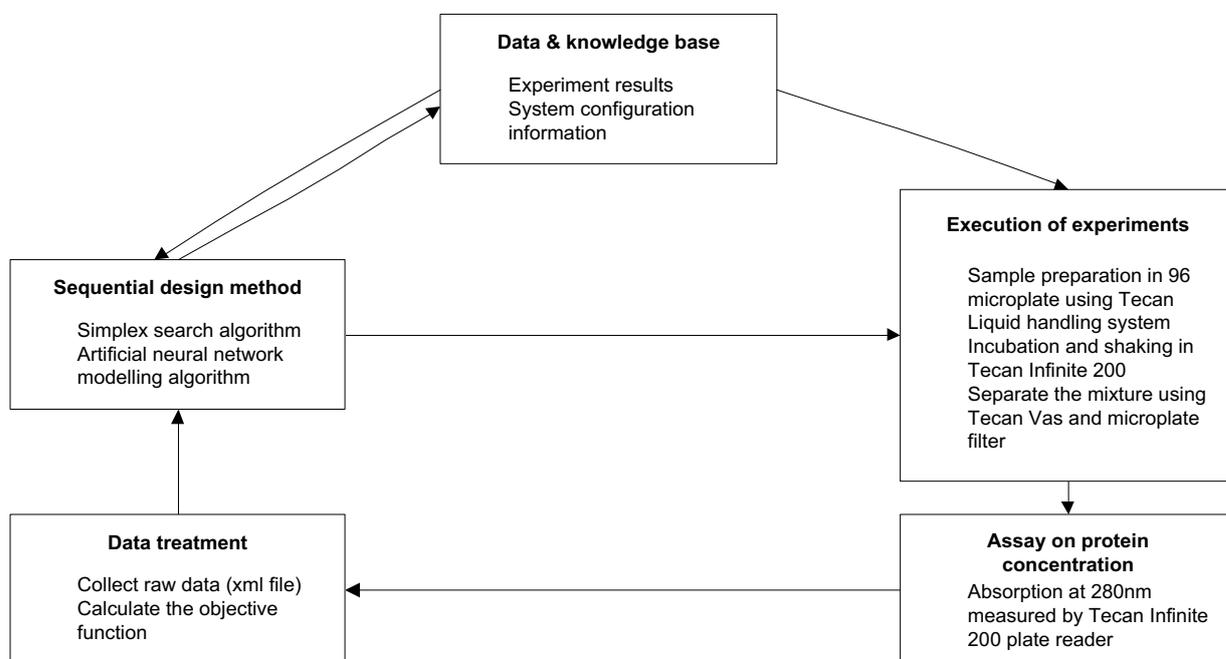


Figure 3.3 Multi-agent architecture used in IAPBD for precipitation process design

The experiments carried out include sample preparation in a 96 micro filter plate, incubation and shaking by plate reader, and liquid and solid separation of the samples by a vacuum pump sucking out the liquid from the bottom of the micro filter plate (Figure 3.3). All of these have been carried out by a Tecan LHS and driven by designated experimental procedure software in the EVOware.

The plate reader is located next to the Tecan and within the reach of the robotic arm. The lysozyme optical density is measured at 280nm UV absorption and the data stored in an XML file. Data treatment software that converts the raw data into lysozyme concentration based on calibration curves is also stored in the database.

The sequential design method uses the experimental data to calculate the objective function to decide if it is necessary to design a next round of experiments and use simplex search algorithm or artificial neural network modelling algorithm to design the next round experiments required. The experiment results and the newly designed experimental conditions are stored in the database for further use.

### 3.3.1 Multi-agent based architecture for precipitation process design

Based on the architecture explained in Chapter 2, the architecture for lysozyme precipitation process development has been established as shown in Figure 3.4.

*The Coordinate Agent* retrieves the historical experiment data from database and writes the new executed experiment results into the same database. It also translates the information in the database in the message format to blackboard.

*The Execution Agent* controls the robotic arm (Tecan Freedom 200) to add the lysozyme solution, phosphate buffer and ammonium sulphate solution into the microwell plate. The different pH values of each experiment were achieved by mixing suitably calculated volumes of monosodium phosphate and disodium phosphate solution. Then the agent

controls the LHS to put the microwell plate into incubator and controls the shaking speed and time. After incubation and shaking, the agent controls a vacuum pump on the platform to separate the precipitates and liquid. Finally the agent controls the robotic arm to take the sample of permeation to the plate reader to detect lysozyme concentration.

*The Assay Agent* controls the plate reader directly via its API. It communicates with the Coordinate Agent when the experiments carried out for calibration if there is insufficient data from the Coordinate Agent. Based on the calibration results, the agent reads data from the plate reader and calculates the lysozyme concentration of each well.

*The Experiment Design Agent* comprises two algorithms, which are the simplex search algorithm and the artificial neural network algorithm. Based on these algorithms, the agent publishes the newly designed experiments on blackboard for execution.

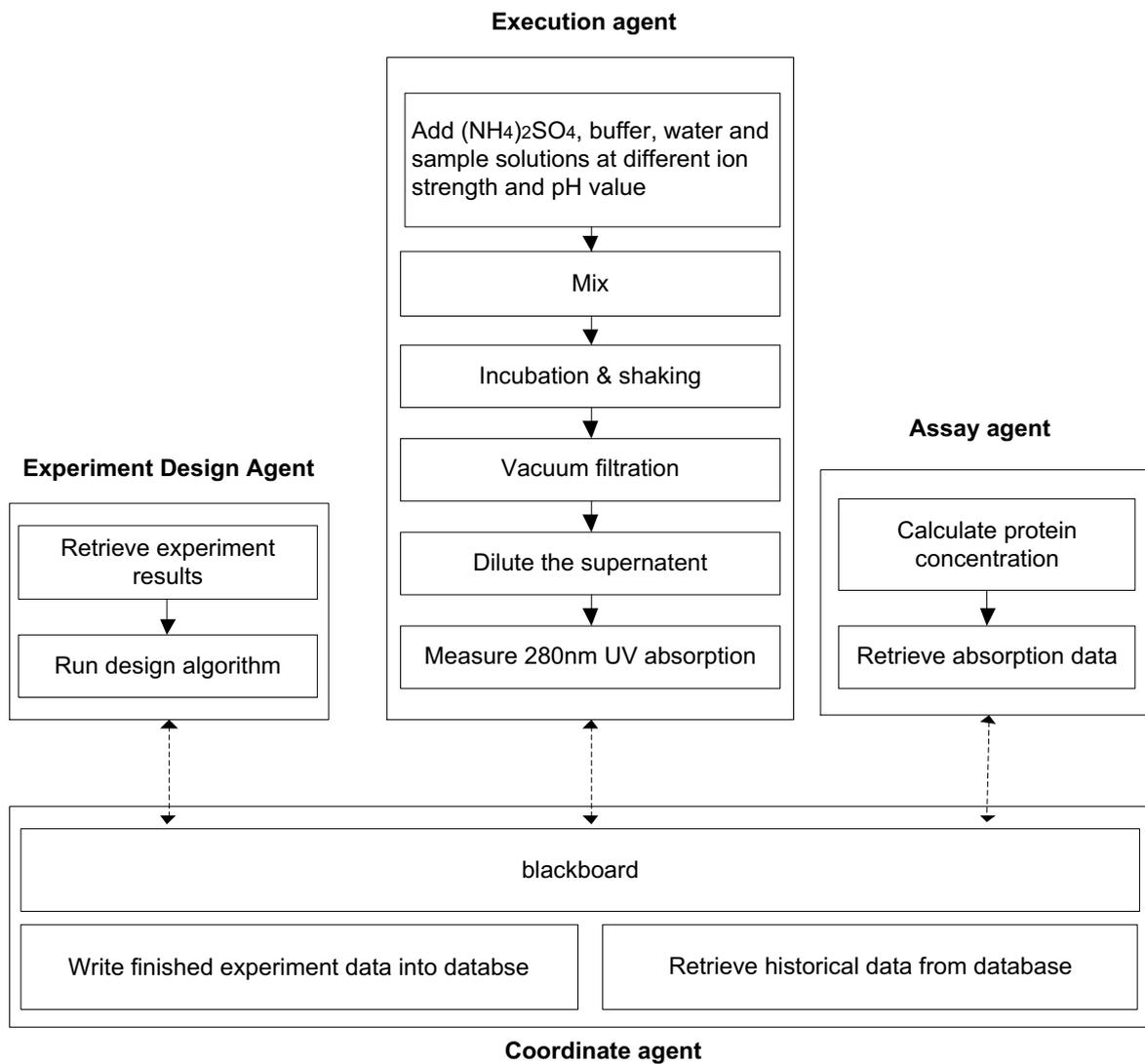


Figure 3.4 Multi-agent architecture for lysozyme precipitation process development

### 3.3.2 Experiment Design Agent: simplex method

The simplex algorithm, first proposed by John Nelder & Roger Mead (Olsson & Nelson 1975), is designed for minimizing an objective function in a multi-dimensional space. The technique was soon applied to solve bioprocess optimization problem (Ghodbane & Guiochon 1988; Konstantinidis et al. 2012; Horstkotte et al. 2010). The algorithm is significantly powerful as it only uses function values instead of the whole profile of response surface. It does not require the derivative term of objective function like majority search-based optimization algorithms. This feature makes the simplex algorithm popular in searching for optimal solution in experiment associated applications. Every iteration only one or few experiments are designed, which drastically reduces the experiment number compared with other experiment design methods. The procedures of the algorithm are detailed below.

0. Generate initial experiments.

Given the number of input parameters is  $n$ ,  $(n+1)$  experiments are generated where value of each input parameter can be randomly picked within the defined feasible range. The experiments are transformed into the vectors according to the value of inputs and outputs:

$$\mathbf{e} = [i_1, i_2, \dots, i_n, o_1, o_2, \dots, o_m]$$

These experiments are initial simplex. Add unfinished experiments in the simplex into set,  $S$ .

1. Execute
  - a. Remove the completed experiments from  $S$ .
  - b. Execute the experiments in set  $S$ .
2. Sort.
  - a. Check whether experiments in current simplex have been done. If not, add unfinished experiment to set  $S$ , and go to step 1.

- b. The objective function is  $f(\mathbf{e})$ . According to the value of objective function, sort the experiments

$$f(e_1) \leq f(e_2) \leq \dots \leq f(e_{n+1})$$

- c. Compute the variance of all the experiments ( $ss$ ) in the current simplex

$$ss = \frac{1}{n+1} \sum_{i=0}^{n+1} \left( f(e_i) - \frac{1}{n+1} \sum_{i=0}^{n+1} f(e_i) \right)^2$$

- d. If  $ss$  is less than the given accurate factor,  $i$ , then output the result and terminate the algorithm, else go to step 4.

3. Calculate the centre of the gravity of all points

$$e_0 = \frac{\sum_{i=1}^n e_i}{n}$$

4. Reflection

- a. Compute the reflected point

$$e_r = e_0 + (e_0 - e_{n+1})$$

- b. Check whether  $e_r$  has been done or not. If not, add  $e_r$  into  $S$ , and go to step 1.

- c. If the point is better than the worst, but not better than the best, i.e.

$$f(e_1) \leq f(e_r) \leq f(e_n)$$

Then obtain a new simplex by replacing the worst point with the reflected point, and go to step 2

5. Expansion

- a. If the reflected point is the best point so far which means  $f(e_r) \geq f(e_n)$ , then calculate the expanded point,

$$e_e = e_0 + 1.5(e_0 - e_{n+1})$$

- b. Check whether  $e_r$  has been finished or not. If not, add  $e_e$  into S, and go to step 1.
- c. If the expanded point is better than the reflected point,  $f(e_e) \geq f(e_n)$ , then obtain a new simplex by replacing the worst point with the expanded point, and go to step 2.
- d. Else, obtain a new simplex by replacing the worst point with the reflected point, and go to step 2.

6. Contraction

- a. Compute contracted point,

$$e_c = e_{n+1} + (e_0 - e_{n+1})$$

- b. Check whether  $e_r$  has been completed or not. If not, add  $e_c$  into S, and go to step 1.
- c. If the contracted point is better than the worst point, then obtain a new simplex by replacing the worst point with the contracted point, and go to step 2.
- d. Else go to step 8

7. Reduction

- a. For all but the best point, replace the point with

$$e_i = e_1 + 0.5(e_i - e_1)$$

Go to step 2.

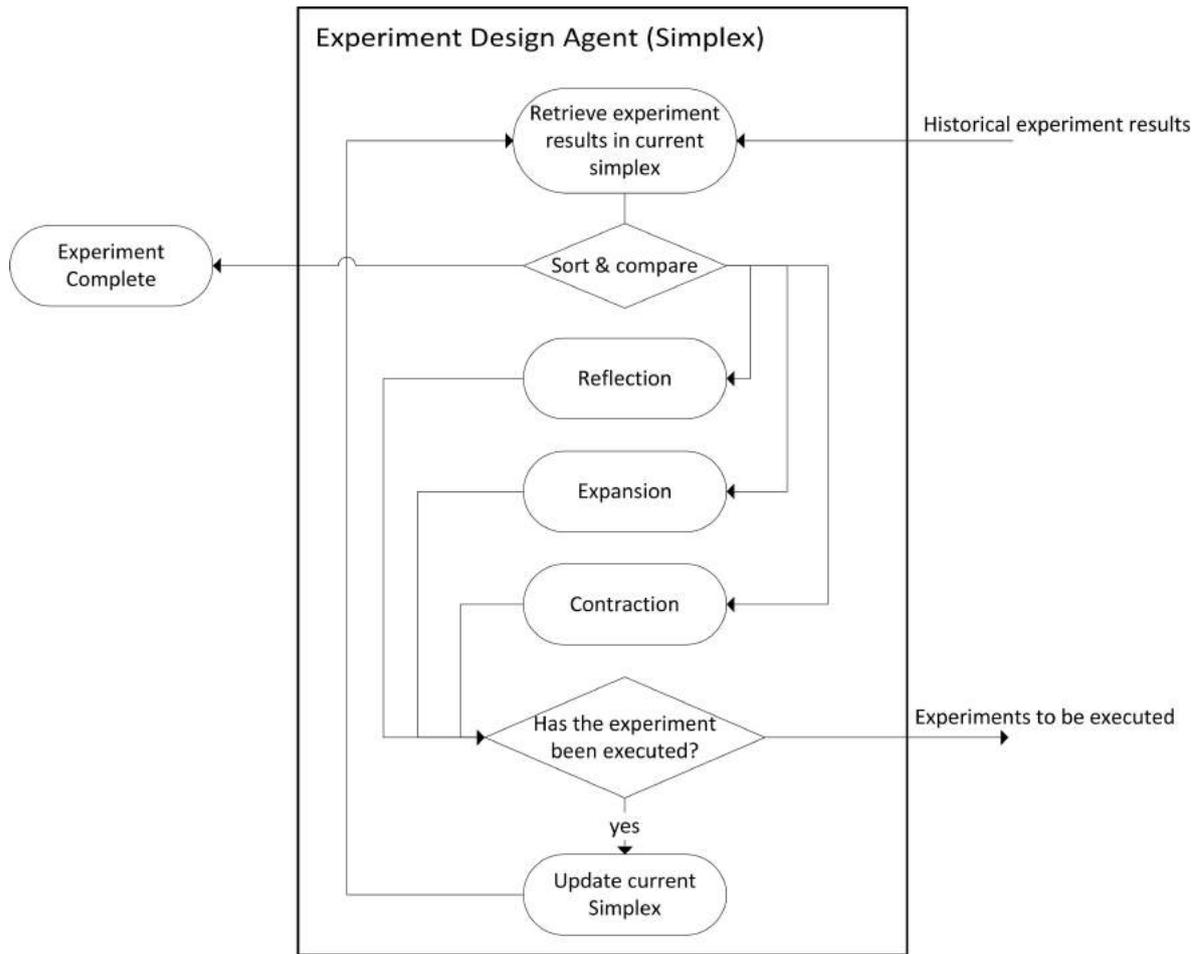


Figure 3.5 The experiment design agent to facilitate the simplex search algorithm

### 3.3.3 Experiment Design Agent: ANN

ANNs provide a range of powerful new techniques for solving problems in sensor data analysis, process identification, and control and have been used in a diverse range of chemical engineering applications (Pirdashti et al. 2013). Inspired by the way biological nervous systems, ANN is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. Feed-forward ANNs are one of popular architectures of neural network. As shown in Figure 3.6, Feed-forward ANNs allow signals to travel one way only, from inputs to output. There is no loop, i.e. the output of any layer does not affect that same layer. They are widely used to model non-linear systems (Khosravi et al. 2011).

The Back-Propagation Algorithm was invented to train ANNs, i.e. to adjust the weight of each neuron in such a way that the error between the desired output and actual output is reduced (Werbos 1990). The process requires that the ANN compute the error derivative of the weights (EW). EW describes how the error changes as each weight is increased or decreased slightly. The Gradient Descent algorithm can be then applied to train the ANN iteratively.

The Back-Propagation is used as the teaching methods of the ANN in this case study. The activation function for each neuron is sigmoid function as follows.

$$f(v) = \frac{1}{1 + e^{-v}}$$

Thus the output value for each neuron in the network can be expressed as follows.

$$X_i^k = f\left(\sum_j W_{ij}^k X_j^{k-1}\right)$$

where  $X_i^k$  is the output value of  $i$ th neuron at  $k$ th layer, and  $W_{ij}^k$  is the weight value between  $i$ th neuron in  $k$ th layer and  $j$ th neuron in  $(k-1)$ th layer.

The steepest descent method is used here to adjust the weight values in the network. First, the error function is defined in the following expression.

$$e = \frac{1}{2} \sum_i (X_i^m - Y)^2$$

where  $Y$  is the result from the experiment and  $X_i^m$  is the output given from the network. The adjustment for the weight value for one iteration can be expressed as follows.

$$\Delta W_{ij} = -\mu \frac{\partial e}{\partial W_{ij}}$$

The following equation explains how to compute new weight value from the last iteration.

$$W_{ij}(t + 1) = W_{ij}(t) - \mu \frac{\partial e}{\partial U_i^k} X_j^{k-1}$$

A single hidden layer artificial neural network is represented in Figure 3.6. There are two nodes in the input layer representing salt concentration and pH value. 10 neurons are in the hidden layer. The output layer contains one neuron since there is only one response parameter (lysozyme concentration in the supernatant).

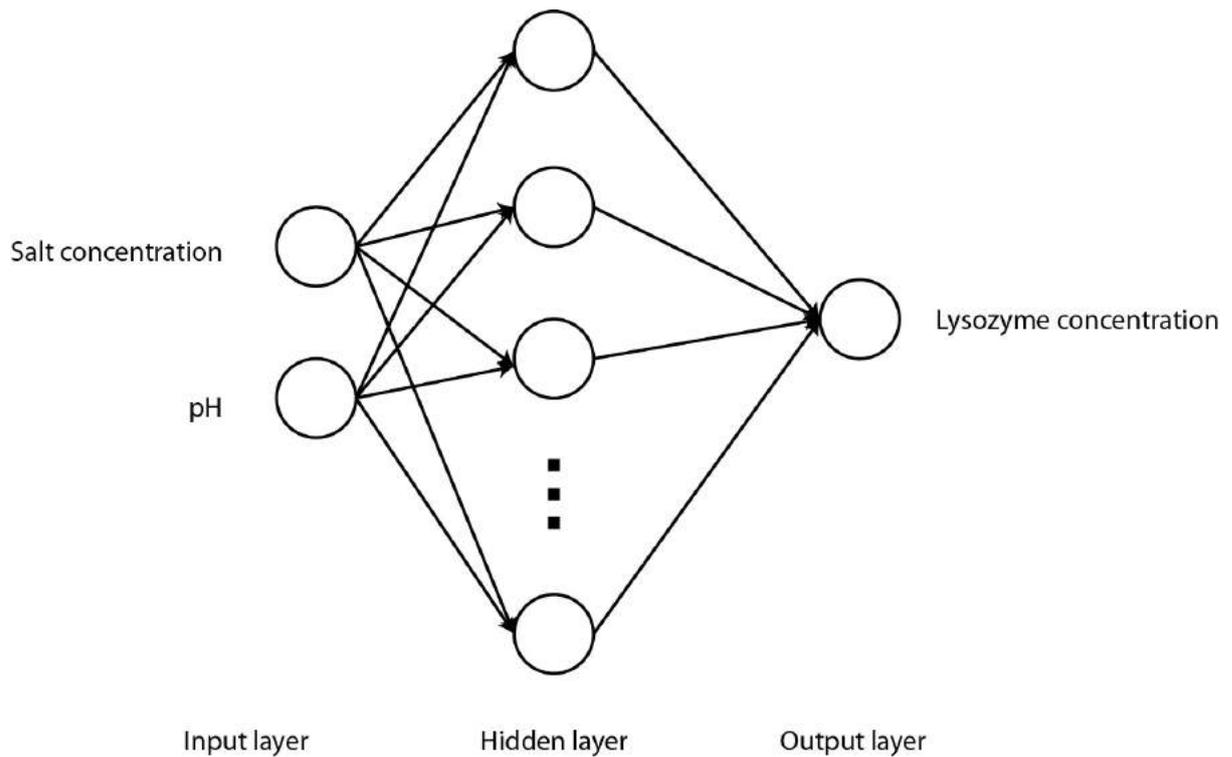


Figure 3.6 Schematic of a single hidden layer neural network

After defining the mechanism of the neural network, in order to save the experimental materials, the flowchart (Figure 3.7) demonstrates the procedures for the system to execute experiments in a batch-by-batch manner. The first three batches of experiments are firstly carried out on IAPBD as training set, validation set and testing set. The ANN is then trained by training set and validation set. If value of error function, which is calculated based on data in testing set, is larger than pre-defined threshold, another batch of experiments will be carried out on IAPBD as new testing set, while old testing set is used for training of ANN until ANN provides satisfactory prediction of lysozyme concentration.

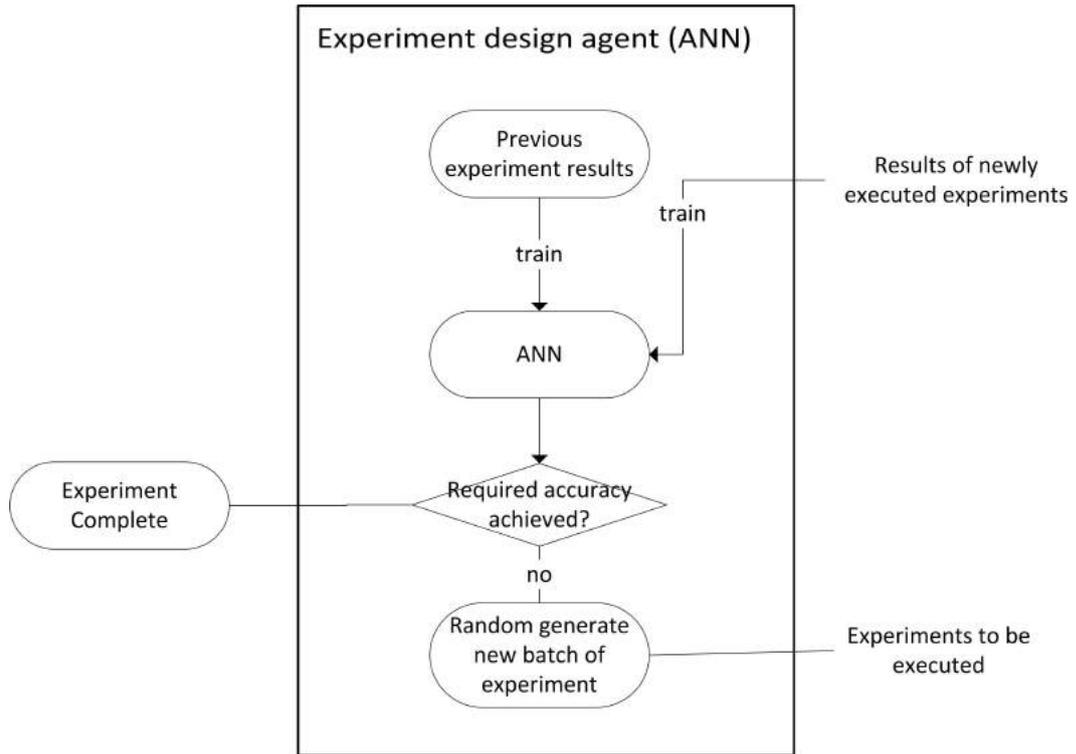


Figure 3.7 The Experiment Design Agent to facilitate the simplex search algorithm

### 3.4 The lysozyme precipitation process development

The Execution Agent and Assay Agent drove the TECAN liquid handling system and the plate reader automatically. The communication between the coordinator agent and other agents was smooth in facilitating data exchange. Two lysozyme precipitation studies were carried out. The Experiment Design Agent used the simplex search algorithm in the first study and the ANN algorithm in the second study.

#### 3.4.1 Precipitation optimization using the simplex algorithm

The initial simplex was selected randomly to mimic a new process design task when no prior knowledge was available. The Experiment Design Agent then designed sequentially 16 new experiments based on the simplex algorithm. The optimal point was found after 16 iterations. The maximum value of the objective function was 0.70, where the ammonium sulphate concentration was 0.8 mol/L and the pH was 7.1. The corresponding lysozyme concentration was 6.05 mg/mL. As shown in Figure 3.8, the objective value jumped from 0.14 to 0.50 after only the first set of iterations, which proves that the simplex algorithm is very effective when searching for the extreme point in the given space. When the ammonium sulphate concentration was lower than 0.8 mol/L, the solubility of lysozyme stayed almost the same so that the objective function value decreased. When the ammonium sulphate concentration was higher than 0.8 mol/L, the rapid decrease in the solubility of lysozyme made the objective function decrease. Also, the impact of pH on the solubility was less than that of the ammonium sulfate concentration. Figure 3.9 shows the response surface extrapolated from these 18 data points. It confirms that the optimum point found reached a relatively high objective function value.

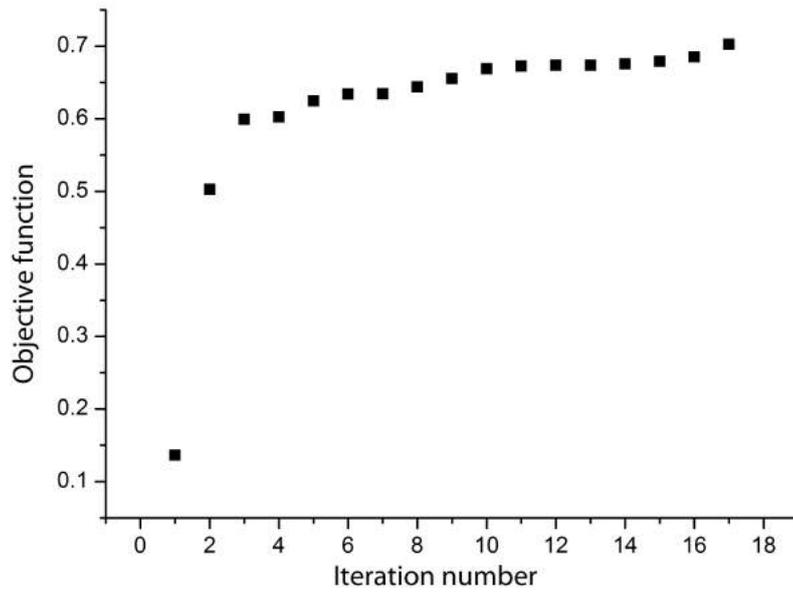


Figure 3.8 The objective function increases with the number of iterations in the simplex search algorithm

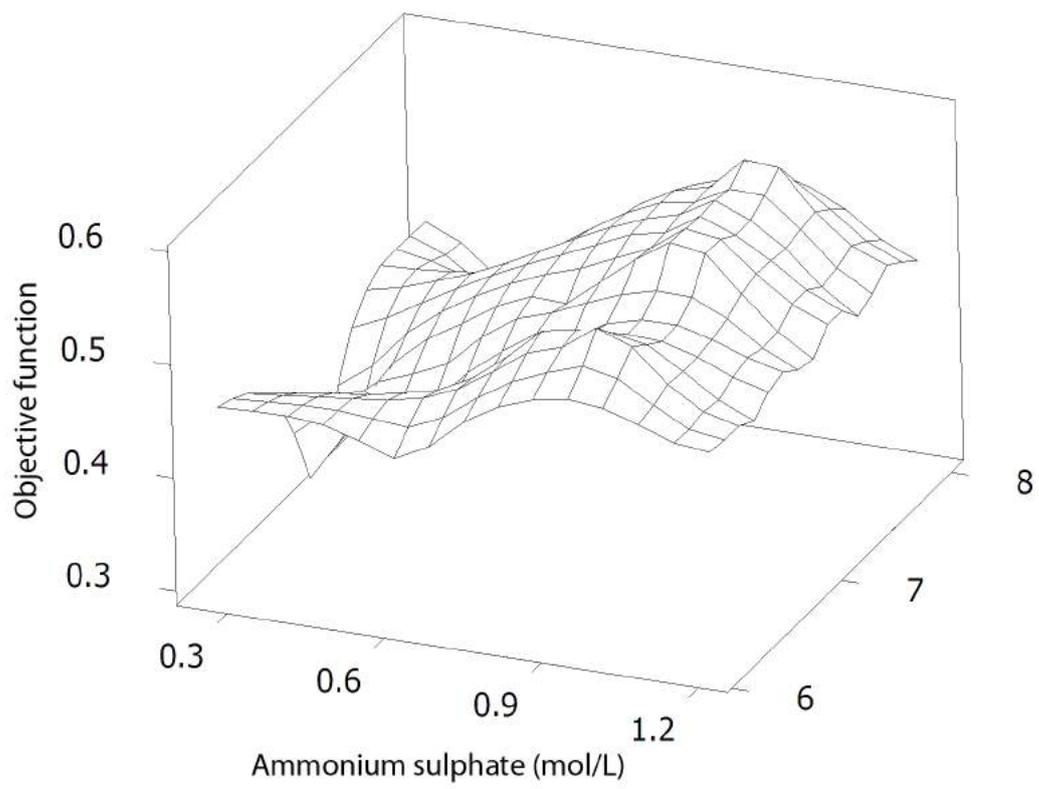


Figure 3.9 The estimated surface of the objective function based on all 18 experimental points

### 3.4.2 Precipitation optimization using the ANN algorithm

The initial five experiments were selected randomly. After four iterations, the model achieved the desired accuracy. The response surface of the predicted lysozyme concentration in the soluble phase that was achieved is shown in Figure 3.10. When the ammonium sulphate concentration was higher than 0.8 mol/L, the lysozyme concentration decreased rapidly, similar to the results from the simplex algorithm. At same time, pH had little impact on the solubility of lysozyme. According to the model generated by the network, the highest objective function value was found at 0.7 mol/L of ammonium sulphate concentration and pH of 7.0, as shown in Figure 3.11.

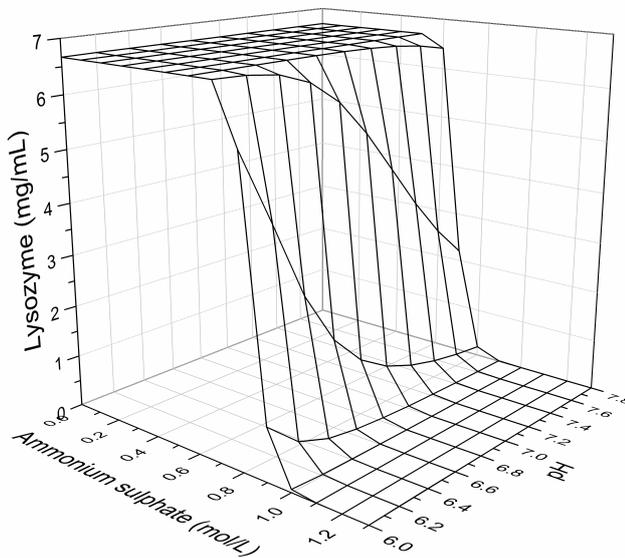


Figure 3.10 The predicted lysozyme concentration response surface at various ammonium sulfate concentrations and pH values by the ANN method

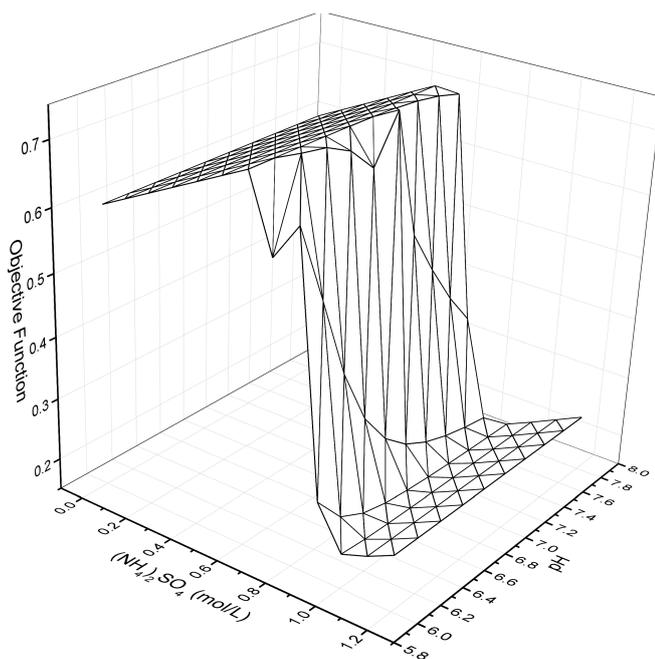


Figure 3.11 The predicted objective function surface from the ANN method.

The results from two precipitation studies demonstrated the capability of the Experiment Design Agent. Both of the sequential experiment design methods delivered the process design solution successfully. At each iteration, the communication was also very efficient since the time used for communication, data analysis, and experiment design in the above studies was less than 1 min, which significantly saved the process design time.

### 3.5 Conclusions

A first prototype of the intelligent automation platform to perform precipitation experiments automatically has been established. The platform is able to drive liquid handling robotic arm, vacuum pump, shaker and plate reader together to perform all of the procedures such as assembly adding, shaking, incubation, uv absorption measurement automatically without human involvement. The closed loop learning was achieved by using simplex search

algorithm and neural network modelling that the experiments were selected iteratively until the final design solution was found. The 5 agents worked together to select the initial experiments, drive the LHS to execute experiments and the plate reader to measure the protein concentrations, and control the simplex algorithm or the ANN algorithm to carry out data analysis then identify the next batch of experiments. The whole experimental procedure works smoothly.

The case study demonstrated two different algorithms: simplex algorithm and artificial neural network. The simplex algorithm shows its power when searching extreme points in a defined space, while the neural network modelling approach gives a better understanding of the effect of lysozyme solubility at various ion strengths and pH values. The 'intelligent' experimental design by these algorithms can keep the number of experiments low. The results obtained from the platform are reliable and same as those from the manual operations (data not shown).

The intelligent automation platform that integrates experiment devices, database, and algorithms for experiment design has been demonstrated as a powerful tool for high throughput process development.

The IAPBD has successfully achieved the automation of high throughput precipitation experimentation. The realisation of the automation has made the time spent on the data analysis and experiment design negligible. The outcome on time reduction in data analysis and experiment design has fulfilled the goal of the intelligent automation platform. The time saving in this simple experiment means the potential time saving in complex bioprocess development may be even more significant. The elimination of time delay during experiment procedure and utilization of intelligent experiment design in intelligent automation platform has reduced time needed and make the platform a very effective process optimization tool.

## **Chapter 4 IAPBD implementation in a monoclonal antibody precipitation process design**

### **4.1 Introduction**

The advance in mammalian cell culture has increased the monoclonal antibody production titre to 10 g/L or higher (Low et al. 2007). The efficiency of upstream processes presents challenges for downstream antibody purification processes to achieve both product quality and cost efficiency. In a typical downstream purification process, the impurities from mammalian cell culture includes proteins, host genomic DNA, endotoxin and other potential hazards such as allergy molecules (Wurm 2004), all of which must be removed after downstream processes for the therapeutic antibody product. The residual in the final antibody product must be controlled within an extremely low level for safety purpose according to the requirements from FDA. Currently, the downstream processes for mAb purification usually starts with protein A chromatography because the affinity chromatography with protein A has been proven to be the most effective purification step for mAb and is capable of achieving 95% purity after one single step depending on the compositions of feedstock (Shukla et al. 2007; Sommerfeld & Strube 2005; Roque et al. 2007).

The disadvantage of protein A chromatography, however, is significant during large scale high titres feedstock. Protein A resin has relatively low binding capacity, while its cost is very high. In the large scale downstream purification processes, a large amount of expensive protein A resin is required to capture MAb. Furthermore, the reduced wall support in a very large diameter column and physical properties of the beads will slow the flow rate due to the increasing pressure drop (Stickel & Fotopoulos 2001) and could reduce the robustness of the process. Thus, the volume of column has to be controlled within the limit. This leads to the parallel batch mode is widely adopted in MAb purification by increasing the number of the

columns. However, this approach demands a large initial capital investment and higher operation cost. A lot of research works have tackled the challenge by improving the binding capacity or beads rigidity (Hober et al. 2007). The manufacture cost will still be high due to a large number of processing cycles, high concentrations of column-fouling materials such as cell debris or lipids, and more frequent use of harsh cleaning agents (Tugcu et al. 2008). Therefore, a cost effective purification solution which is able to reach comparable quality results of protein A chromatography, as well as easy to fit into current large scale manufacture system, is extremely desirable (Shukla & Thömmes 2010).

The precipitation process for MAb purification presents a promising solution. The precipitation process is the technique that uses the differences of protein solubility to precipitate components to separate proteins. It has been reported to be promising in industrial production due to ease of operation and clean products in the later refining stages (Thömmes & Etzel 2007). Ammonium sulphate is extensively used in precipitation process because of both its high salting out effect and small impact for protein denature (Cheng et al. 2006).

Traditionally, in order to meet the quality requirements, polishing processes are demanded after protein A chromatography to further remove impurities, e.g. host cell protein, genomic DNA etc. (Li et al. 2005; Shukla et al. 2007). Ion exchange chromatography (IEX) is an important polishing step. It separates proteins based on protein charge differences. Therefore, ionic interactions, the type of resins, buffer condition and protein properties have significant effect on process efficiency.

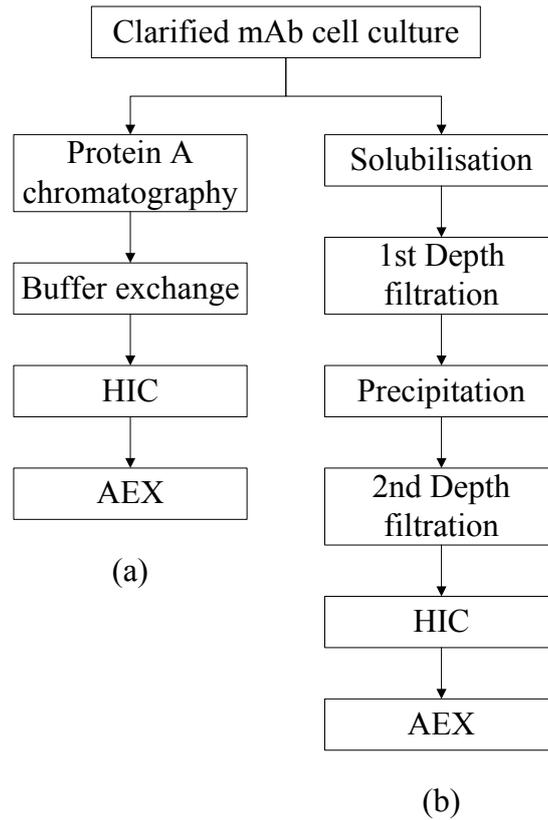


Figure 4.1 Two downstream purification processes: (a) protein A chromatography platform; (b) precipitation based platform. (Ji 2012)

Furthermore a recent work proposed a mAb purification process that includes two-step precipitation and anion exchange chromatography (AEX) (Ji 2012). The precipitation and depth filtration are used to remove impurities before AEX. The two process routes are shown in Figure 4.1. Based on Ji's model based optimization approach, the precipitation based purification platform achieved very comparable results and the operating cost has been significantly reduced based on his estimation.

A “step by step” approach is often used to optimize a sequence of unit operations. Each step is assumed to be independent from other steps. Thus, each step has been optimized individually during the process design. However, some studies showed that the interactions do exist among those unit operations (King et al. 2007; Rayat et al. 2010). For example the

interaction between the homogenisation for cell disruption to release an intracellular protein and the centrifugation for removal cell debris was reported (Siddiqi et al. 1996). As both of the precipitation and the chromatography process are targeting protein separation, some of the impurity proteins may be removed only by the precipitation, some by the chromatography and some by both. The different step one operation will lead to different feed material for step two. Hence the interaction between the two steps is likely to exist at some extends in principle.

Another reason of using the “step by step” approach may be due to the practical difficulty encountered in manually organising the two-step operations at pilot scale. With the advance in the microscale experimentation technology and the automation technology, multi-step process experimentation becomes manageable.

In the two-step precipitation and AEX process, some impurities may be removed relatively easily via AEX comparing with precipitation, while the other can be precipitated out with ease. In that case, the optimal condition may be achieved by studying related unit operations together so that each unit operation can have best performance. By examining two different optimization approaches, as shown in Figure 4.2, the interaction exist between these two steps can be investigated. Approach (a) is the optimization of each unit operation separately. Approach (b) is to use the optimal conditions in AEX obtained from Approach (a) to purify the feed material first, and then optimise the precipitation step. My hypothesis is the results from these two approaches will be the same if there is no interaction between them. .

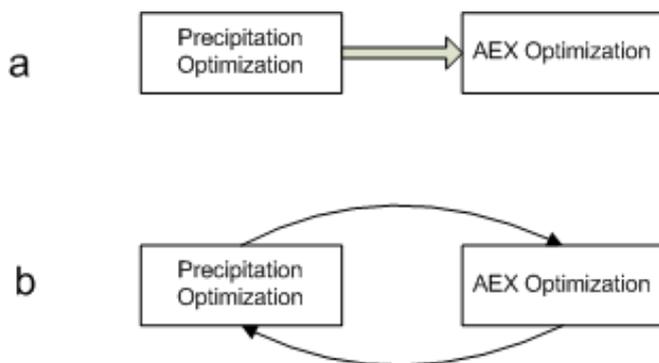


Figure 4.2 Two different approach for process optimization that involves more than one unit operations

Another purpose for the case study is to demonstrate that the platform is able to integrate with the automated devices that do not have the APIs. HPLC is the workhorse for protein quantification. It will be used to quantify mAb and impurity protein concentrations in the case study. Since the HPLC systems do not have the programmable interface, the architecture in the Assay Agent in the previous case study is not feasible anymore. Instead of controlling the HPLC directly from the agent, an online watcher program has been developed to monitor the operation status of HPLC system and to report back the raw data to Assay Agent.

In the case study, a mAb precipitation optimization will be firstly carried out to further evaluate the platform, which is integrated with a separated device (e.g. HPLC). Experimental results will be benchmarked with those from Ji's work (Ji 2012), which were done manually. The selected precipitation condition will be used to perform a two-step mAb purification process, which include precipitation and AEX. The operation condition for AEX will be chosen according to result from Ji's study (Ji 2012).

The next step is to investigate the interaction between precipitation and AEX. The feedstock will go through AEX first using the same pH and ion strength. Then, another round precipitation optimization will be performed on the platform. If our assumption is valid, higher yield and higher purity of mAb is expected.

In summary, there are two purpose of the case study as discussed before: 1) verification of the platform integrated with instruments, which do not have programmable interface (e.g. HPLC); 2) examination of the interaction between precipitation and AEX in mAb purification process.

## **4.2 Materials and Methods**

### 4.2.1 Materials

#### 4.2.1.1 Cell culture and mAb production

A humanised IgG4 type monoclonal antibody (mAb) was expressed extracellular by the CHO cyol cell line which was developed and authorised by Lonza (Slough, UK) for culturing in Biochemical Engineering Department at UCL. The pI of the mAb is 7.4 to 7.8 measured by isoelectirc focusing gel. The cell culture broth was produced by my colleague following standard procedures (Popova et al. 2015). The cell culture was harvest at mAb concentration around 1g/L. It was then processed at 10,000 rpm by Eppendorf Cenrifuge 5810R for 30min and then filtered through a 0.22 mm filter.

The clarified cell culture broth was stored in a -70 °C freezer. For each set of experiments, the same batch of cell culture was used. The concentrated material was diluted by PBS to required mAb concentration in the following studies.

#### 4.2.1.2 Chemicals

Sodium monobasic phosphate, sodium dibasic phosphate, and ammonium sulphate et al. were purchased from Sigma Chemical Co. Ltd. (Dorset, UK). All chemicals were reagent grade. Ethanol and other HPLC solution were HPLC analysis grade.

#### 4.2.1.3 Antibody standard

MAB standard were produced by Protein A chromatography. MabSelect Hitrap 5ml column (GE Healthcare, Sweden) was equilibrated in pH 7.4, 0.15 M sodium chloride, 20 mM sodium phosphate buffer with 1 ml/min flowrate on AKTA Basic (GE Healthcare, Sweden). The feedstock was buffer exchanged to pH 7.4 loading conditions in Millipore 5 kDa concentration tube (Millipore Limited, Dundee, UK) and loaded onto the column. Three column volumes of the loading buffer was followed to wash the column and one column volume of the loading buffer with 10% (v/v) isopropanol was used to wash out hydrophobic impurities. Then, mAb was eluted by 100 mM, pH 3.5 sodium citrate buffer at 1 ml/min. Then immediately add 50 µl per 1 ml elution 1 M pH 9.0 Tris buffer to neutralize and protect antibody. Loading, washing, elution and fraction collection were executed and monitored by AKTA Basic with Unicorn software (GE Healthcare, Sweden).

#### 4.2.2 Methods

##### 4.2.2.1 Micro scale precipitation

The mAb precipitation was carried out in the 96 microwell filter plate with 0.45 µm Durapore membrane (Millipore Limited, Dundee, UK) by Tecan robot (Tecan Group Ltd., Switzerland). Since the mAb formed lipid-like precipitate, it required higher mixing speed and only filtration can successfully separate solids from liquid. The precipitation conditions chosen were pH from 5.0 to 8.5 with an interval of 0.5, ammonium sulphate concentration from 0.6 mol/L to 2.2 mol/L with an interval of 0.2 mol/L. Acetate buffer was used for pH 5.0 to 5.5 and phosphate buffer for pH 6.0 to 8.5. The total volume of small-scale precipitation was 250 µl, including salt solution, 20 mM pH buffer, distilled water and mAb solution. 25 µl pH buffer was first aspirated from 200 mM stock solution and added into each microwell, followed by required volume of 4 M ammonium sulphate stock solution and water. The microwell plate was then mixed in the Eppendorf thermomixer for 15 mins to avoid unevenly distributed solution before dispensing 50 µl feedstock. The precipitation

plate was incubated at 20 °C with a microwell lid on and mixed in Eppendorf thermomixer at 1000 rpm for 2 hours. Then microwell plate was placed into vacuum separator, Te-VacS (Tecan Group Ltd., Switzerland) with a 96 microwell receive plate underneath. Clear permeate was transferred to Agilent 96 micro-well plate by robotic arm for HPLC analysis.

#### 4.2.2.2 Anion exchange chromatography

Pre-packed 1mL HiTrap Capto Q (GE Healthcare, Sweden) was used in the case study for anion exchange chromatography. AKTA basic was used to connect columns and control purification steps. The conditions of loading buffer were pH 7 by 5mM sodium citric buffer and sodium chloride with concentration of 1mS/cm. 5 column volumes of loading buffer was applied to HiTrap column first. The antibody feedstock was buffer exchanged to pH 7 loading conditions and loaded to column at 1ml/min flowrate. Fraction collection was then used to collect elute with 1 ml each tube for 1 ml column. Collected elution at main peak was kept together in +4°C fridge.

#### 4.2.2.3 Bradford total protein assay

Bradford protein assay was used to analyze the total protein concentration in the samples. BSA standard and Bradford protein assay kit were bought from Sigma-Aldrich (Dorset, UK). PBS was used to dilute samples and worked as blank. All samples were diluted to be in 0-40 µg/ml range, which was the standard and assay working range. 1 ml of standards, controls and samples were transferred to cuvettes and then 1 ml Bradford reagent was added. Each cuvette was well mixed and covered from light for 10 minutes at room temperature (Bradford, 1976). Both 2 ml cuvette and transparent 96 microwell plates, if samples were transferred from cuvettes, can be used to take samples and test at UV 595 nm. The total protein concentration was then calculated based on the corresponding calibration curve. Figure 4.3 shows the calibration curve of total protein concentration using Bradford assay.

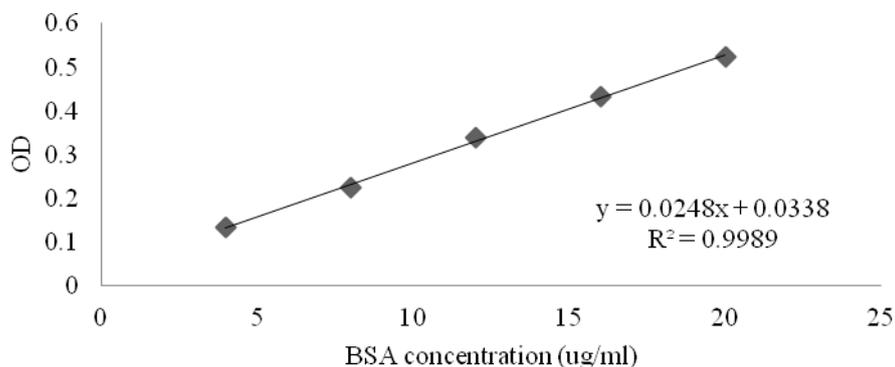


Figure 4.3 Calibration curve of total protein concentration using Bradford assay

#### 4.2.2.4 mAb HPLC analysis

The mAb concentration was measured by a 1 ml protein G HiTrap column (GE Healthcare, Sweden) connected to an Agilent 1200 series HPLC system (Agilent Technologies, UK). The equilibrium and loading buffer was pH 7.4, 20 mM sodium phosphate. 100  $\mu$ L sample was injected into column with an autosampler. The sequence and methods were pre-programmed in HPLC software Chemstation. UV 280 nm signal was recorded and used to measure the peak area. The mAb concentration was calculated based on a calibration curve, which was generated by several mAb concentration samples, range from 0 mg/ml to 1.5 mg/ml, diluted from mAb standard. Before samples being loaded into HPLC, they were transferred to 96 microwell filtration plates and centrifuged at 1,000 rpm for 10 minutes. The permeates were transferred to Agilent 96 microwell plates for HPLC. The flow rate of HPLC was kept at 2 ml/min with upper pressure 85 bar limit. Total analysis time is 6 minute with mAb peaking at around 3 minute. The calibration curve of mAb concentration by peak area using HPLC is shown in Figure 4.4.

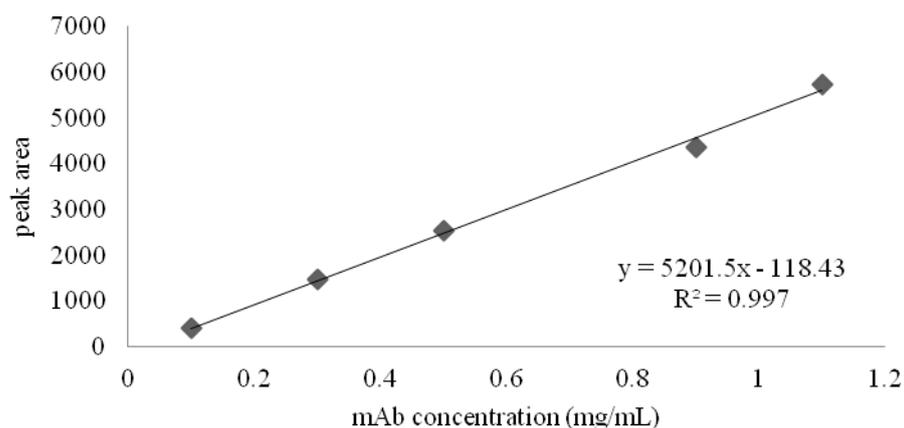


Figure 4.4 Calibration curve of mAb concentration by peak area using HPLC

#### 4.2.2.5 Impurities analysis

The overall impurities included host cell protein, cell culture media protein and other impurities. These impurities were measured by HPLC at UV 280 nm as the first flow through peak. The feedstock was used as the standard. Several samples diluted from standard were made according to different dilution rates. The Bradford total protein assay was used to measure the total protein concentration in each diluted sample. The impurity concentration for each sample was calculated by subtracting the corresponding mAb concentration from the total protein in that sample. The calibration curve was then regressed from the impurity concentration and HPLC peak area. The regression goodness of fit had R-square of 0.99 and random sample tests were validated by Bradford assay. The calibration curve of impurity concentration by peak area using HPLC is given in Figure 4.5.

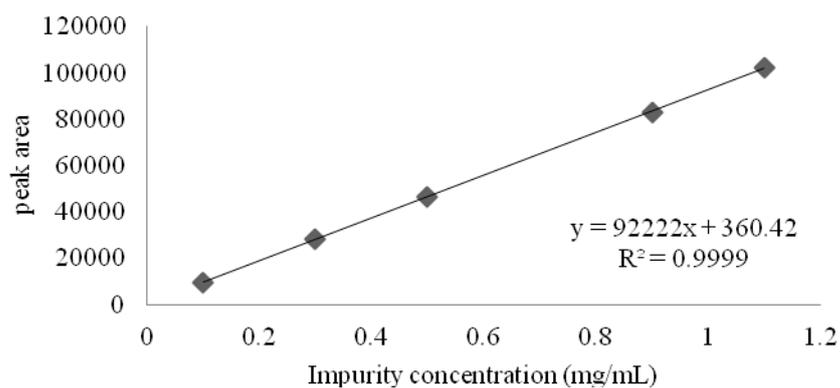


Figure 4.5 Calibration curve of impurity concentration by peak area using HPLC

### 4.3 IAPBD implementation for mAb precipitation experimentation

The mAb precipitation process has similar experimental procedures as the lysozyme precipitation process. The key difference is that the assay for mAb quantification uses HPLC instead of the plate reader. HPLC is a stand-alone instrument. Its software does not provide the API for integration. Hence the new development illustrated in this section is the Assay Agent Design. The other agents use the software developed in Chapter 3.

#### 4.3.1 Assay Agent

The Assay Agent in this case study is different from that in previous chapter. Instead of Tecan platereader, HPLC, which is controlled by Agilent ChemStation (Agilent, US), is used to measure protein concentration. Unfortunately Agilent ChemStation does not have public APIs, which allow the third party software to control HPLC. Therefore, a new approach was proposed and developed to communicate with HPLC in the case study.

There are two challenges that need to be overcome in order to achieve the process automation. The first one is to get the real time experimental information, e.g. status, assay

data etc., without direct access to ChemStation. The other one is to convert the raw data such as retention time and peak area into protein concentration given that the retention time of one protein is not exactly same on each HPLC run. A file watcher was designed and programmed in the Assay Agent. The structure of Assay Agent for HPLC communication and data analysis is shown in Figure 4.6.

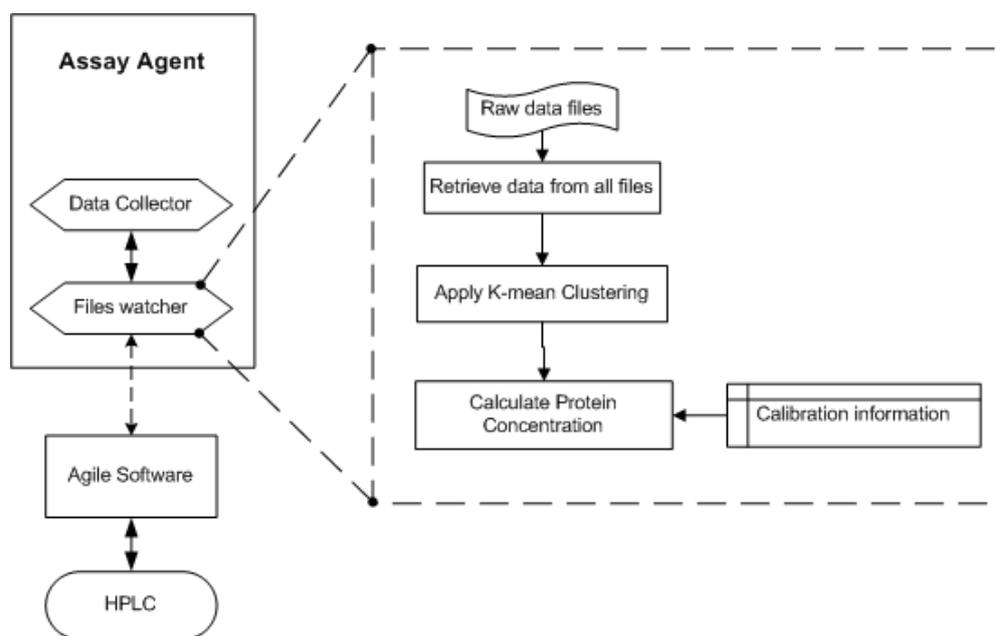


Figure 4.6 Structure of assay agent in the case study

The first challenge is solved by continuously monitoring the location where ChemStation stores output files. In this way, the agent is able to retrieve the assay data by analyzing the latest created data file in real time.

K-mean clustering algorithm is introduced to overcome the second challenge. By monitoring the data file generated by ChemStation, the Assay Agent is only able to retrieve the raw data, which are pairs of retention time and peak area. Although the retention times for a certain type of protein are supposed to appear at same time at each HPLC run, they are

not necessary to be exactly same. Besides, noises on the HPLC spectra sometimes are recognized as peaks by ChemStation. Thus, an algorithm is needed to 1) translate the raw data into protein concentrations; 2) eliminate the noises in raw data. K-mean clustering algorithm embedded in Assay Agent is chosen to gather the similar peaks together based on retention times. K-mean algorithm aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. As shown in Figure 4.7, 280 nm absorption peaks in HPLC forms two clusters around 0.3min and 3 min, which stand for impurities and mAb respectively. Therefore, a cluster algorithm is required so that the agent can calculate the accurate protein concentration by correctly identifying the mAb's peaks from impurity's ones during HPLC assay.

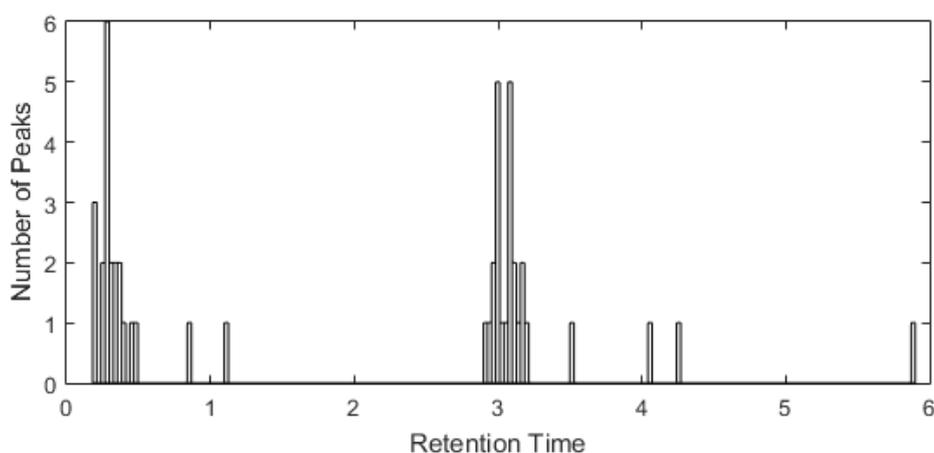


Figure 4.7 HPLC retention time distribution among 21 experiment runs

The k-mean algorithm is to firstly define k centres, and then take each point belonging to a given data set to associate with the nearest centre until no data point is pending. By iteratively updating the centre points from new formed cluster, the algorithm aims to minimize squared error function given by:

$$J(V) = \sum_{i=1}^C \sum_{j=1}^{C_i} (\|x_i - v_j\|)^2$$

where,

$\|x_i - v_j\|$  is the Euclidean distance between  $x_i$  and center point  $v_j$  of cluster  $j$ .

$C_i$  is the number of data points in the  $i$ th cluster.

$C$  is the number of cluster centers.

The detailed steps for the algorithm in our platform are:

- 1) Select 2 cluster centers according to predefined retention time stored in the database
- 2) Calculate the distance between each data point (peak) and cluster centers.
- 3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
- 4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where,  $c_i$  is the number of data points in  $i$ th cluster.

- 5) Recalculate the distance between each data point and new obtained cluster centers.
- 6) If no data point was reassigned then stop, otherwise repeat from step 3).

Fail safe and quality assurance mechanism is included in the agent. After each peak in the dataset has been labelled by clustering algorithm, the agent will convert peak areas into protein concentrations according to calibration information stored in the database. In the case that one or more protein concentration is missing (i.e. peaks are not recognized by the algorithm) for an experiment, the agent will be paused, and experimental data will be examined manually. In addition, the outlier detection algorithm as explained in Chapter 2 is implemented to insure the data quality of experiment results.

From initial test runs, the algorithm was proven to be fast and robust. The mAb and impurity concentrations calculated by the Assay Agent automatically were same as the ones that were done manually.

#### 4.3.2 Execution Agent

Same as previous case study, the execution agent controls the robotic arm (Tecan Freedom 200) to add the feedstock, phosphate buffer and ammonium sulphate solution into the microwell plate. The different pH values of each experiment ranged from 6 to 8.5 were achieved by mixing suitably calculated volumes of monosodium phosphate and disodium phosphate solution for phosphate buffer. And the pH from 5.0 to 5.5 was achieved by mixing acetate buffer and sodium acetate solution. Then control it to put the microwell plate into Eppendorf thermomixer and control the shaking speed and time. After incubation and shaking, it controls a vacuum pump on the platform to separate the precipitates and liquid. Finally control the robotic arm to take the sample of permeation to the plate reader to detect mAb concentration.

#### 4.3.3 Experiment Design Agent

The Experiment Design Agent is same as one from previous case study. This case study aimed to find out the optimal condition for mAb precipitation other than to explore the response surface for the objective function. Hence, instead of model based approaches (e.g. ANN), simplex search algorithm is chosen in this chapter. The objective function is defined by:

$$J = w_1 * \frac{c_0 - c}{c_0} + w_2 * \frac{c_0 - c}{T - c - c_t}$$

Where,

$w_1$  and  $w_2$  are weighting factors for yield and purity. In this case study,  $w_1$  equals 0.4, and  $w_2$  equals 0.6.

$c_0$  is initial mAb concentration in feedstock.

$c$  is mAb concentration in supernatant after precipitation.

$c_t$  is impurity concentration in supernatant after precipitation.

T is total protein concentration in feedstock.

Thus,  $\frac{c_0 - c}{c_0}$  is mAb yield, while  $\frac{c_0 - c}{T - c - c_t}$  is mAb purity.

#### 4.3.4 Coordinate Agent

The Coordinate Agent retrieves the historical experiment data from database and writes the new executed experiment results into the same database. It also translates the information in the database in the message format to blackboard.

#### 4.4 The mAb precipitation process development

The execution agent drove the TECAN LHS and HPLC automatically. The Assay Agent gathered real time data from HPLC and translated the raw data into mAb and impurity concentration successfully. The communication between the coordinator agent and other agents was smooth in facilitating data exchange. The Experiment Design Agent used the simplex search algorithm in both experiments. Two mAb precipitation studies were carried

out.

#### 4.4.1 1<sup>st</sup> experiment: mAb precipitation optimization on initial feedstock

The aim of the mAb precipitation optimization is to verify the IAPBD from the operation and the precipitation experiment results.

Three initial points to form the initial simplex was selected randomly. The initial simplex shown in Fig. 4.8 (A) is quite away from the optimal point. One point was found at the iteration 2 located on the [1,4, 8] (1.4 mol/L of Ammonium Sulphate and pH value of 8) via *reflection*, and then another point at (2.2, 8) via *expansion*. Following the procedures explained in Chapter 3, the Experiment Design Agent found the optimal point after 5 iterations as shown in Figure 4.8.

The maximum value of the objective function was 0.81, where the ammonium sulphate concentration was 2.2 mol/L and the pH was 8.

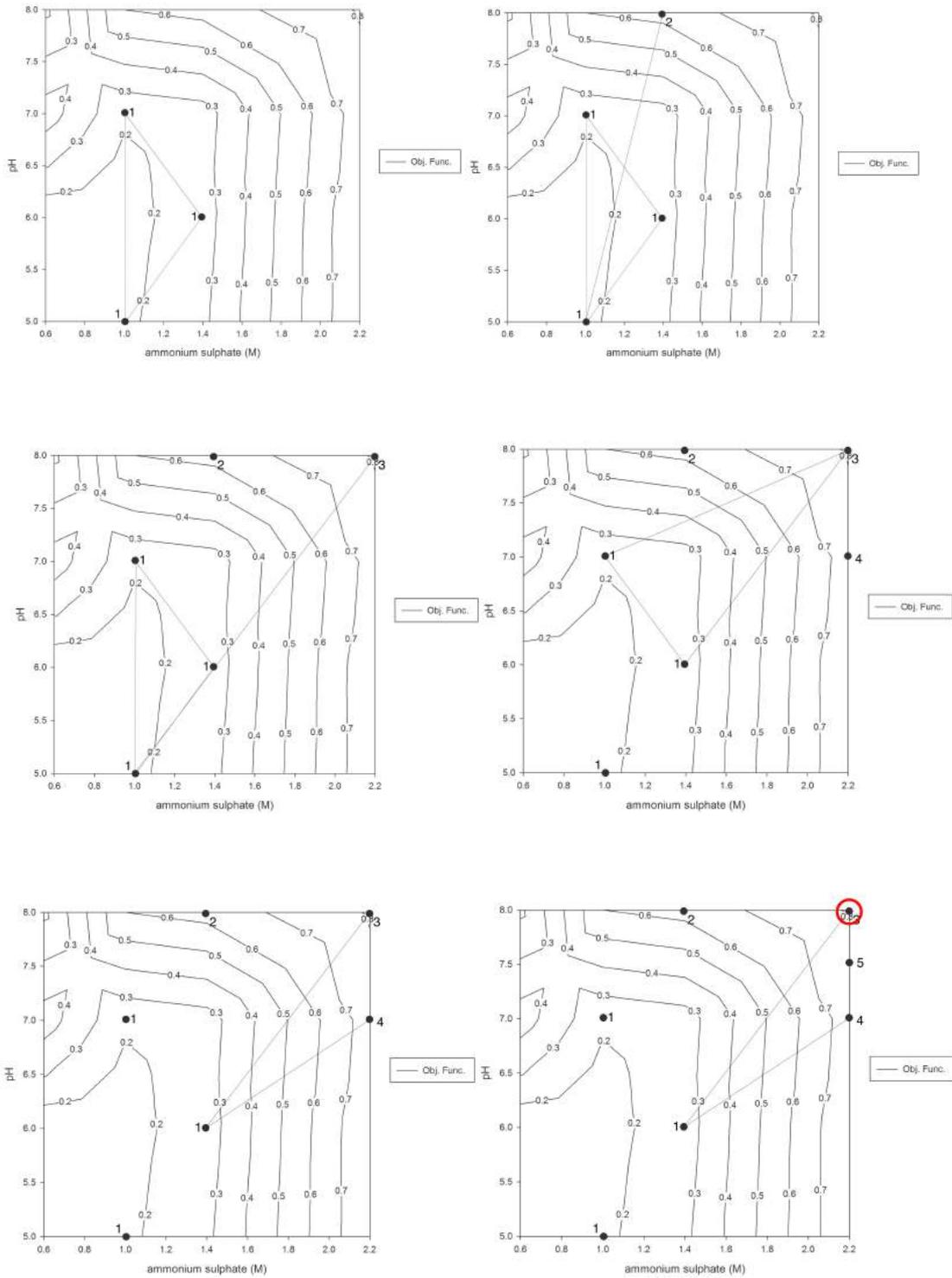


Figure 4.8 Illustration of simplex search algorithm on the objective function's surface. The designed experiment points are shown as the black dots. The number near the black dot is the iteration number. The final optimal point is shown in small red circle.

The corresponding mAb concentration in the permeate was 0.001 mg/mL, whereas impurity concentration was 0.813 mg/mL.

As shown in Figure 4.9, when the ammonium sulphate concentration was lower than 1.3 mol/L and pH was lower than 6.5, little precipitation occurred and the solubility of both mAb and impurity stayed almost the same so that the objective function value stayed almost the same. When the ammonium sulphate concentration was higher than 1.3 mol/L, the solubility of mAb decreased rapidly while the impurity concentration stayed almost at the same level. So the objective function increased when ammonium sulphate concentration was higher than 1.3 mol/L. The impact of pH on the solubility was less than that of the ammonium sulphate concentration. The solubility of both mAb and impurity decreased a little when pH was larger than 7.

As shown in Figure 4.8, the objective value jumped from 0.09 to 0.80 after only the first set of iterations, which again proves that the simplex algorithm is very effective when searching for the extreme point in the given space. As shown in Fig. 4.10, the objective function surface has strong nonlinearity. It will require a large set of data to characterize such process. This case study has demonstrated that the simplex algorithm has the potential to reduce the experiment load without compromise the quality of process design.

The raw data was benchmarked with the results in Ji's research (Ji 2012). The mAb concentrations under same condition of ammonium sulphate concentration and pH value in supernatant is comparable with his data.

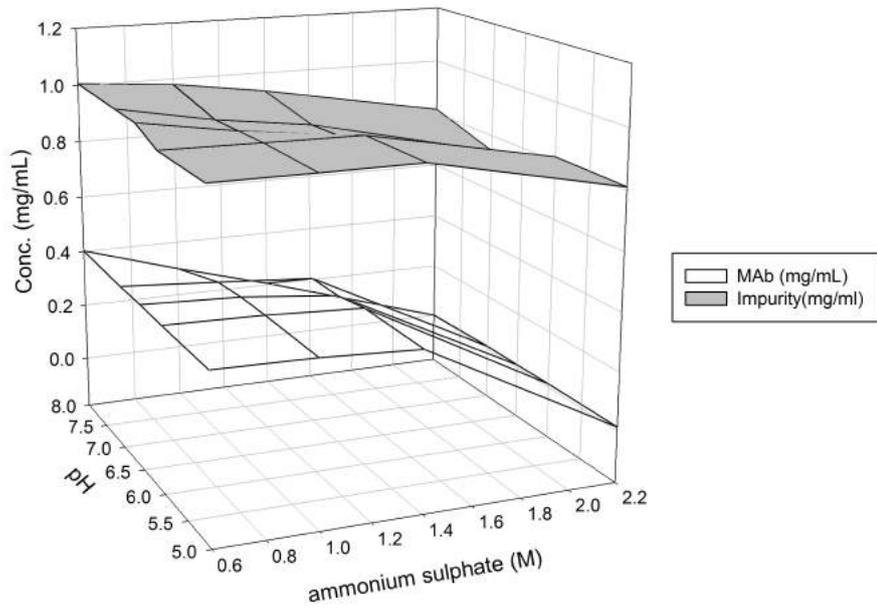


Figure 4.9 mAb and impurity concentration surface under different pH and ammonium sulphate concentration in 1<sup>st</sup> run.

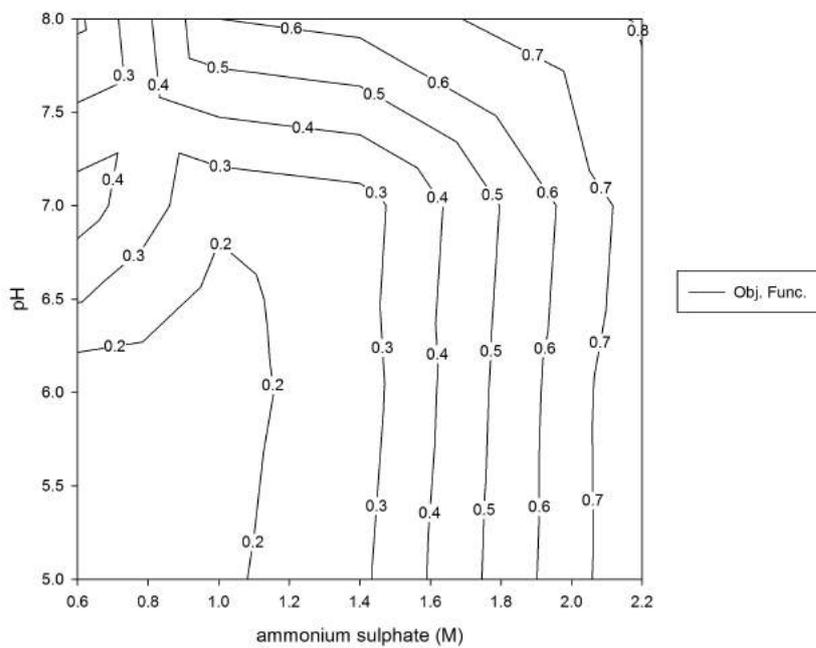


Figure 4.10 Objective function surface under different pH and ammonium sulphate concentration in 1<sup>st</sup> run.

#### 4.3.2 2<sup>nd</sup> experiment: mAb precipitation optimization on elute from AEX

The feed material was collected from the elution of AEX process at the conditions as explained in Section 4.2.2.2. The same procedures applied in the 2<sup>nd</sup> experiment run on the automated platform. Starting with 3 random selected pH and ammonium sulphate concentration, the optimal objective function was found after 6 iterations. The maximum objective function was 0.87 where pH value was 5 and ammonium sulphate concentration was 2.2 mol/L.

Because of the AEX purification, the remaining impurity concentration was about half of the initial concentration in feedstock. Besides the difference in impurity concentration, the surface of both mAb and impurity were almost identical to that in 1<sup>st</sup> experiment run as expected. Both mAb and impurity were more sensitive with ammonium sulphate concentration. The mAb concentration dropped rapidly when ammonium sulphate concentration was larger than 1.3 mol/L. The impact of pH was smaller than that of salt concentration. The solubility of both mAb and impurity decreased when pH value was more than 7.

However, there were some differences between the 1st run and the 2nd run. The impurity was less sensitive to the salt concentration when the salt concentration was less than 1.4 mol/L. As seen in Figure 4.12, the objective function surface was smoother when the salt concentration is less than 1.4 mol/L. Furthermore, the impact of pH was even less in the 2<sup>nd</sup> run comparing to the 1<sup>st</sup> run. Thus, the maximum objective function was found at pH value of 5 instead of 8 in the 2<sup>nd</sup> run.

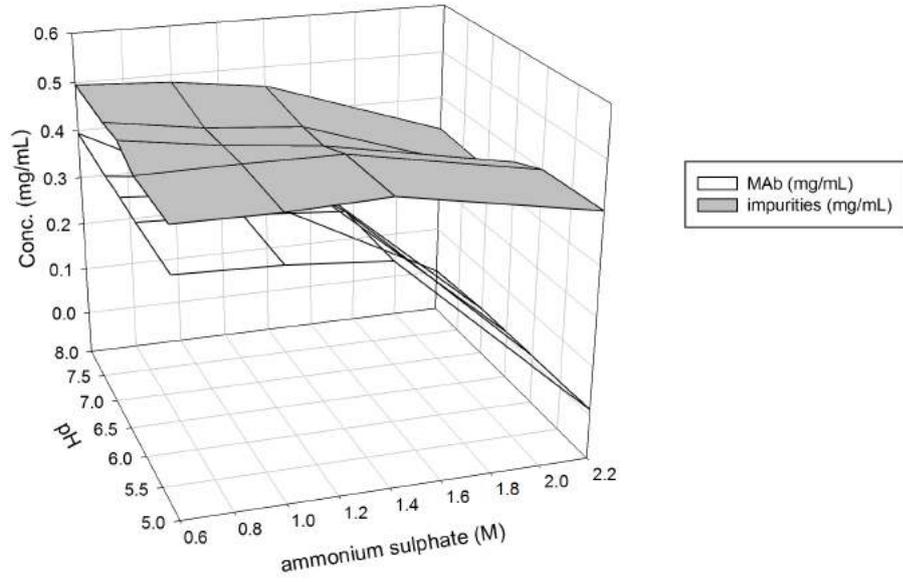


Figure 4.11 mAb and impurity concentration surface under different pH and ammonium sulphate concentration in 2<sup>nd</sup> run

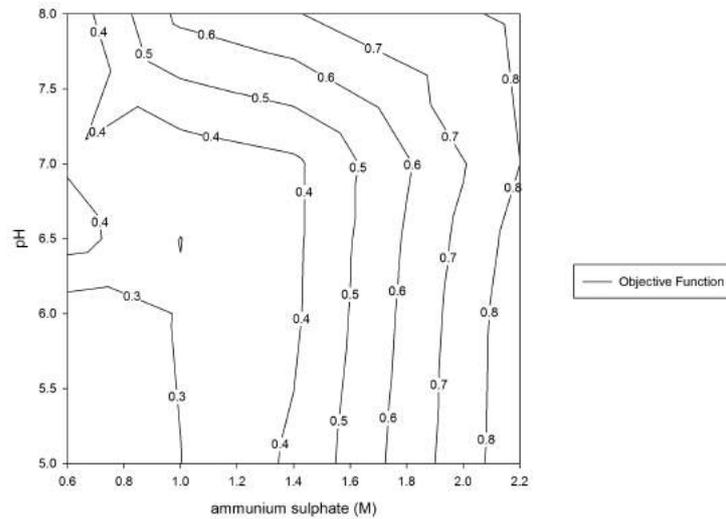


Figure 4.12 objective function surfaces with different pH and ammonium sulphate concentration in 2<sup>nd</sup> run

### 4.3.3. Comparison study of a two-step mAb purification process

The optimal precipitation conditions as shown in Table 4.1 obtained from two precipitation experiment runs were used in a two-step purification process to further evaluate final purity and yield of mAb.

Table 4.1 Conditions selected for two-step purification process

	<b>Conditions</b>	<b>1st run</b>	<b>2nd run</b>
Precipitation	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (mol/L)	2.2	2.2
	pH	8	5
AEX	Conductivity (ms/cm)	1	
	pH	7	

Two experiments were carried out manually at which scale using different conditions defined in Table 4.1, and results are shown in Figure 4.13. The purity in the According to the results, the purity in the 2<sup>nd</sup> run has increased 6.3% while the yield has decreased only 2.6%. The objective function value from the 2<sup>nd</sup> run is also slightly higher that that from the 1<sup>st</sup> run.

Therefore, our assumption that there are interactions among unit operations is valid. There are a group of impurity proteins in the feedstock that can be removed by either the precipitation or the AEX column. Hence the precipitation conditions will impact on the removal of this group of proteins at different extends. The performance of bioprocess can be further improved by considering more than one unit operation during optimization process. The integrated optimization approach includes one shown in Figure 4.2. The output of latter optimized operations can be an input for the previous operation for optimization. The bioprocess thus can be optimized by several iterations illustrated in Figure 4.2. However, this approach may be time demanding. The better solution is that all operations are processed in one step if the combined experiment procedure is feasible at microscale. In this case, all four factors can be regarded as input parameters for IAPBD. Because of the high throughput technology, the capacity of the platform can then be better utilized to achieve

higher efficiency of bioprocess development.

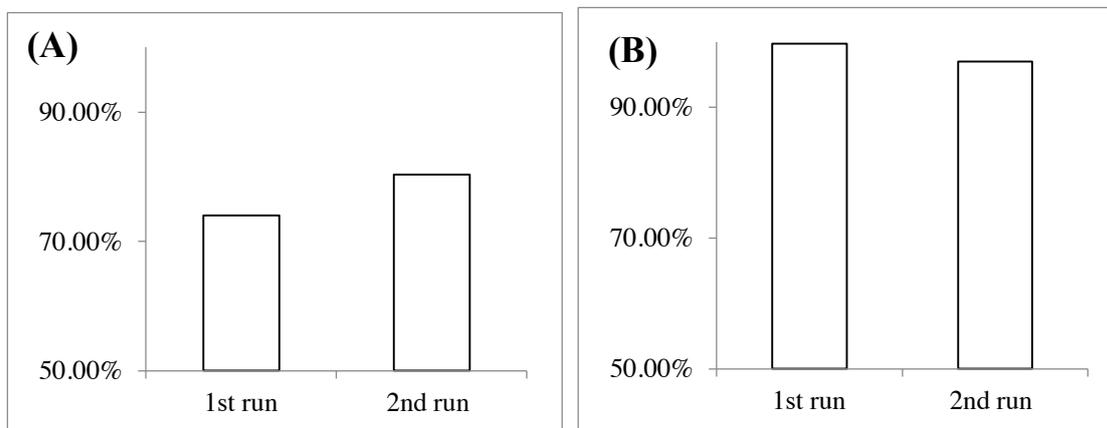


Figure 4.13 (A) mAb purities from the conditions selected by 1<sup>st</sup> run and 2<sup>nd</sup> run. (B) mAb yield from the conditions selected by 1<sup>st</sup> run and 2<sup>nd</sup> run

#### 4.5 Conclusions

The closed loop learning was achieved by using simplex search algorithm that the experiments were selected iteratively until the final design solution was found. The platform selects initial experiments from the database, which records the previous experimental data, execute experiments and measurement by automated devices, and carry out data analysis by design algorithm to identify the next batch of experiments.

The capacity of IAPBD has been increased by integrating the instruments, which have no open APIs. Although the platform cannot drive the HPLC directly, it is now capable of monitoring and analysing raw data from HPLC in real time. K-mean algorithm has been applied to translate raw data of retention time and peak area into protein concentrations. The results obtained from the 1<sup>st</sup> set of precipitation experiments demonstrate that the platform can provide reliable results same as those from manual operations.

The comparison in the two-step mAb purification study showed the interaction between the precipitation and the AEX chromatography exists. The two-step process optimization may

be realised by the IAPBD using the simplex search algorithm.

## **Chapter 5 IAPBD implementation for fermentation process optimization**

### **5.1 Introduction**

Fermentation is the core technology in biochemical engineering and bioprocessing. Developing automated and efficient methods for fermentation process development has significant importance for biopharmaceutical industry. Microscale fermentation has been studied substantially recently and a range of micro scale devices is commercially available. High-throughput approaches for fermentation process optimization have been demonstrated as a feasible and efficient method (Halim et al. 2013). They offer significant time and resource savings while allowing exploration of a wide range of process variables (Fernandes et al. 2006). In the form of micro deep-well plate, shaken microscale bioreactors are popular because they are inexpensive, widely available and easy to be facilitated on conventional LHS (Duetz 2007). The automated approach has been reported to significantly decrease the experimental material up to 1000-fold and to increase the throughput up to 100-fold in fermentation studies (Micheletti & Lye 2006; Rios-Solis et al. 2013; Du et al. 2014).

The high throughput experimentation makes experiment design and data analysis a potential bottleneck for fermentation studies. Currently, brute force method, which is to explore every possibility within the design space, is used as experiment design method in many cases. This will require a large number of experiments to be investigated. Although the implementation of microscale technologies can dramatically decrease the experimental material, it is still desirable to apply advanced experiment design and analysis methods during fermentation studies in order to reduce the time consumption of both experiment execution and data analysis. There are a few fermentation studies reported using sophisticated experiment design methods, e.g. DoE and ANN (Medeiros et al. 2000; Weuster-Botz 2000; Kennedy et al. 1992). Those methods were used offline and separated from the automated instruments. However, none of those applications has linked the method with the automated instruments. However, IAPBD provides a promising approach that can integrate those methods with

automated instruments like LHS to perform online experiment design and data analysis.

In this case study, the benefits of IAPBD are exemplified for optimization of the production of a whole cell monoamine oxidase (MAO) biocatalyst. An engineered form of the monoamine oxidase from *Aspergillus niger* (MAO-N-D5) expressed in *E. coli* is used as the biocatalyst to convert amine 3-azabicyclo octane to imine and the reaction is shown in Figure 5.1. The product imine (2) is commercially interesting as it can be used as a building block in the synthesis of the Hepatitis C viral serine protease inhibitor compounds telaprevir and bocepravir (Znabet et al. 2010; Turner 2011). The secondary amine 3-azabicycloctane (3-AZA) (1) has been identified as an appropriate substrate for imine production. MAO-N-D5 expression and specific activity will be vital to achieve a high reaction turnover and a minimum amount of side products formed (Köhler et al. 2010). The case study will only focus on final biomass concentration. Many factors will influence cell growth in fermentation process such as medium composition and operating parameters. Such a process optimization will require substantial amount of effort in general and the statistical DoE method is often used.

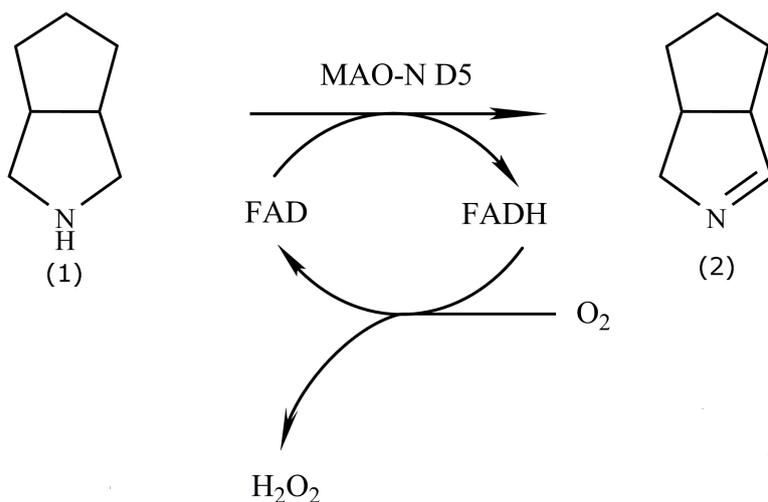


Figure 5.1 MAO-N-D5 mediated oxidation of secondary amine 3-azabicyclo octane (1) to the corresponding imine (2).

Another purpose of the case study is to examine Execution Agent's communication mechanism with some instruments that cannot be controlled directly. In many cases, the software provided by LHS supplier does not have public APIs for third party software integration. For example, Tecan Gemini was used to control an older Tecan LHS model in this case study, and Gemini does not have a programmable interface. Thus, Execution Agent of IAPBD has been modified in order to inject important data (e.g. volumes of a specific reagent needed for each micro well) into Gemini during operation. Similar to the situation in the case study in Chapter 4, the plate reader used to measure biomass concentration cannot be controlled directly by the Assay Agent either. So a file watcher algorithm will be used for the Assay Agent to communicate with the plate reader.

In this chapter IAPBD is able to utilize the sequential DoE method to achieve high biomass growth at micro scale, followed by the large scale comparison study. Furthermore, communication mechanism applied in Execution Agent will be examined as well.

## 5.2 Materials and Methods

### 5.2.1 Reagent and cells

Molecular biology enzymes were obtained from New England BioLabs (NEB, Hitchin, UK). Competent *E. coli* BL21-Gold (DE3) cells were obtained from Stratagene (Amsterdam, NL). All other reagents were obtained from Sigma-Aldrich (Gillingham, UK) unless noted otherwise, and were of the highest purity available.

### 5.2.2 Shake flask MAO Fermentation

Competent *E. coli* BL21-Gold (DE3) cells were transformed with the plasmid pET16b MAO-N-D5 using the heat shock technique described by the supplier (Stratagene, Amsterdam, NL). An overnight culture of the transformed cells was obtained in a 100 ml shake flask (10 ml working volume) of LB broth (10 g l<sup>-1</sup> tryptone, 5 g l<sup>-1</sup> yeast extract, 10 g l<sup>-1</sup> NaCl) containing 150 mg ml<sup>-1</sup> ampicillin. Growth was performed at 30 °C with orbital shaking at 250 rpm using a SI 50 orbital shaker (Stuart Scientific, Redhill, UK). The total volume of this culture was used to inoculate a 1 litre shake flask (100 ml working volume) which was left to grow for 8 hours.

### 5.2.3 Microscale automated fermentation

A Tecan Genesis platform was used for process automation and setup with all required polypropylene 96-DSW culture plates, 96-Standard round well plates, reagent troughs and disposable microconductive tips (Tecan, Reading, UK). A sterile environment was maintained by using UV sterilisation within a Bigneat class 2 biosafety cabinet equipped with a laminar airflow system and HEPA filtration (Bigneat Ltd, Hampshire, UK). The UV sterilisation step was found to be appropriate as LB-glycerol agar plates left open for a seven-hour period after a 20 minute sterilisation process showed no signs of contamination

(Baboo et al. 2012). Figure 5.2 (a) shows a real picture of the actual Tecan microscale platform with all the components. Figure 5.2 (b) shows a schematic representation of these in the Tecan Gemini software used for the programming of operations. The automated platform comprised 3 microscale shakers and incubators which were used as fermentation units using 96 DSW where temperature and shaking speed could be varied. The 96-DSW microwell plate format was adopted in this work for the fermentation steps as this geometry was found to ensure the best conditions for oxygen transfer (Duetz 2007; Baboo et al. 2012).

Microscale fermentations were run in parallel and in triplicate with a fermentation volume ranging from 0.5 to 1.5 ml. After the media preparation step, the wells were inoculated using 10 % v/v of an overnight shake flask culture as described above. Shaking speeds ranged from 400 - 1000 rpm. In order to obtain fermentation kinetic data 10  $\mu$ l samples were withdrawn from each well at selected time intervals and diluted in 90  $\mu$ l of 100 mM phosphate buffer using a 96-SRW microwell plate prior to  $OD_{600}$  analysis in the microwell plate spectrophotometer. A calibration factor was used to convert all microwell plate readings to standard spectrophotometer readings (Baboo et al. 2012).

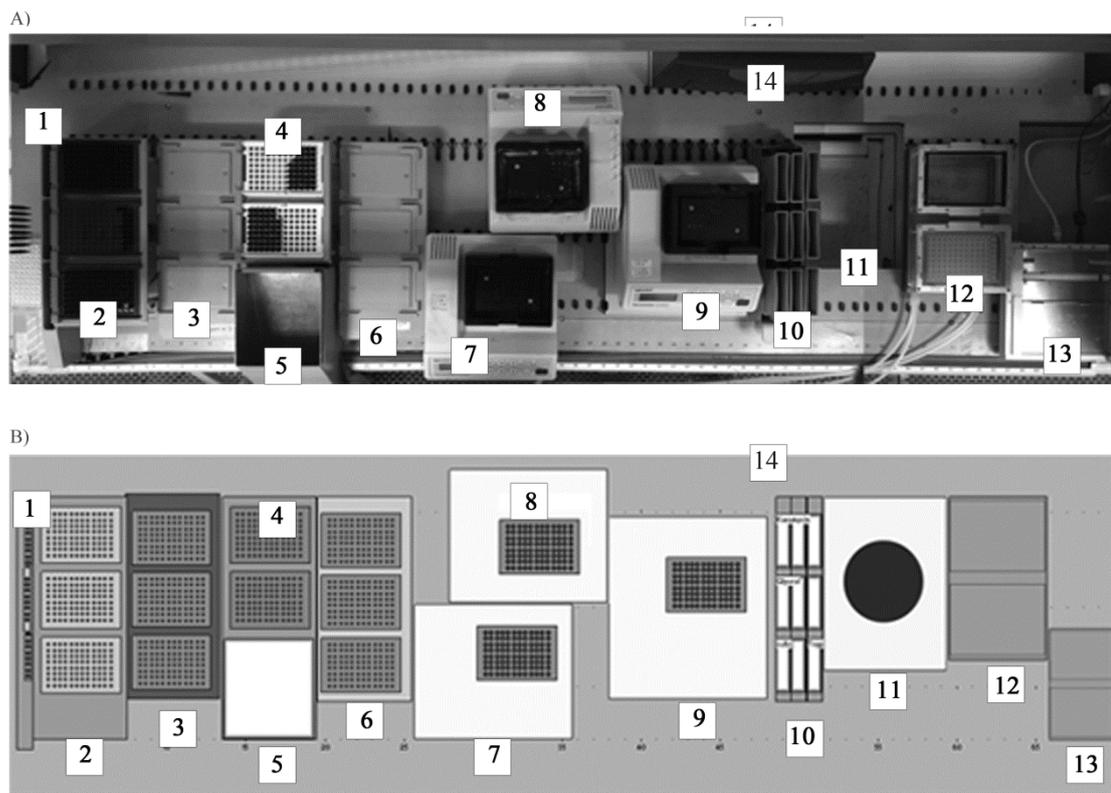


Figure 5.2 Detail of the microwell automated system. (a) Photograph of the Tecan robot deck and ancillary equipment. (b) Deck layout as represented in the Tecan EvoWare software: 1) tip washing station, 2) 1000 and 200  $\mu\text{l}$  sterile tip racks, 3) microplate racks for biomass sampling, 4) 20  $\mu\text{l}$  sterile tip racks, 5) waste disposal unit, 6) microplate rack for colorimetric assay preparation, 7-9) microwell fermentation stations, 10) fermentation media containers, 11) microplate centrifuge, 12) microplate microfiltration and low pressure unit, 13) HPLC automated rack container, 14) microplate spectrophotometer.

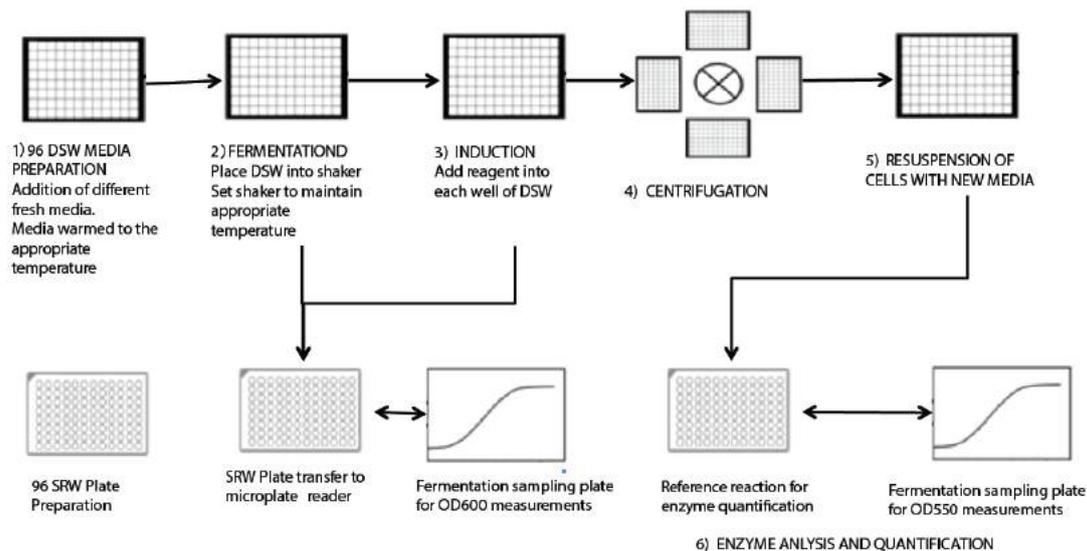


Figure 5.3 Schematic representation of the automated platform which performs microscale fermentation and biomass concentration measurement

### 5.3 MAO-N-D5 expressing *E coli* fermentation optimization

A sequential DoE on 12 factors, which include both operating factors and medium chemicals (As shown in Table 5.1), has been used in the case study for fermentation optimization. The *induction time* is included as an input factor because the experiment results were also used for a case study for enzyme concentration optimization.

Table 5.1 The ranges of operating factors and medium chemicals

<b>Factors</b>		<b>High</b>	<b>Low</b>
Peptone	% w/v	3	0.5
Yeast	% w/v	2	0.5
Glycerol	% w/v	1.5	0.2
MgSO <sub>4</sub>	% w/v	0.2	0.01
NH <sub>4</sub> Cl	% w/v	0.2	0.01
Na <sub>2</sub> HPO <sub>4</sub>	% w/v	0.2	0.01
K <sub>2</sub> PO <sub>4</sub>	% w/v	1	0.1
IPTG	mM	1	0.1
Shaking	rpm	1000	400
Fill Volume	mL	1.5	0.5
Induction Time	hr	8	2
Temperature	°C	44	37

In this case study, the sequential DoE approach was chosen for fermentation process optimization. DoE methods, especially factorial designs and mixture design, have been proven to be effective to explore the relations among several variables in fermentation process optimization (Medeiros et al. 2000; Kennedy et al. 1992; Weuster-Botz 2000). A full factorial experiment is the approach whose experimental units take on all possible combinations of high and low levels across all factors. It unveils not only the effect of each factor on the response variable, but also the effects of interactions among these factors on response variable. However, the number of required experiments grows exponentially with the number of factors. Given the number of factors  $k$ ,  $2^k$  experiments need to be carried out, which usually is too many to be logistically feasible. For example, a two-level full factorial design of three factors is to examine all possible combination among high and low levels

of all factors as shown in Figure 5.4(A). Thus eight experiments ( $2^3$ ) will be investigated in total.

By sacrificing certain degree of regression model resolution, fractional design becomes more attractive in most cases. Sometimes the fractional factorial design with relatively low resolution is more preferable at early stage for the purpose of factors screening. A fractional factorial design consists of a carefully chosen subset (fraction) of the experiments of a full factorial design. For example, only four experiments are chosen instead of eight in full factorial design as shown in Figure 5.4(B). But some three-way interactions terms are confounded with two-way interactions. It has been observed that the system is usually dominated by main effects and low-order interactions in most cases (Wu & Hamada 2011). Thus it is most likely that single factor effects and two-factor interactions are the most significant response in factorial experiment. So fractional factorial design reduces experimental materials without scarifying model's predictability.

Fractional designs are expressed by resolution. Resolution is a term, which describes the degree to which estimated main effects are confounded with higher-level interactions. In general, the resolution of a design is one more than the smallest order interaction that some main effect is aliased with. For example, a resolution-3 design means that main effects are confounded with some 2-level interactions. Low-resolution design is often chosen as factor screening. For example, in Figure 5.4(B), it is a resolution-4 design because some three-way interactions terms are confound with main effects.

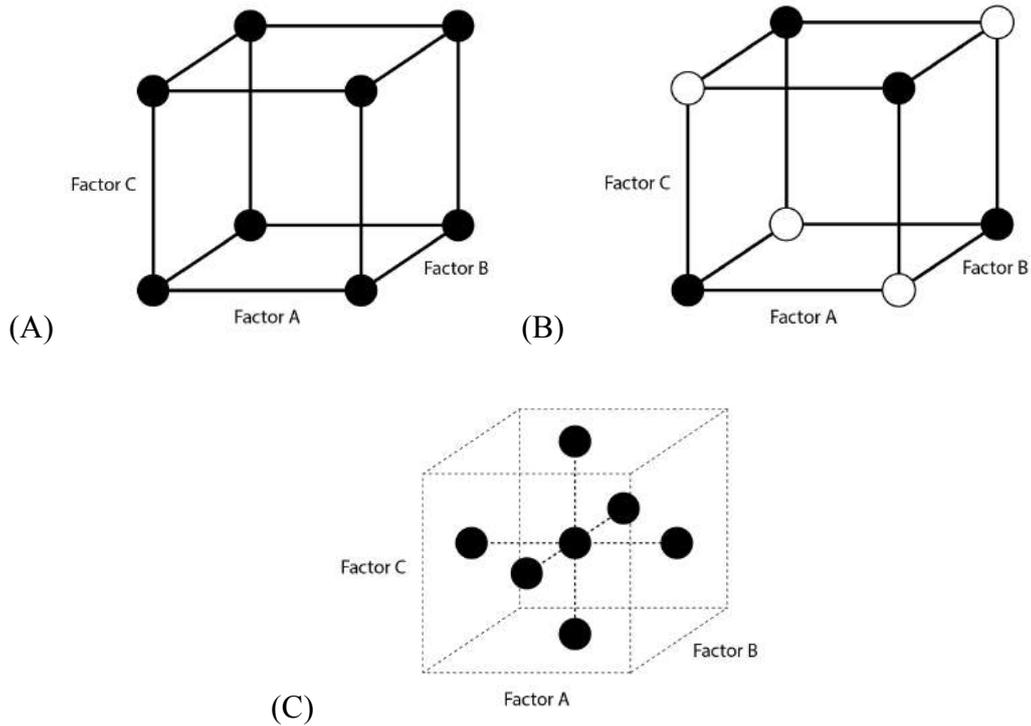


Figure 5.4 (A) Two-level full factorial design of three factors in which all possible combinations among high and low levels of three factors are required to be examined. (B) Two-level fractional design in which some combinations are ignored to reduce the number of experiment runs and some three-way interactions terms are confounded with two-way interactions. (C) Box-benhnken response surface design for three factors in which the convexity of response surface is studied by introducing central points in experiment runs.

While factorial design leads to a model with only one-degree polynomial model, response surface design introduces higher-degree polynomial terms in the model to further estimate the convexity of the response surface. Besides experiments of combination of high/low levels of each factor that required by 2-level factorial design, response surface design includes some central (medium) points. For example, in Figure 5.4(C), additional seven experiments will be examined apart from the other experiments located at vertices in factorial design.

Analysis of variance (ANOVA) is usually used to investigate and model the relationship

between a response variable and one or more explanatory variables (factors). The technique essentially consists of partitioning variation due to the factors and error. Each factor consists of two or more levels. ANOVA tests the null hypothesis that the population means of each level are equal, versus the alternative hypothesis that at least one of level means are not all equal.

DoE can be executed in sequential manner to reduce the number of experiment runs. It is not usually the case that all the factors investigated have significant effect on response. Thus, a factor screening design to eliminate the irrelevant or insignificant factors will often result in reduction of experiment runs in higher resolution factorial design.

## **5.4 IAPBD implementation for fermentation optimization**

### 5.4.1 Coordinate Agent

Same as in previous cases, the Coordinate Agent retrieves the historical experiment data from database and writes the new executed experiment results into the same database. It also translates the information in the database in the message format to blackboard.

### 5.4.2 Assay Agent

Similar to the Assay Agent in Chapter 4, the agent retrieves data via a watcher program by monitoring specific file folder. Tecan Gemini is used in the case study to control both LHS and assay devices. Unlike Tecan Evoware that is used in first case study (Chapter 3), the software does not have API for third party integration, so the agent can only access experimental data by monitoring output files from Tecan Gemini.

### 5.4.3 Execution Agent

As shown in Figure 5.5, the Execution Agent controls the devices by monitoring and sending data to Tecan Gemini in real time. Although Gemini does not provide API for direct third party integration, it provides option to open a third party executable file and take return values into its current running script. Thus, a communicator is introduced to send relevant parameters, which is volume of each media solution that required to be dispensed into each well, to the Gemini script when needed. Different from the case in Chapter 3, the script for Tecan LHS cannot be generated automatically and executed by LHS in real time. Thus, researcher still has to compose scripts manually to define every procedure that is needed during experimentation. However, because values of some parameters can be updated in real time from the communicator, no human intervention is needed during operation.

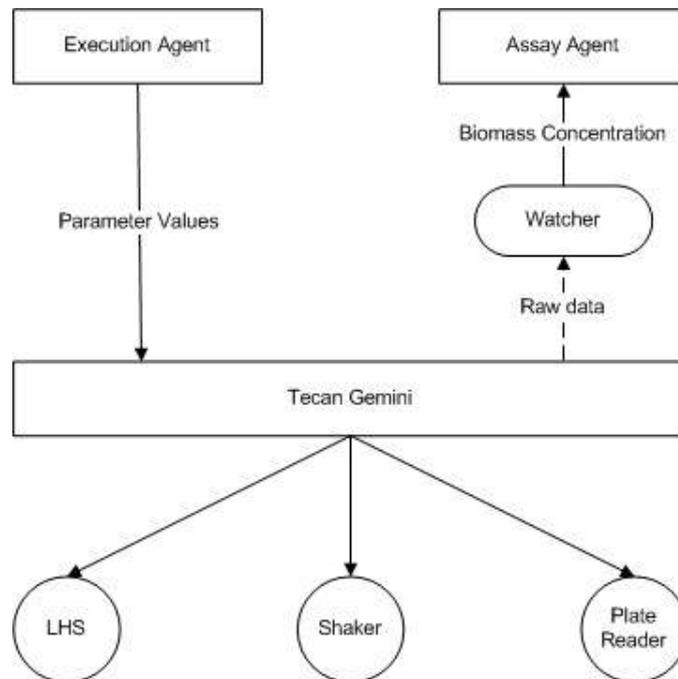


Figure 5.5 Interactions among Assay Agent, Execution Agent and experiment devices.

#### 5.4.4 Experiment Design Agent

The sequential DoE method, which is explained in previous section, is integrated into Experiment Design Agent. In this case, the flowchart of the Experiment Design Agent is shown in Figure 5.6. Starting with low-resolution fractional factorial designs, the method will select significant factors. Based on selected factors, extra experiments will be selected to further investigate the response surface of the chosen factors. Finally the optimal value set of factors will be archived based on response surface study.

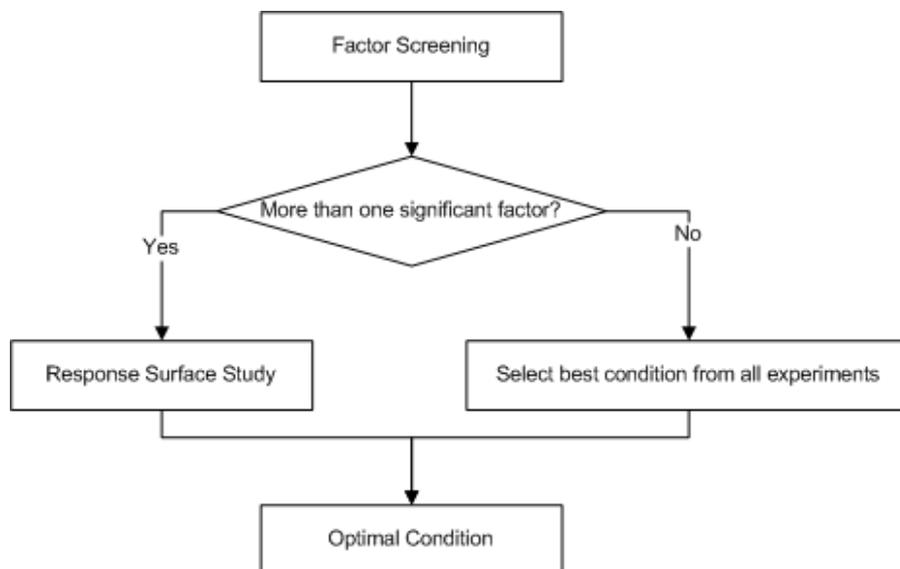


Figure 5.6 Flowchart of Experiment Design Agent

The DoE algorithm was realized from the existing functions in Matlab's Statistics and Machine Learning Toolbox. The algorithm was then compiled into .Net libraries in order to integrate into Experiment Design Agent.

Table 5.2 Matlab toolboxes and functions used in Experiment Design Agent

Toolbox Name	Functions	Usage in IAPBD
Statistics and Machine Learning Toolbox	fracfact	Fractional Factorial Design
	bbdesign	Response Surface Design
	manova	Analysis of Variances
Optimization Toolbox	fmincon	Find optimal value of process model from DoE algorithm
Matlab Compiler SDK	libraryCompiler	Compile Matlab functions into .net library, which can be called by IAPBD

### 5.5 IAPBD demonstration through fermentation optimization

For the fermentation optimization, the following results using above described methods have been achieved. The Execution Agent and the Assay Agent drove the TECAN liquid handling system and the plate reader via Tecan Gemini. The communication between the coordinator agent and other agents was smooth in facilitating data exchange. The Experiment Design Agent used the sequential DoE for process optimization.

A global analysis of engineering variables and media components were performed. The Experiment Design Agent designed 32 experiment runs based on fractional factorial design, as shown in Table 5.3. As shown in Table 5.1, there are 12 factors included in the case study. The full 2-level factorial design will result in  $2^{12}$  experiment runs, which is obviously not practical. Thus a low-resolution factor screening is more preferable at early stage.

Table 5.3 Fractional factorial design for factor screening. -1: low level, 1:high level.

No.	Peptone	Yeast	Glycerol	MgSO <sub>4</sub>	NH <sub>4</sub> Cl	Na <sub>2</sub> HPO <sub>4</sub>	K <sub>2</sub> PO <sub>4</sub>	IPTG	Shaking	Fill Volumn	Induction Time	Temperature
1	-1	-1	-1	1	1	1	1	-1	-1	1	1	-1
2	-1	-1	1	-1	1	-1	1	1	1	-1	1	-1
3	-1	-1	1	1	-1	1	-1	1	-1	-1	1	-1
4	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1
5	-1	1	-1	1	-1	-1	-1	1	1	1	-1	1
6	-1	1	1	-1	-1	-1	1	-1	-1	1	1	-1
7	1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1
8	1	-1	-1	1	-1	-1	1	-1	1	-1	-1	1
9	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1
10	1	-1	1	-1	-1	1	-1	-1	1	1	1	1
11	1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1
12	1	1	-1	-1	-1	1	1	1	-1	-1	1	1
13	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1
14	1	1	1	1	1	-1	-1	-1	-1	-1	1	1
15	1	1	1	1	1	1	1	1	1	1	-1	-1
16	1	1	1	1	1	1	1	1	1	1	1	1
17	1	-1	-1	-1	1	-1	-1	1	-1	1	-1	-1
18	-1	1	1	-1	-1	-1	1	-1	-1	1	1	1
19	-1	-1	-1	1	1	1	1	-1	-1	1	-1	1
20	-1	-1	1	1	-1	1	-1	1	-1	-1	1	1
21	1	-1	-1	1	-1	-1	1	-1	1	-1	-1	-1
22	-1	1	-1	-1	1	1	-1	-1	1	-1	1	-1
23	-1	1	-1	1	-1	-1	-1	1	1	1	-1	-1
24	-1	-1	1	-1	1	-1	1	1	1	-1	1	1
25	1	-1	-1	-1	1	-1	-1	1	-1	1	1	1
26	-1	1	1	-1	-1	-1	1	-1	-1	1	-1	-1
27	-1	-1	1	1	-1	1	-1	1	-1	-1	-1	-1
28	1	-1	-1	1	-1	-1	1	-1	1	-1	1	1
29	-1	1	-1	1	-1	-1	-1	1	1	1	1	-1
30	-1	-1	1	-1	1	-1	1	1	1	-1	-1	1
31	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1
32	1	1	1	1	1	-1	-1	-1	-1	-1	1	-1

96 experiment runs (3 replicates for each experimental condition) were carried out by Tecan LHS. The biomass concentrations were measured afterwards. Among all 32 runs, the biomass concentration ranged from 0.09g/L to 4.86g/L.

Table 5.4 ANOVA of fraction factorial design.<sup>1</sup>

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	12	43.09	3.5908	30.32	0
Peptone (A)	1	0.0026	0.0026	0.02	0.883
Yeast (B)	1	0.0091	0.0091	0.08	0.784
Glycerol (C)	1	1.5123	1.5123	12.77	0.002
MgSO <sub>4</sub> (D)	1	0.155	0.155	1.31	0.267
NH <sub>4</sub> Cl (E)	1	1.7508	1.7508	14.78	0.001
Na <sub>2</sub> HPO <sub>4</sub> (F)	1	0.114	0.114	0.96	0.339
K <sub>2</sub> PO <sub>4</sub> (G)	1	0.7215	0.7215	6.09	0.026
IPTG (H)	1	0.4877	0.4877	4.12	0.057
Shaking (J)	1	35.3548	35.3548	298.49	0
Fill Volume (K)	1	1.2701	1.2701	10.72	0.004
Induction Time (L)	1	0	0	0	0.999
Temperature (M)	1	0.0159	0.0159	0.13	0.718
Error	19	2.2505	0.1184		
Total	31	45.3404			

The ANOVA (Table 5.4) was performed after experimentation. According to ANOVA result, only 4 factors among 12 factors have statistically significant effect on biomass concentration (response factor) based on 95% confidence interval. This is also confirmed by factor plot shown in Figure 5.7.

---

<sup>1</sup> ANOVA result shown in the table is reproduced by Minitab (Minitab Inc. Pennsylvania, U.S.) based on raw data stored in platform's database. In the table, DF stands for degree of freedom. "Adj SS" and "Adj MS" stand for adjusted sums of squares and mean squares.

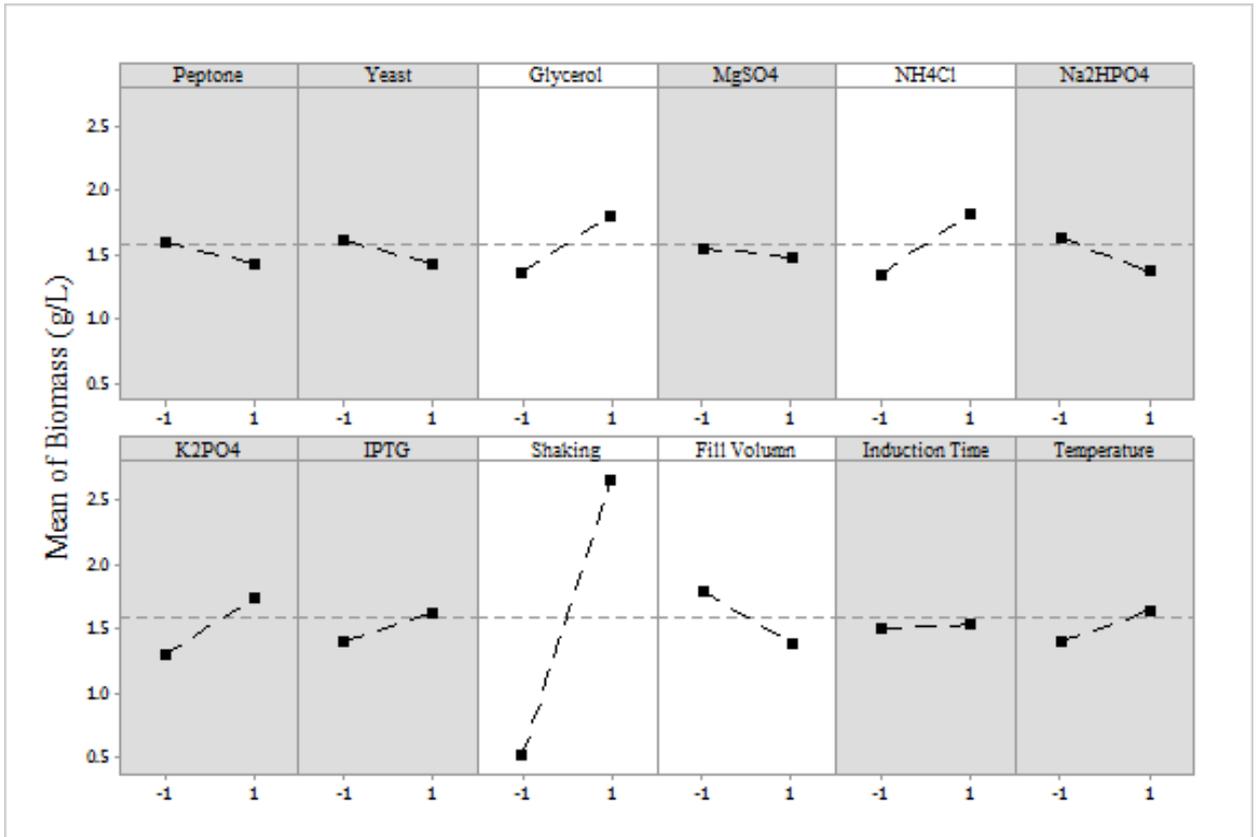


Figure 5.7 Main factor plot of biomass concentration. A gray background represents a factor considered insignificant in the model.

A response surface design was then generated by the Experiment Design Agent based on these four factors. According to Box–Behnken method, the design is a two-level full factorial design combined with central points for each factor. Thus, there are 27 experiment runs in total as shown in Table 5.5. Apart from the selected 4 factors, the values of the other eight factors that led to the highest biomass concentration were chosen based on the previous experiment results.

Table 5.5 Response surface design for selected four factors

No.	Glycerol	NH <sub>4</sub> Cl	Shaking speed	Fill volume
1	0	0	1	1
2	-1	0	0	-1
3	0	0	1	-1
4	-1	1	0	0
5	-1	0	-1	0
6	0	-1	1	0
7	0	0	0	0
8	0	1	0	-1
9	1	0	0	-1
10	1	-1	0	0
11	1	1	0	0
12	0	-1	-1	0
13	0	0	-1	1
14	0	-1	0	-1
15	-1	0	0	1
16	0	1	-1	0
17	0	0	-1	-1
18	1	0	0	1
19	0	1	0	1
20	1	0	-1	0
21	0	0	0	0
22	1	0	1	0
23	-1	0	1	0
24	0	1	1	0
25	0	-1	0	1
26	0	0	0	0
27	-1	-1	0	0

Table 5.6 Factors included in the response surface design

<b>Factors</b>		<b>High (1)</b>	<b>Medium (0)</b>	<b>Low (-1)</b>
Glycerol	% w/v	1.5	0.85	0.2
NH <sub>4</sub> Cl	% w/v	0.2	0.1	0.01
Shaking	rpm	1000	700	400
Fill Volume	mL	1.5	1	0.5

Execution Agent and Assay Agent then carried out the 27 experiment runs on Tecan LHS. The final model was generated based on ANOVA performed in Experiment Design Agent. The result of ANOVA is shown in Table 5.7. As seen in Table 5.7, the quadratic terms have no statistical significance based on 95% confidence interval, while some interaction terms proved to have significant effect on biomass concentration, e.g. ‘glycerol\* NH<sub>4</sub>Cl’, ‘glycerol\*shaking speed’, and ‘NH<sub>4</sub>Cl\*shaking speed’.

Table 5.7 ANOVA for response surface design on four factors: glycerol (A), NH<sub>4</sub>Cl (B), shaking speed (C), fill volume (D).<sup>2</sup>

Source		DF	Adj SS	Adj MS	F-Value	P-Value
Model		14	13.5124	0.9652	144.71	0
	<b>Linear</b>	4	13.1592	3.2898	493.24	0
	A	1	0.5001	0.5001	74.98	0
	B	1	0.0752	0.0752	11.28	0.006
	C	1	12.1374	12.1374	1819.75	0
	D	1	0.4465	0.4465	66.95	0
	<b>Square</b>	4	0.0067	0.0017	0.25	0.903
	A*A	1	0.0008	0.0008	0.12	0.737
	B*B	1	0	0	0.01	0.938
	C*C	1	0.0001	0.0001	0.02	0.899
	D*D	1	0.0036	0.0036	0.53	0.479
	<b>2-Way Interaction</b>	6	0.3464	0.0577	8.66	0.001
	A*B	1	0.152	0.152	22.79	0
	A*C	1	0.1096	0.1096	16.44	0.002
	A*D	1	0.0061	0.0061	0.91	0.358
	B*C	1	0.045	0.045	6.75	0.023
	B*D	1	0.0051	0.0051	0.76	0.399
	C*D	1	0.0286	0.0286	4.28	0.061
Error		12	0.08	0.0067		
	Lack-of-Fit	10	0.0677	0.0068	1.1	0.566
	Pure Error	2	0.0123	0.0062		
Total		26	13.5924			

---

<sup>2</sup> ANOVA result shown in the table is reproduced by Minitab (Minitab Inc. Pennsylvania, U.S.) based on raw data stored in platform's database. In the table, DF stands for degree of freedom. "Adj SS" and "Adj MS" stand for adjusted sums of squares and mean squares.

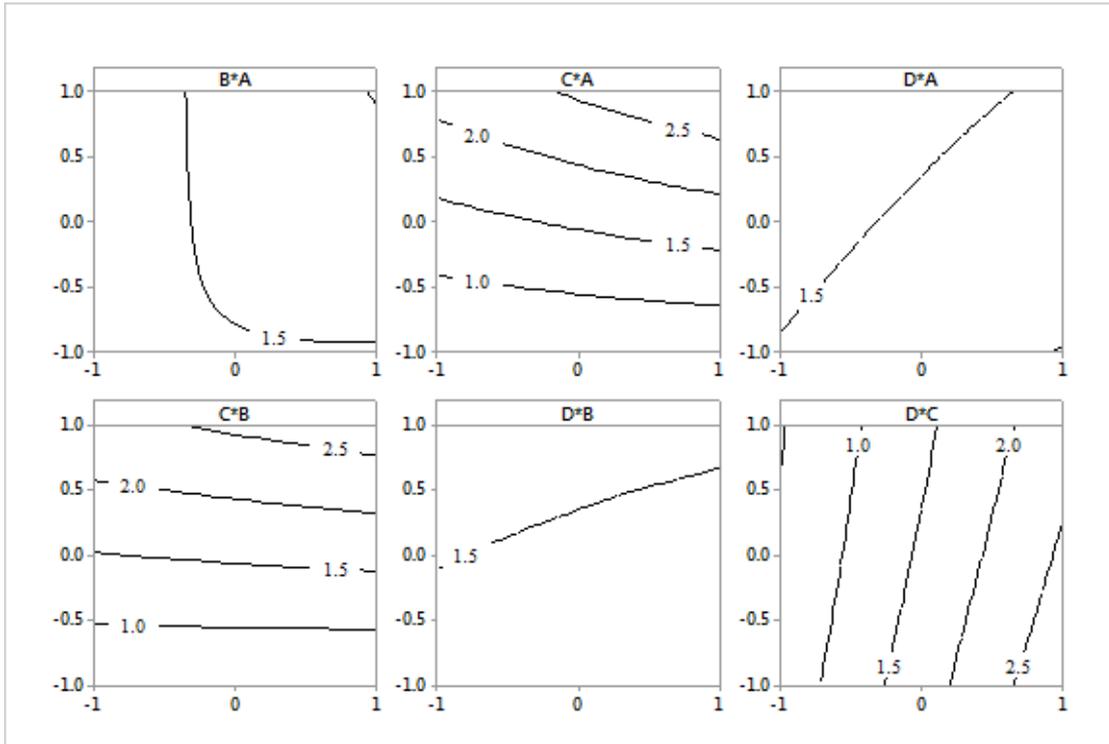


Figure 5.8 Response surface plot of biomass concentration on four factors: glycerol (A), NH<sub>4</sub>Cl (B), shaking speed (C), fill volume (D).

Using stepwise method, the polynomial model has been found in Equation 5.1 with R squared value of 98.96%.

$$\text{Biomass} = 1.5675 + 0.2041A + 0.0792B + 1.0057C - 0.1929D + 0.1950A*B + 0.1656A*C + 0.1061B*C - 0.0845C*D$$

where A stands for the concentration of glycerol, B for the concentration of NH<sub>4</sub>Cl, C for the shaking speed, and D for the fill volume.

The optimal conditions for fermentation process have been found based on the equation as shown in Table 5.8.

Table 5.8 Optimal media and operating conditions for microscale fermentation

<b>Tryptone (g/L)</b>	30
<b>Yeast (g/L)</b>	15
<b>Glycerol (g/L)</b>	90
<b>MgSO<sub>4</sub> (g/L)</b>	2
<b>NH<sub>4</sub>Cl (g/L)</b>	1
<b>Na<sub>2</sub>HPO<sub>4</sub> (g/L)</b>	2
<b>K<sub>2</sub>PO<sub>4</sub> (g/L)</b>	1
<b>Induction Time (h)</b>	None
<b>Temperature (°C)</b>	30
<b>Induction concentration of IPTG (mM)</b>	None
<b>Shaking (rpm)</b>	1000
<b>Fill volume (ml)</b>	0.5

### 5.6 Comparison between tradition implementation and IAPBD

IAPBD has demonstrated its strength to significantly reduce time consumption for microscale fermentation studies on automated LHS. As shown in Figure 5.9, the advantage of IAPBD is not obvious when performing the 1<sup>st</sup> round of experiments. It is because the script that controls cannot be generated dynamically as in Chapter 3 due to lack of APIs for LHS in this case study. Thus, same as traditional approach, researchers still have to manually write script to define each step that LHS need to act for experimentation, which takes about 1-2 days. However, the data analysis time have been shortened from half day into less than 1 minute due to the integration with data analysis methods in IAPBD. The real advantage of IAPBD appears in the 2<sup>nd</sup> iteration. Because communicator can inject updated parameters into scripts, no script coding and debugging is needed this time. The total time consumption of the second round is only 1 day instead of 2.5 days in traditional approach.

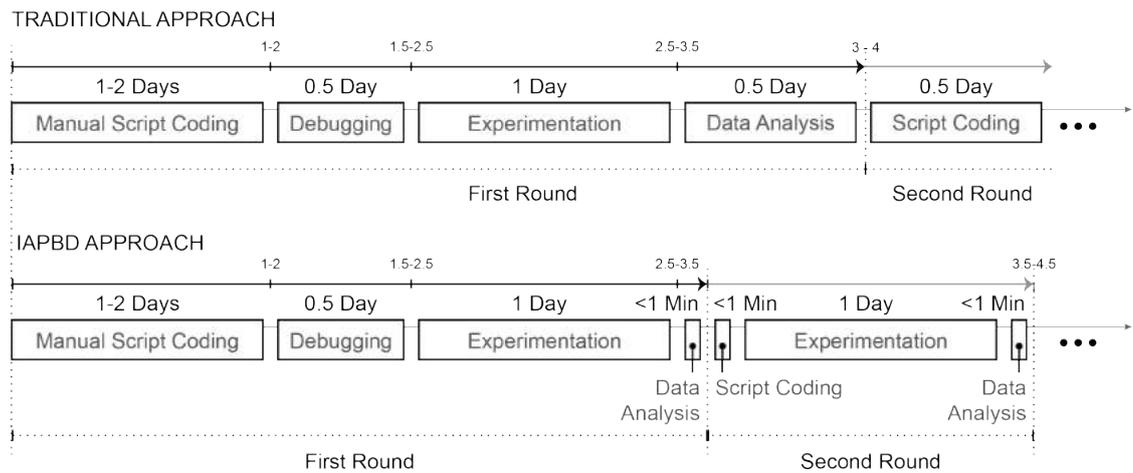


Figure 5.9 Time consumption in the traditional approach and IAPBD

### 5.7 Effect of dissolved oxygen in fermentation

As shown in Equation 5.1, biomass growth was promoted by any factor that would increase oxygen availability like lower fill volume, higher shaking speeds and higher temperature. In order to further study this factor in isolation from all the other factors, Dr. L. Rios-Solis carried out an experiment run manually. Using a microscale DOT probe constructed in our department, the DOT was measured for a microscale fermentation using the selected conditions from DoE results. Figure 5.10 shows the resulting DOT profile as well as the biomass growth kinetics. The microscale fermentation in Figure 11 was oxygen limited as the DOT fell to 0% after 45 min when the early exponential phase began, and remained at 0% until the stationary phase. This is in agreement with recently published works that the limitation of dissolved oxygen plays an import role for final biomass growth in microscale fermentation (Kirk & Szita 2013).

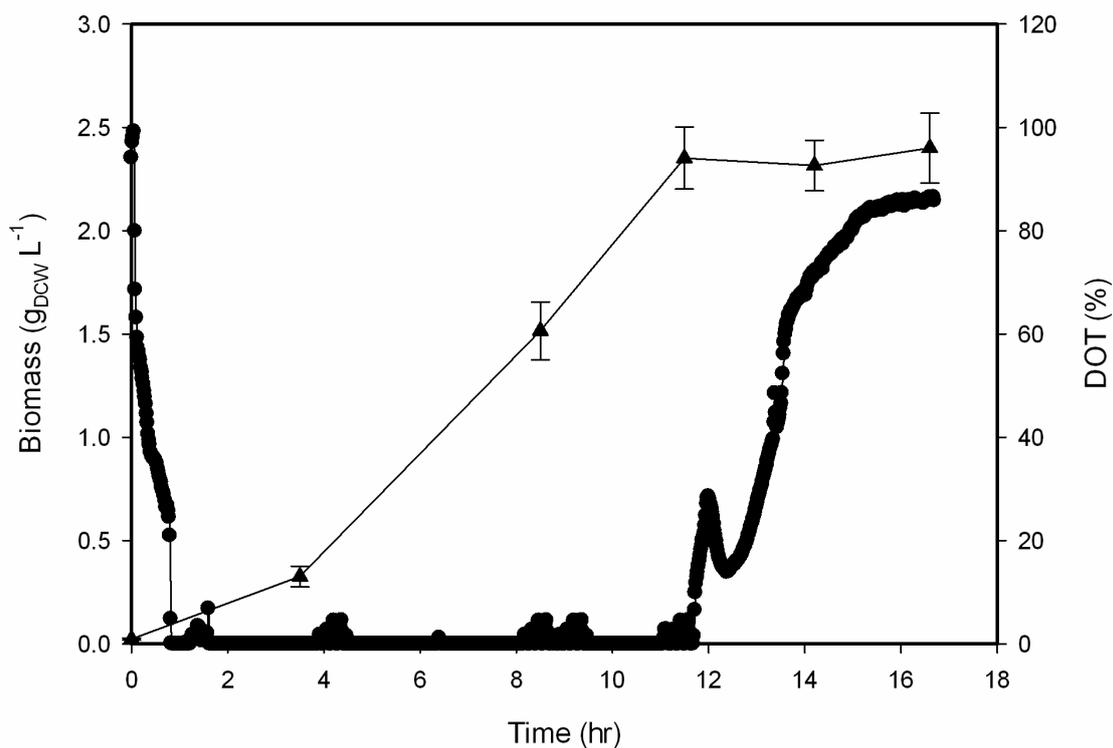


Figure 5.10 Microscale fermentation following (•) DOT (%) and (▲) biomass (g<sub>DCW</sub> l<sup>-1</sup>) using the selected fermentation conditions shown in where A stands for the concentration of glycerol, B for the concentration of NH<sub>4</sub>Cl, C for shaking speed, and D for fill volume.

### 5.8 Pilot scale verification

A 5L fermentation experiment was performed to evaluate the optimized condition from microscale fermentation studies. The media component used in the 5L fermentation is shown in Table 5.8. A fermentation using LB media was also carried out as the control.

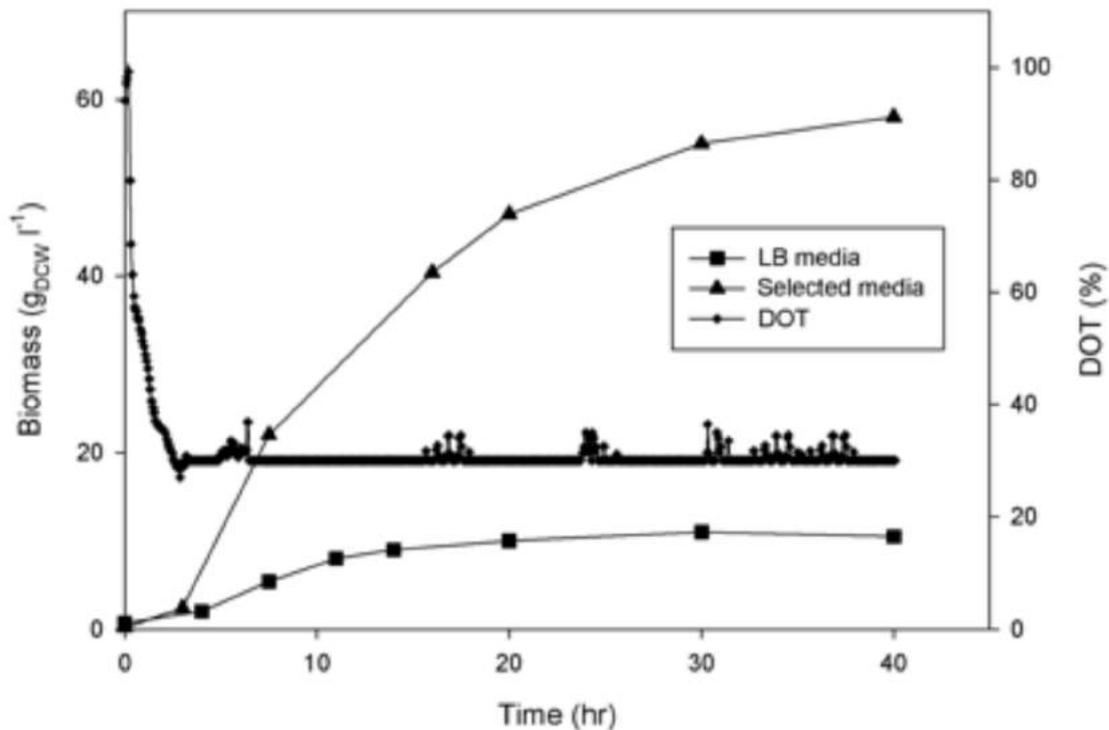


Figure 5.11 The biomass growth in the 5L fermentation using (▲) selected media from Table 10 and (■) LB. (◆) is the DOT for the selected media

As shown in Figure 5.11, the fermentation with the optimized conditions reached a biomass of  $58 \text{ g}_{\text{DCW}} \text{ l}^{-1}$  in 40 hr with a specific growth rate of  $0.65 \text{ hr}^{-1}$ , while the fermentation with LB reached a biomass 5.5 times lower with a specific growth rate of  $0.23 \text{ hr}^{-1}$ . The better performance with optimized conditions further proved that the microscale fermentation driven by the automation platform is capable of upstream process study.

## 5.9 Conclusions

### 5.9.1 Evaluation of IAPBD

The prototype of the intelligent automation platform to perform fermentation experiments automatically has been established. The platform is able to drive liquid handling robotic arm,

shaker and plate reader together to perform all the procedures without human involvement.

The closed-loop learning was achieved by using the sequential DoE method that the experiments were selected iteratively until the final design solution was found. The platform selects initial experiments based on factorial DoE, executes experiments and assays by automated devices, and carries out data analysis by experiment design algorithm to identify the next batch of experiments. The case study has proven that besides simplex search and neural network, various algorithms can be applied in the platform based on the requirement of research. The DoE algorithm shows its power when modelling the relationship among several variables and response factor. Starting with simple media optimization, the platform is capable to model the whole microscale fermentation process involving nearly 20 different variables.

The integration of experiment devices, database, and algorithms for experiment design demonstrates a powerful tool for high throughput process development for not only downstream unit operation but also upstream fermentation process. Although current devices can provide high throughput, the brute force experiment design may still lead to unnecessary waste of time and material. The ‘intelligent’ experimental design by these algorithms, however, can keep the number of experiments low. The elimination of time delay during experiment procedure and utilization of intelligent experiment design in intelligent automation platform has reduced time needed and makes the platform a very effective process optimization tool.

### 5.9.2 Fermentation process optimization

The case study showed operating conditions play more important roles in biomass growth in microwell plate. Shaking speed and fill volume have significant effect for microscale *E. coli* fermentation. This is because these two factors have positive relation with dissolved oxygen in media. Later experiments confirmed that cell growth largely limited by dissolved oxygen in deep well plate. Although it does not matter which types of carbon and nitrogen source

supplied in the media, the quantity of nutrition is essential to *E. coli* because quantities of glycerol and  $\text{NH}_4\text{Cl}$  are proved to be the statistically significant factors.

## **Chapter 6 Conclusions and future work**

### **6.1 Conclusions**

Bioprocess development is very labour intensive, requiring many experiments to characterize each unit operation in the process sequence to achieve product safety and process efficiency. Recent advances in microscale biochemical engineering have led to automated high throughput experimentation. The activities in bioprocess development are implemented sequentially in which 1) liquid handling system performs the wet lab experiments; 2) standalone analytical devices detect the data; and 3) specific software is used for data analysis and experiment design. The experiment design, data analysis and process understanding requires substantial engineer's time and becomes one of the bottlenecks in automated experimentation when a high number of experiments is needed to explore a large design space in order to discover a high process performance. A whole bioprocess often involves a sequence of unit operations. The whole bioprocess design approach will require very large amount of experimental work to understand process interactions.

This thesis has addressed the challenge to reduce the development time and cost without compromising the process efficiency. To increase the process efficiency, a large design space with wide ranges has been explored through micro scale high throughput experimentation. To reduce the development time and cost, a closed-loop learning approach has been adopted in bioprocess design in which the design objectives are driven iteratively by intelligent data acquisition algorithms. In this way fewer number of information rich experiments are needed yet without compromising the process performance. In addition, artificial intelligence technology has been used to transform all manual operations in data processing, experiment design, process evaluation, programming and data management into automation. A framework that brings all of the elements performed manually into an automated fashion has been established to deliver the closed-loop learning approach. This

includes the physical devices that carry out experiment preparation, experiment execution, assays, and different types of software that drives each device. In addition utilization of intelligent algorithms for process evaluation, prediction and optimization as well as automation is designed to reduce the process development time and enhance the process efficiency. Based on the framework, a novel prototype of Intelligent Automation Platform for Bioprocess Development (IAPBD) has been built using multi-agent architecture. A relational database and four agents including Coordinate Agent, Experiment Design Agent, Execution Agent and Assay Agent have been designed to perform individual tasks and worked as a team to deliver the bioprocess design objectives. The multi-agent architecture used a blackboard mechanism. The architecture has been chosen to control all of the software to realize the closed-loop learning. The blackboard mechanism is employed as a communication mechanism to connect the four agents so that they are able to communicate with each other during operations. The blackboard is a place that all agents can publish and read all relevant data in XML format.

Starting with the objective and a set of initial experiments from the user or database, IAPBD is able to drive the robotic arm to perform defined initial experiments, and drive the analytical devices to detect the data after the experiments are completed. Then it will pick up the data from the analytical devices, and carry out data processing and process evaluation based on the optimization objective to achieve the design solution. After evaluation of the design solution, the platform will decide to stop or continue to design the next round experiments for optimization.

The data structure of experiment and data structure of device information have been designed and presented to fit the experiment data sharing purpose. Data are stored using Entity-Relationship model in the database for easy information search and communication. The Experiment Execution Agent has been designed to manage the experiment tasks and the connections with robotic arm LHS and other automated devices such as labware. The agent use programmable interfaces in the automated devices, “translator” algorithm and “watcher” algorithm to connect with the hardware. The Experiment Design Agent is responsible of

designing experiments and publishing them on blackboard. It has a set of experiment design algorithms to use and a general interface to communicate with these algorithms. The Assay Agent is tasked to manage the analytical instruments. The agent also used programmable interface and “watcher” algorithm to connect with the analytical devices. For HPLC, the k-mean cluster algorithm has been developed to turn the HPLC raw data into the measurement. The Coordinate Agent manages the database by retrieving historical experiment data and configuration data, writing new executed experiment information into the database, and manages the information published on blackboard. The communication among the agents is realised by sending and reading messages with defined data structure in XML format to the blackboard. This can eliminate manual operations to improve efficiency and accuracy of experimentation and data analysis during bioprocess design.

The prototype has been evaluated firstly by a lysozyme precipitation process design, which involves typical microscale experimental procedures such as mixing, shaking, sample preparation and high throughput data detection instruments. All of the devices used have programmable interfaces so agents can control them directly. The detailed designs for each agent have been illustrated through this example.

The goal of the experiment is to find an optimal design that maximizes the yield of lysozyme and maximizes the ammonium sulphate concentration by examining pH and ammonium sulphate concentration. An objective function that combines the yield and the ammonium sulphate concentration has been mathematically formulated. The case study is deliberately chosen for its simplicity so as to illustrate the functions of IAPBD. The experiment has several common operations in most of biochemical experiments such as mixing, shaking, incubation, filtration for sample preparation and a plate reader as a high throughput data detection instrument. All of the devices used have programmable interfaces so agents can control them directly. The detailed designs for each agent have been illustrated through this example. In this case study, the multi-agent architecture is proven to be able to drive liquid handling robotic arm, vacuum pump, shaker and plate reader together to perform all the procedures such as assembly adding, shaking, incubation, UV absorption measurement

automatically. Both the simplex search algorithm and the artificial neural network algorithm for experiment design have been implemented in the Experiment Design Agent. The closed loop learning was achieved by using simplex search algorithm and artificial neural network modelling in which the experiments were selected iteratively until the final design solution was found based on the objective function.

Starting with three randomly chosen experiments, the Experiment Design Agent used simplex search method and found the optimal point after 16 iterations. The maximum value of objective function was 0.70, and the ammonium sulphate concentration was 0.8 mol/L and the pH was 7.1. The Experiment Design Agent using the ANN method found optimal point after 4 iterations with 20 experiments executed by IAPBD. The optimal point was at 0.7 mol/L of ammonium sulfate concentration and pH of 7.0, which is very close to the one found by simplex search algorithm. These case studies showed that IAPBD is capable of integrating automated devices and intelligent algorithm together to complete biochemical experiments for process development automatically without human intervention. The success of these case studies proved the concept of IAPBD.

The second case study was designed to verify the capacity of IAPBD when integrating with isolated instruments, which do not have programmable interface, and to investigate the interaction between two step operations, i.e. anion exchange chromatography and precipitation for monoclonal antibody purification. In this work HPLC was required to quantify antibody concentrations but HPLC does not have the programmable interface for connection.

The integration with HPLC was achieved by two algorithms. The first one is a “file watcher” algorithm that monitors the location where ChemStation outputs its raw data files, and reports the latest raw data to the Assay Agent once there is a new file generated. The second is K-mean clustering algorithm that is used to translate the raw data of retention time and peak area into the protein concentration.

After the integration with HPLC has achieved, two sets of experiments were performed on

IAPBD to examine the interaction between the anion exchange chromatography and the precipitation.

The first experiment was a simple mAb precipitation optimization carried out on IAPBD with simplex search algorithm. The results achieved by IAPBD have been verified by previous work done manually. This has verified that IAPBD can perform the precipitation process optimization automatically and accurately. In the next experiment, the feedstock went through AEX first using a set of conditions optimized manually. The material collected from the elution was used as the feedstock for the 2<sup>nd</sup> round mAb precipitation optimization. The impurity of the new feedstock has been reduced almost 50%. The results from two experiments turned out to be slightly different. According to the results, in the 2<sup>nd</sup> experiment, there was 6.3% potential gain in mAb purity while only 2.6% yield was sacrificed. Thus, the interaction between precipitation and AEX does exist based on the experiments. IAPBD carried out two sets of precipitation experiments automatically and discovered the interaction between precipitation and anion exchange chromatography exist so that the integrated process design for two steps may achieve better results.

IAPBD operated smoothly during the case study. It controlled all the relevant automated instruments to execute the experiments automatically, and the “File Watcher” algorithm for connecting with HPLC has been proven robust and efficient to retrieve the data from HPLC.

The third case study is designed to use the sequential Design of Experiment method to optimize the production of a whole cell biocatalyst in fermentation. A different automation platform, Tecan Gemini, was used to satisfy the sterilisation requirement for fermentation. However the LHS has no programmable interface, which provides a challenge for communication that may happen in other devices. In order to communicate with LHS in sample preparation step, a “translator” algorithm acted as a communicator was programmed in Execution Agent to send relevant information to Gemini, i.e. the software that drives Tecan LHS.

The algorithm of the sequential DoE methods has been coded as one of the experiment

design method in the Experiment Design Agent. The fermentation optimization aiming to maximize the biomass concentration hence the intracellular biocatalyst titre has been achieved automatically by IAPBD. IAPBD performed a general factor screening with 12 factors including both media components and operating conditions. An optimal solution for medium composition and operating conditions was found successfully after two iterations, based on a polynomial model established by the Experiment Design Agent. It was then confirmed by the large scale fermentation experiment that the optimal solution has outperformed normal LB medium and the biomass has been increased 5.5 folds. IAPBD worked smoothly during the operations, and it proved to be time efficient.

In summary, a prototype of IAPBD has been built by using the multi-agent architecture and the closed-loop learning framework for bioprocess design has been realised. IAPBD is capable of interacting with robotic arm LHS, isolated instruments to perform bioprocess optimization automatically at microscale. Three case studies further demonstrated that IAPBD's capacity in both upstream and downstream process design. The benefit of this novel IAPBD is its automation that reduces the bioprocess design time significantly and frees engineer's time for other intellectual tasks. With further development and evaluation, its potential to reduce the time and cost of the whole bioprocess development can be realised.

## **6.2 Future work**

This thesis is the first step to develop the intelligent automation platform for bioprocess design. Although the prove-concept has been achieved, further development is anticipated to improve the technology toward commercialisation to realise its true value. Four areas have been identified and recommendations on further researches have been made.

### **6.2.1 Applications of microscale chromatography optimization on IAPBD**

IAPBD has the potential to be used in any bioprocess unit operations when they can be

studied in microwell plate and robotic LHS. In downstream process studies, the thesis only covered precipitation in two case studies. However, chromatography plays an even more important role in current downstream purification process. The number of parameters in chromatography process (e.g. binding buffer conditions, elution buffer conditions, flow rate, etc.) is much more than that in precipitation. The benefit of automated execution and analysis is obvious since the number of experiments required for process development grows exponentially with number of parameters. As mentioned in Chapter 1, several researchers have reported that chromatography can be studied at microscale using robotic LHS. IAPBD has potential to gain better performance comparing with manual operation as the number of parameters increases. Thus using IAPBD for chromatography process parameter screening and design optimization will be an interesting research activity to evaluate its capability.

### 6.2.2 Optimized planning and scheduling functionalities during experiment execution

The efficiency of experiment execution and assay operations can be further improved if advanced planning and scheduling functionalities is introduced in IAPBD. Currently, only one type of experiment has been tested at one time on IAPBD. For example, while IAPBD is running mAb precipitation experiments, no other experiments are planned to run deliberately. In fact, most time is spent on incubation or mixing in bioprocess experiments when LHS stays idle. It is possible to assign some sample preparation tasks to LHS during this time period so that IAPBD is able to carry out more experiments and multiple tasks in bioprocess design may be achieved at the same time. This can potentially reduce whole process development time significantly.

Advanced planning and scheduling techniques will be required for IAPBD when handling several different types of experiments in order to fully utilize the capacity of those automated instruments. This is because the time consumption of each experimental task is different and some tasks are dependent on the completion of other tasks. For example, in

mAb precipitation experiments, feedstock cannot be dispensed into microwell plate before the buffer solution has been added. The techniques will be able to identify which tasks can be executed in parallel from all tasks to be performed. Actually, similar problems have been studied intensively in the fields like manufacturing scheduling and supply chain management. A few of methods have been development in recent years to tackle those challenges, e.g. mixed integer programming and genetic algorithms. Thus, it is promising to integrate IAPBD with those advanced planning techniques to improve the utilization of the resources in bioprocess development.

### 6.2.3 Integration with knowledge engineering techniques

The knowledge and data base module in IAPBD can be further improved by integrating sophisticated knowledge engineering techniques. Currently, IAPBD's database only accumulates historical experiment data that was generated from itself. However, the amount of data available from other sources has increased exponentially since the expansion of new biopharmaceutical development. The development time can be significantly reduced if IAPBD is able to take advantage of all relevant data from other sources (e.g. previous research papers).

Sophisticated knowledge engineering and data mining techniques will help IAPBD better understand the data available. Currently, data in bioprocess studies are heterogeneous, sometimes inconsistent and incomplete. This makes it difficult to be systematically utilized for analysis for bioprocess development. Thus, techniques that can systematically analyse data would benefit IAPBD significantly. Some applications of implementing knowledge engineering into bioprocess development have been reported recently (Zhang, Hunter, & Zhou, 2013). The technique is able to represent the data into a novel data structure for experimental data that can be analysed systematically. Logic reasoning based algorithm has been introduced to generate suggestions for bioprocess development according to historical data.

Historical objective functions from the advanced knowledge base can support researchers make decisions on design objectives. This will enable IAPBD's Experiment Design Agent works more closely with the knowledge base to make decision easier and quicker.

#### 6.2.4 Whole process design and optimization on IAPBD

Each unit operation was studied separately currently on IAPBD, although the second case study in the thesis tried to investigate the interaction between two different operations. The manufacture of biopharmaceutical products is normally composed of several bioprocess including upstream and downstream operations. In most cases, tradeoffs among different processes and objectives are required when expanding to whole process optimization. The whole process design and optimization cannot be achieved by simply adding together all designs based on the single unit operation.

However, if most unit operations now can be studied in microwell plate on robotic instruments, it should be possible to perform multiple process steps on IAPBD continuously from upstream fermentation to downstream purification. The capacity of IAPBD will be further utilized since the number of parameters will increase dramatically if the whole process is studied at the same time.

Alternatively, since IAPBD's knowledge base is able to store all the historical data and statistical models, the whole process can be described by a group of statistical models with multiple design space and constraints. Thus, multi-agent architecture can be expanded as described in Chapter 1. We can develop more process design agents who possess specific process knowledge of unit operations. The whole process design can be achieved through discussion and argumentation among all these agents.

## References

- Allmendinger, R., Simaria, A.S. & Farid, S.S., 2014. Multiobjective evolutionary optimization in antibody purification process design. *Biochemical Engineering Journal*, 91, pp.250–264.
- Angluin, D., 1988. Queries and concept learning. *Machine Learning*, 2, pp.319–342.
- Atkin, K.E. et al., 2008. Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of variants of monoamine oxidase from *Aspergillus niger*. *Acta crystallographica. Section F, Structural biology and crystallization communications*, 64(Pt 3), pp.182–5.
- Baboo, J.Z. et al., 2012. An automated microscale platform for evaluation and optimization of oxidative bioconversion processes. *Biotechnology progress*, 28(2), pp.392–405.
- Bergander, T. et al., 2008. High-throughput process development: Determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin. In *Biotechnology Progress*. pp. 632–639.
- Bergsdorf, C. et al., 2006. A cost-effective solution to reduce dead volume of a standard dispenser system by a factor of 5. *Journal of biomolecular screening*, 11(4), pp.407–412.
- Boyd, J., 2002. Robotic laboratory automation. *Science*, 295(5554), pp.517–518.
- Bray, T. et al., 1997. Extensible markup language (XML). *World Wide Web Journal*, 2(4), pp.27–66.
- Carr, R. et al., 2003. Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. *Angewandte Chemie (International ed. in English)*, 42(39), pp.4807–10.
- Chai, S.C. et al., 2013. *Practical Considerations of Liquid Handling Devices in Drug*

*Discovery*, INTECH Open Access Publisher.

Chen, K.Y. & Chen, C.J., 2010. Applying multi-agent technique in multi-section flexible manufacturing system. In *Expert Systems with Applications*. pp. 7310–7318.

Chen, P.P.-S., 1976. The entity-relationship model---toward a unified view of data. *ACM Transactions on Database Systems*, 1(1), pp.9–36.

Cheng, Y.C. et al., 2006. Kinetics and equilibria of lysozyme precipitation and crystallization in concentrated ammonium sulfate solutions. *Biotechnology and Bioengineering*, 94, pp.177–188.

Chhatre, S. et al., 2010. An automated packed Protein G micro-pipette tip assay for rapid quantification of polyclonal antibodies in ovine serum. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 878, pp.3067–3075.

Chhatre, S., Bracewell, D.G. & Titchener-Hooker, N.J., 2009. A microscale approach for predicting the performance of chromatography columns used to recover therapeutic polyclonal antibodies. *Journal of Chromatography A*, 1216(45), pp.7806–7815.

Chhatre, S. & Titchener-Hooker, N.J., 2009. Review: Microscale methods for high-throughput chromatography development in the pharmaceutical industry. *Journal of Chemical Technology and Biotechnology*, 84, pp.927–940.

Coffman, J.L., Kramarczyk, J.F. & Kelley, B.D., 2008. High-throughput screening of chromatographic separations: I. method development and column modeling. *Biotechnology and Bioengineering*, 100, pp.605–618.

Curling, J.M., 1980. *Methods of plasma protein fractionation*, Academic Pr.

Davidsson, P. & Wernstedt, F., 2002. A multi-agent system architecture for coordination of just-in-time production and distribution. *The Knowledge Engineering Review*, 17(04), pp.317–329.

Davis, R. & Smith, R.G., 1983. Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence*, 20, pp.63–109.

Devlin, J.P., 1997. *High throughput screening: the discovery of bioactive substances*, CRC Press.

Du, C.J. et al., 2014. Evaluation of CV2025  $\omega$ -transaminase for the bioconversion of lignin breakdown products into value-added chemicals: synthesis of vanillylamine from vanillin 1. *Biocatalysis and Biotransformation*, 32(5-6), pp.302–313.

Duetz, W. a, 2007. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. *Trends in microbiology*, 15(10), pp.469–75.

Duran, F.C., Cotta, C. & Fernández, A.J., 2009. *Nature-inspired algorithms for optimisation*, Available at:

<http://www.springerlink.com/index/10.1007/978-3-642-00267-0>  
<http://books.google.co.uk/books?hl=en&lr=&id=557PPQbkfNYC&oi=fnd&pg=PA1&dq=nature+inspired+algorithm&ots=KBhS9RwhMC&sig=gGhQ06L7x43h8ona9AVn30tlv0>.

Eslamimanesh, A. et al., 2011. Artificial neural network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. *Chemical Engineering Science*, 66(13), pp.3039–3044.

Felton, M.J., 2003. Product Review: Liquid handling: Dispensing reliability. *Analytical chemistry*, 75(17), p.397–A.

Ferber, J. & Drogoul, A., 1992. Using reactive multi-agent systems in simulation and problem solving. *Distributed artificial intelligence: Theory and praxis*. Available at: [http://books.google.com/books?hl=en&lr=&id=uBn6At5xzOoC&oi=fnd&pg=PA53&dq=Using+reactive+multi-agent+systems+in+simulation+and+problem+solving&ots=dFW\\_imxRNK&sig=uG6hVPEQOBIK4SkuIDh7naRIQBA\papers2://publication/uuid/D0F903](http://books.google.com/books?hl=en&lr=&id=uBn6At5xzOoC&oi=fnd&pg=PA53&dq=Using+reactive+multi-agent+systems+in+simulation+and+problem+solving&ots=dFW_imxRNK&sig=uG6hVPEQOBIK4SkuIDh7naRIQBA\papers2://publication/uuid/D0F903)

05-1A2C-4BAA-8FF8-175984BA5E30.

Fernandes, P. et al., 2006. Microlitre/millilitre shaken bioreactors in fermentative and biotransformation processes—a review. *and Biotransformation*, 24, pp.237–252. Available at: <http://www.informahealthcare.com/doi/abs/10.1080/10242420600667684>.

Ferreira-Torres, C., Micheletti, M. & Lye, G.J., 2005a. Microscale process evaluation of recombinant biocatalyst libraries: Application to Baeyer-Villiger monooxygenase catalysed lactone synthesis. *Bioprocess and Biosystems Engineering*, 28(2), pp.83–93.

Ferreira-Torres, C., Micheletti, M. & Lye, G.J., 2005b. Microscale process evaluation of recombinant biocatalyst libraries: Application to Baeyer-Villiger monooxygenase catalysed lactone synthesis. *Bioprocess and Biosystems Engineering*, 28, pp.83–93.

Franceschini, G. & Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63, pp.4846–4872.

Galini-Luciani, D., Nguyen, L. & Yazdani, M., 2007. Is PAMPA a useful tool for discovery? *Journal of Pharmaceutical Sciences*, 96, pp.2886–2892.

Gao, Y. et al., 2010. Application of agent-based system for bioprocess description and process improvement. *Biotechnology Progress*, 26, pp.706–716.

Genesereth, M.R. & Ketchpel, S.P., 1994. Software agents. *Communications of the ACM*, 37, pp.48–53. Available at: <http://www-leland.stanford.edu/group/CIFE/online.publications/WP032.pdf>  
[papers://e411e0f6-a7ed-4df0-b2a1-432575887559/Paper/p2424](http://papers://e411e0f6-a7ed-4df0-b2a1-432575887559/Paper/p2424).

Ghodbane, S. & Guiochon, G., 1988. A simplex optimization of the experimental parameters in preparative liquid chromatography. *Chromatographia*, 26, pp.53–59.

Gribbon, P. & Andreas, S., 2005. High-throughput drug discovery: What can we expect from HTS? *Drug Discovery Today*, 10, pp.17–22.

- Habib, G., Zhou, Y. & Hoare, M., 2000. Rapid monitoring for the enhanced definition and control of a selective cell homogenate purification by a batch-flocculation process. *Biotechnology and bioengineering*, 70(2), pp.131–142.
- Halim, M. et al., 2013. Microscale methods to rapidly evaluate bioprocess options for increasing bioconversion yields: application to the  $\omega$ -transaminase synthesis of chiral amines. *Bioprocess and biosystems engineering*.
- Hentz, N.G. & Knaide, T.R., 2014. Effect of liquid-handling accuracy on assay performance. *Journal of laboratory automation*, 19, pp.153–62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24029722>.
- Hober, S., Nord, K. & Linhult, M., 2007. Protein A chromatography for antibody purification. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 848, pp.40–47.
- Horstkotte, B., Duarte, C.M. & Cerdà, V., 2010. Response functions for SIMPLEX optimization of flow-injection analysis and related techniques. *TrAC - Trends in Analytical Chemistry*, 29, pp.1224–1235.
- Hüser, J. et al., 2006. High- throughput Screening for Targeted Lead Discovery. *High-Throughput Screening in Drug Discovery, Volume 35*, pp.15–36.
- Huuskonen, J., Salo, M. & Taskinen, J., 1997. Neural network modeling for estimation of the aqueous solubility of structurally related drugs. *Journal of pharmaceutical sciences*, 86, pp.450–454.
- Jackson, N.B., Liddell, J.M. & Lye, G.J., 2006. An automated microscale technique for the quantitative and parallel analysis of microfiltration operations. *Journal of Membrane Science*, 276, pp.31–41.
- Ji, Y., 2012. Model based process design for bioprocess optimisation: case studies on precipitation with its applications in antibody purification.

Karlsson, D. et al., 2004. Model-based optimization of a preparative ion-exchange step for antibody purification. *Journal of Chromatography A*, 1055(1-2), pp.29–39.

Kennedy, M.J., Prapulla, S.G. & Thakur, M.S., 1992. Designing fermentation media: A comparison of neural networks to factorial design. *Biotechnology Techniques*, 6, pp.293–298.

Khosravi, A. et al., 2011. Comprehensive review of neural network-based prediction intervals and new advances. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 22(9), pp.1341–1356.

King, J.M.P., Titchener-Hooker, N.J. & Zhou, Y., 2007. Ranking bioprocess variables using global sensitivity analysis: a case study in centrifugation. *Bioprocess and biosystems engineering*, 30(2), pp.123–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17235605> [Accessed June 18, 2014].

King, R.D. et al., 2004. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427, pp.247–252.

Kirk, T. V. & Szita, N., 2013. Oxygen transfer characteristics of miniaturized bioreactor systems. *Biotechnology and Bioengineering*, 110, pp.1005–1019.

Knevelman, C. et al., 2010. High-throughput screening techniques for rapid PEG-based precipitation of IgG4 mAb from clarified cell culture supernatant. *Biotechnology Progress*, 26, pp.697–705.

Knuth, D.E., 1973. The Art of Computer Programming. *Reading MA*, 2(116), p.401. Available at: <http://neerc.ifmo.ru/school/io/archive/20090516/problems-advanced-07.pdf>.

Köhler, V. et al., 2010. Enantioselective Biocatalytic Oxidative Desymmetrization of Substituted Pyrrolidines \*\*. , (100 mm), pp.1–4.

Kong, F. et al., 2012. Automatic liquid handling for life science: a critical review of the

current state of the art. *Journal of laboratory automation*, 17(3), pp.169–85. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22357568> [Accessed June 10, 2014].

Konstantinidis, S. et al., 2012. The hybrid experimental simplex algorithm--an alternative method for “sweet spot” identification in early bioprocess development: case studies in ion exchange chromatography. *Analytica chimica acta*, 743, pp.19–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22882820> [Accessed June 18, 2014].

Kostov, Y. et al., 2001. Low-cost microbioreactor for high-throughput bioprocessing. *Biotechnology and Bioengineering*, 72, pp.346–352.

Kumar, V., 1998. Introduction to Robotics.

Lek, S. et al., 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, 90(1), pp.39–52.

Li, F. et al., 2005. Current Therapeutic Antibody Production and Process Optimization. *BioProcessing J*, pp.1–8. Available at: [http://www.novaseptic.ca/references/files/pmc\\_url/\\$file/rp1017en00.pdf](http://www.novaseptic.ca/references/files/pmc_url/$file/rp1017en00.pdf).

Low, D., O’Leary, R. & Pujar, N.S., 2007. Future of antibody purification. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 848, pp.48–63.

Lye, G.J. et al., 2003. Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends in Biotechnology*, 21(1), pp.29–37.

Ma, G. et al., 2009. Mimic of a large-scale diafiltration process by using ultra scale-down rotating disc filter. *Biotechnology progress*, 26(2), pp.466–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19938066>.

Mannhold, R. et al., 2006. *High-throughput screening in drug discovery*, John Wiley & Sons.

- Markin, R.S., 1992. Laboratory automation systems. An introduction to concepts and terminology. *American journal of clinical pathology*, 98(4 Suppl 1), pp.S3–S10.
- Maturana, F.P. & Norrie, D.H., 1996. Multi-agent Mediator architecture for distributed manufacturing. *Journal of Intelligent Manufacturing*, 7, pp.257–270.
- Medeiros, A.B.P. et al., 2000. Optimization of the production of aroma compounds by *Kluyveromyces marxianus* in solid-state fermentation using factorial design and response surface methodology. *Biochemical Engineering Journal*, 6, pp.33–39.
- Micheletti, M. & Lye, G.J., 2006. Microscale bioprocess optimisation. *Current Opinion in Biotechnology*, 17(6), pp.611–618.
- Mishra, K.P. et al., 2008. A review of high throughput technology for the screening of natural products. *Biomedicine and Pharmacotherapy*, 62, pp.94–98.
- Neal, G. et al., 2003. Ultra scale-down approach for the prediction of full-scale recovery of ovine polyclonal immunoglobulins used in the manufacture of snake venom-specific fab fragment. *Biotechnology and Bioengineering*, 81(2), pp.149–157.
- Nierstrasz, O., 1989. A Survey of Object-Oriented Concepts. In *ObjectOriented Concepts Databases and Applications*. pp. 3–21. Available at:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.5247&rep=rep1&type=pdf>.
- Nii, H.P., 1986. The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures. *AI Magazine*, 7, pp.38–53.
- Niktari, M. et al., 1990. The monitoring and control of protein purification and recovery processes. In *Separations for Biotechnology 2*. Springer, pp. 622–631.
- Nygaard, K. & Dahl, O.-J., 1978. The development of the SIMULA languages. *ACM SIGPLAN Notices*, 13(8), pp.245–272.

Olafsson, S., 2006. Metaheuristics. *Handbooks in operations research and management science*, 13, pp.633–654.

Olsson, D.M. & Nelson, L.S., 1975. The Nelder-Mead Simplex Procedure for Function Minimization. *Technometrics*, 17, pp.45–51. Available at: <http://www.jstor.org/stable/1267998>.

Pereira, D. a & Williams, J. a, 2007. Origin and evolution of high throughput screening. *British journal of pharmacology*, 152(1), pp.53–61.

Petersen, J. & Nguyen, J., 2005. Comparison of absorbance and fluorescence methods for determining liquid dispensing precision. *JALA - Journal of the Association for Laboratory Automation*, 10(2), pp.82–87.

Pirdashti, M. et al., 2013. Artificial neural networks: applications in chemical engineering. *Reviews in Chemical Engineering*, 29(4), pp.205–239. Available at: <http://www.degruyter.com/view/j/revce.2013.29.issue-4/revce-2013-0013/revce-2013-0013.xml>.

Popa-Burke, I. et al., 2009. Establishing quality assurance criteria for serial dilution operations on liquid-handling equipment. *Journal of biomolecular screening*, 14(8), pp.1017–30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19675311>.

Popova, D. et al., 2015. Representative mammalian cell culture test materials for assessment of primary recovery technologies: A rapid method with industrial applicability. *Biotechnology journal*, 10(1), pp.162–170.

Puskeiler, R., Kaufmann, K. & Weuster-Botz, D., 2005. Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD). *Biotechnology and Bioengineering*, 89, pp.512–523.

Rayat, A.C.M.E., Micheletti, M. & Lye, G.J., 2010. Evaluation of cell disruption effects on primary recovery of antibody fragments using microscale bioprocessing techniques.

*Biotechnology Progress*, 26(5), pp.1312–1321. Available at:

<http://dx.doi.org/10.1002/btpr.450>.

Rhode, H. et al., 2004. An improved method for checking HTS/uHTS liquid-handling systems. *Journal of biomolecular screening*, 9(8), pp.726–33. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/15634800>.

Rios-Solis, L. et al., 2013. Non-linear kinetic modelling of reversible bioconversions: Application to the transaminase catalyzed synthesis of chiral amino-alcohols. *Biochemical Engineering Journal*, 73, pp.38–48.

Roque, A.C.A., Silva, C.S.O. & Taipa, M.Â., 2007. Affinity-based methodologies and ligands for antibody purification: Advances and perspectives. *Journal of Chromatography A*, 1160, pp.44–55.

Rose, D., 1999. Microdispensing technologies in drug discovery. *Drug Discovery Today*, 4(9), pp.411–419.

Saitoh, S. & Yoshimori, T., 2008. Fully Automated Laboratory Robotic System for Automating Sample Preparation and Analysis to Reduce Cost and Time in Drug Development Process. *JALA - Journal of the Association for Laboratory Automation*, 13, pp.265–274.

Sasaki, M. et al., 1998. Total laboratory automation in Japan: Past, present and the future. In *Clinica Chimica Acta*. pp. 217–227.

Sasaki, M. & Sugiura, T., 2000. Does the Future Shine Bright for TLA? *Journal of the Association for Laboratory Automation*, 5(3), pp.40–43.

Settles, B., 2010. Active Learning Literature Survey. *Machine Learning*, 15, pp.201–221.

Settles, B. & Craven, M., 2008. An analysis of active learning strategies for sequence labeling tasks. *Proceedings of the Conference on Empirical Methods in Natural Language*

*Processing EMNLP 08*, p.1070. Available at:

<http://portal.acm.org/citation.cfm?doid=1613715.1613855>.

Shukla, A.A. et al., 2007. Downstream processing of monoclonal antibodies-Application of platform approaches. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 848, pp.28–39.

Shukla, A.A. & Thömmes, J., 2010. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends in Biotechnology*, 28, pp.253–261.

Siddiqi, S.F., Titchener-Hooker, N.J. & Shamlou, P.A., 1996. Simulation of particle size distribution changes occurring during high-pressure disruption of bakers' yeast.

*Biotechnology and bioengineering*, 50(2), pp.145–150.

Sommerfeld, S. & Strube, J., 2005. Challenges in biotechnology production—generic processes and process optimization for monoclonal antibodies. *Chemical Engineering and Processing: Process Intensification*, 44, pp.1123–1137.

Stefik, M. & Bobrow, D.G., 1985. Object-Oriented Programming : Themes and Variations. *AI Magazine*, 6(4), pp.41 – 53.

Stickel, J.J. & Fotopoulos, A., 2001. Pressure-flow relationships for packed beds of compressible chromatography media at laboratory and production scale. *Biotechnology Progress*, 17, pp.744–751.

Susanto, A. et al., 2009. High Throughput Screening for the Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems.

*Chemical Engineering & Technology*, 32, pp.140–154. Available at:

<http://doi.wiley.com/10.1002/ceat.200800350>.

Sycara, K.P., 1998. Multiagent Systems. *AI Magazine*, 19, p.79. Available at:

<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1370>.

- Szita, N. et al., 2005. Development of a multiplexed microbio reactor system for high-throughput bioprocessing. *Lab on a chip*, 5, pp.819–826.
- Taylor, P.B. et al., 2002. A standard operating procedure for assessing liquid handler performance in high-throughput screening. *Journal of biomolecular screening*, 7(6), pp.554–69. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14599354>.
- Thömmes, J. & Etzel, M., 2007. Alternatives to chromatographic separations. In *Biotechnology Progress*. pp. 42–45.
- Thurow, K. et al., 2004. Laboratory Information Management Systems for Life Science Applications. *Organic Process Research & Development*, 8, pp.970–982. Available at: [http://pubs3.acs.org/acs/journals/doi/lookup?in\\_doi=10.1021/op040017s](http://pubs3.acs.org/acs/journals/doi/lookup?in_doi=10.1021/op040017s).
- Titchener-Hooker, N.J., Dunnill, P. & Hoare, M., 2008. Micro biochemical engineering to accelerate the design of industrial-scale downstream processes for biopharmaceutical proteins. *Biotechnology and Bioengineering*, 100, pp.473–487.
- Tugcu, N., Roush, D.J. & Göklen, K.E., 2008. Maximizing productivity of chromatography steps for purification of monoclonal antibodies. *Biotechnology and Bioengineering*, 99, pp.599–613.
- Turner, N.J., 2011. Enantioselective oxidation of C-O and C-N bonds using oxidases. *Chemical reviews*, 111(7), pp.4073–87.
- Tustian, A.D. et al., 2007. Adapted ultra scale-down approach for predicting the centrifugal separation behavior of high cell density cultures. *Biotechnology Progress*, 23(6), pp.1404–1410.
- Wellman, M.P., 1995. A computational market model for distributed configuration design. *Ai Edam Artificial Intelligence For Engineering Design Analysis And Manufacturing*, 9, pp.125–133. Available at: [http://www.journals.cambridge.org/abstract\\_S089006040000216X](http://www.journals.cambridge.org/abstract_S089006040000216X).

- Werbos, P.J., 1990. Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 78(10), pp.1550–1560.
- Weuster-Botz, D., 2000. Experimental design for fermentation media development: statistical design or global random search? *Journal of bioscience and bioengineering*, 90, pp.473–483.
- Whelan, K.E. et al., 2004. Intelligent software for laboratory automation. *Trends in biotechnology*, 22, pp.440–445. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0167779904002033>.
- Wiendahl, M. et al., 2008. High throughput screening for the design and optimization of chromatographic processes—miniaturization, automation and parallelization of breakthrough and elution studies. *Chemical engineering & technology*, 31(6), pp.893–903.
- Wooldridge, M. & Jennings, N.R., 1995. Intelligent agents: theory and practice. *The Knowledge Engineering Review*, 10, p.115.
- Wu, C.F.J. & Hamada, M.S., 2011. *Experiments: Planning, Analysis, and Optimization*, Wiley. Available at: <http://books.google.co.uk/books?id=SBgehORJ7hkC>.
- Wurm, F.M., 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature biotechnology*, 22, pp.1393–1398.
- Zhang, X., Zhang, S. & He, X., 2004. Prediction of solubility of lysozyme in lysozyme--NaCl--H<sub>2</sub>O system with artificial neural network. *Journal of crystal growth*, 264(1), pp.409–416.
- Znabet, A. et al., 2010. A highly efficient synthesis of telaprevir by strategic use of biocatalysis and multicomponent reactions. *Chemical communications (Cambridge, England)*, 46(42), pp.7918–20.



## Appendix I Raw data for *E. coli* fermentatio

# 1 General factor optimization: 1<sup>nd</sup> iteration

Factor values are normalized from -1 to 1.

Peptone	Yeast	Glycerol	MgSO4	NH4Cl	Na2HPO4	K2PO4	IPTG	Shaking	Fill Volumn	Induction Time	Temperature	Biomass (g)
-1	-1	-1	1	1	1	1	-1	-1	1	1	-1	0.48
-1	-1	1	-1	1	-1	1	1	1	-1	1	-1	3.53
-1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	0.86
-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	2.84
-1	1	-1	1	-1	-1	-1	1	1	1	-1	1	1.91
-1	1	1	-1	-1	-1	1	-1	-1	1	1	-1	0.52
1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	0.21
1	-1	-1	1	-1	-1	1	-1	1	-1	-1	1	2.24
1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	1.98
1	-1	1	-1	-1	1	-1	-1	1	1	1	1	2.53
1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1	0.46
1	1	-1	-1	-1	1	1	1	-1	-1	1	1	0.23
1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1.41
1	1	1	1	1	-1	-1	-1	-1	-1	1	1	0.69
1	1	1	1	1	1	1	1	1	1	-1	-1	2.69
1	1	1	1	1	1	1	1	1	1	1	1	3.20
1	-1	-1	-1	1	-1	-1	1	-1	1	-1	-1	0.42
-1	1	1	-1	-1	-1	1	-1	-1	1	1	1	0.33
-1	-1	-1	1	1	1	1	-1	-1	1	-1	1	0.14
-1	-1	1	1	-1	1	-1	1	-1	-1	1	1	0.22
1	-1	-1	1	-1	-1	1	-1	1	-1	-1	-1	2.32
-1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	2.15

-1	1	-1	1	-1	-1	-1	1	1	1	-1	-1	2.08
-1	-1	1	-1	1	-1	1	1	1	-1	1	1	4.24
1	-1	-1	-1	1	-1	-1	1	-1	1	1	1	0.30
-1	1	1	-1	-1	-1	1	-1	-1	1	-1	-1	0.31
-1	-1	1	1	-1	1	-1	1	-1	-1	-1	-1	0.23
1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	2.34
-1	1	-1	1	-1	-1	-1	1	1	1	1	-1	2.11
-1	-1	1	-1	1	-1	1	1	1	-1	-1	1	3.65
1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	1.16
1	1	1	1	1	-1	-1	-1	-1	-1	1	-1	0.77

## 2 General factor optimization: 2<sup>nd</sup> iteration

glycerol (A), NH<sub>4</sub>Cl (B), shaking speed (C), fill volume (D). Values are normalized from -1 to 1.

A	B	C	D	Biomass (g l <sup>-1</sup> )
0	0	1	1	2.30
-1	0	0	-1	1.56
0	0	1	-1	2.80
-1	1	0	0	1.29
-1	0	-1	0	0.48
0	-1	1	0	2.48
0	0	0	0	1.65
0	1	0	-1	1.85
1	0	0	-1	2.08
1	-1	0	0	1.43
1	1	0	0	2.03
0	-1	-1	0	0.54
0	0	-1	1	0.55
0	-1	0	-1	1.67
-1	0	0	1	1.22
0	1	-1	0	0.49
0	0	-1	-1	0.71
1	0	0	1	1.58
0	1	0	1	1.36
1	0	-1	0	0.59
0	0	0	0	1.55
1	0	1	0	2.87
-1	0	1	0	2.11
0	1	1	0	2.86
0	-1	0	1	1.33
0	0	0	0	1.49
-1	-1	0	0	1.47

## **Appendix II Paper by the Author**