A Collective Intelligence Framework for *In Silico* Representations of Biomolecules and their Activities

Sathish Periyasamy

School of Computer Science & Informatics
and
School of Biosciences
Cardiff University

May 2010

Thesis submitted in partial fulfilment of the requirement for the degree
of
Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 14/12/2010

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ..... (candidate) Date 14/12/2010

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by explicit references.

Signed ..... (candidate) Date 14/12/2010

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date 14/12/2010

ABSTRACT

One of the key challenges in systems biology is the modelling of cellular systems. Advanced models with molecular information that facilitate in the prediction of cellular behaviour under various conditions are fundamental for revealing cellular level characteristics and underlying principles of cellular functions. It has been acknowledged that the success of systems biology depends not only on studies based on specific instance of life, but also on studies based on the principles governing the entire organisational space of life. The modelling of adaptive dynamics is identified as an essential requirement to understand the organisational space of biological systems. The novel Collective Intelligence framework proposed in this thesis offers great potential for modelling multi-scale adaptive dynamics from molecules to cell in physiological timescale. The major contribution to systems biology is based on defining cellular functions in the context of a multi-objective topology and implementing this principle, as an *in silico* model, to study performances of intracellular functions by measuring the activities of diverse species of functional products. The major contribution to computing is identifying a novel Collective Intelligence approach based on information processing strategies of biomolecules and utilising it for modelling intracellular activities. The aim of the thesis is to investigate systems biology approaches in representing biological complexity from molecules to cells and developing computational approaches to bring abstract theories to practical use by: (1) Characterising major biomolecular self-organising mechanisms. (2) Using a bottom-up integrative approach to model intercellular organisational behaviour. (3) Develop a Collective Intelligence based cell modelling and simulation environment to conduct analysis studies. This thesis argues that a system theoretic approach based on Collective Intelligence where the concepts of self-organisation and emergence underlie the approach is ideal to represent the multi-scale and multi-tasking nature of a biological cell from the bottom-up. This thesis proposes a Collective Intelligence based cell modelling and simulation environment using agents to conduct analysis studies on collective behaviour of biomolecules.

This thesis is dedicated to my late father Pitchapillai Periyasamy, who inspired me to do science. He passed away few weeks after my viva.

ACKNOWLEDGEMENTS

I am grateful to my supervisors, Professor Alex Gray and Dr. Peter Kille for making every effort to relieve me from the administrative and financial problems, that I encountered during the course of my research. Although I hardly had the composure to pursue the research, their continuous support was influential in continuing the research.

I owe my deepest gratitude to Professor Alex Gray without whom I would have never achieved this thesis. He provided the opportunity for me to be a candidate for a PhD and has helped me at critical moments.

I would also like to show my gratitude to Dr. Peter Kille with whom I spent most of the time during my research. He has been a figure of motivation, a great listener and gave me the courage and liberty to pursue the desired research questions.

Further, I would like to thank Mrs. Margaret Evans and Dr. Pamela Munn for all the administrative support I received, for attending conferences and arranging meetings with my supervisors. I would also like to thank Mrs. Helen Williams who on many occasions helped me with my student visa applications, monitoring my progress and frequently reminding me of my thesis submission. Her continuous effort for monitoring the progress of research students has compelled me to produce many progress reports, which facilitated to formulate my thesis.

I am indebted to the Repast development team especially Dr. Eric Tatara and Dr. Nicholson T. Collier at the Center for Complex Adaptive Agent Systems Simulation, Decision and Information Sciences Division, Argonne National Laboratory and Mr. Miles Parker the founder of Metascape, LLC for their valuable technical support in implementing the model.

I am most grateful to my late father, mother, sisters, in-laws, cousins, uncles and aunts for their love, blessings and affection. They provided moral support that was influential for the completion of my thesis.

Finally, I would like to show my gratitude to Ms. Subashini Shamala Rajasingham, who has been sharing the successful and frustrating moments of my research from the other side of the globe.

CONTENTS

Chapter 1 Introduction 1
1.1 Overview 1
1.2 Background 1
1.3 Motivation 3
1.4 Research Required 6
1.5 Relevant Work 8
1.5.1 Corresponding Paradigms 8
1.6 Aim, Objectives and Hypothesis of the Research 9
1.6.1 Aim of the Research 9
1.6.2 Objectives of the Research 9
1.6.3 Hypothesis 10
1.7 Research Direction 11
1.8 Research Methodology 13
1.9 Research Contributions 14
1.10 The Scope and Organisation of this Thesis 16
1.11 Concluding Remarks 18
Chapter 2 Multi-Scale Adaptive Dynamics from Molecules to Cell 19
2.1 Overview 19
2.2 Cellular Organisation, Adaptation and Complexity from the Bottom-Up 20
2.2.1 Characteristics of Biomolecules and Cells 21
2.3 Biological Adaptation from the Fundamental and Organisational Perspectives 23
2.4 Organisational Space of Biological Cells 27
2.4.1 Cellular Level Properties 29
2.4.2 Uncertainty in a Cellular Environment 34
2.4.3 The Impact of Time and Energy in Biological Adaptation 36
2.5 Multi-Level Biological Organisation 37
2.5.1 Task Formation and Integration in Cells 39
2.5.2 Units and Levels of Selection 43
2.5.3 Timescales of Biological Adaptation 44
2.6 Concluding Remarks 46
Chapter 3 Modelling Biological Phenomena 47
3.1 Overview 47

3.2 Requirements for Modelling the Adaptive Dynamics from Molecules to Cell 48
3.2.1 Model Requirements 48
3.2.2 Data Requirements 49
3.2.3 Technical Requirements 50
3.2.4 Non-Functional Requirements 50
3.3 The Scope of Systems Biology 51
3.3.1 Top-Down Systems Biology 51
3.3.2 Bottom-Up Systems Biology 52
3.3.3 Discovering General Principles of Biological System Behaviour 52
3.4 Biological Modelling Methodologies 53
3.4.1 Model Development Processes 53
3.4.2 Hierarchical Modelling 55
3.5 Existing Biological Modelling Formalisms 56
3.5.1 Mathematical Modelling Formalisms 58
3.5.2 Computational Modelling Formalisms 62
3.5.3 Hybrid Modelling Formalisms 67
3.6 Evaluation of Modelling Methodologies and Formalisms against Model Requirements 68
3.6.1 Agent Frameworks Applicable for Modelling Cellular Phenomena 73
3.6.2 Selection and Implementation of an Appropriate Approach 74
3.7 Feasibility of Meeting Model Requirements 75
3.7.1 Resource Requirements and Availability 76
3.7.2 Technological Feasibility 82
3.8 Concluding Remarks 83
Chapter 4 Representation of Biomolecules and their Activities within an *In silico* Environment 85
4.1 Overview 85
4.2 Swarming the Internal Organisation of Biological Cells 86
4.2.1 The Principles of Swarming 86
4.2.2 Analysing and Designing Swarm Systems 91
4.2.3 Collective Intelligence Framework 94
4.3 Modelling the Characteristics and Behaviour of Biomolecules 97
4.4 Modelling Interactions of Biomolecules 99
4.4.1 Direct Interactions 102
4.4.2 Indirect Interactions 103
4.4.3 Local Interactions 103
4.5 Modelling Collective Behaviour of Biomolecules 103

4.5.1 Biological Constraints of Intracellular Organisation 104
4.5.2 Physical Constraints of Biomolecular Interactions 105
4.6 The Cellular Environment 110
4.7 Concluding Remarks 111
Chapter 5 A Collective Intelligence Approach to Modelling Intelligent Cellular Organisation 113
5.1 Overview 113
5.2 The Model 114
5.2.1 Purpose 115
5.2.2 Entities, State Variables and Scales 115
5.2.3 Process Overview and Scheduling 117
5.2.4 Design Concepts 117
5.2.5 Initialisation 123
5.2.6 Input Data 123
5.2.7 Sub-models 123
5.3 Concluding Remarks 130
Chapter 6 Swarm Based Cell Modelling and Simulation Environment 131
6.1 Overview 131
6.2 Repast Simphony Simulation Package 132
6.3 Model Implementation 132
6.4 The Master Context 134
6.5 The Sub-Contexts 135
6.5.1 Physical Layer 135
6.5.2 Pathway Layer 136
6.5.3 Cell Layer 136
6.6 Native Biomolecules 137
6.7 Environments 137
6.7.1 Observations 138
6.8 Model Interface 138
6.9 Model Execution 141
6.10 Concluding Remarks 141
Chapter 7 Model Evaluation by Simulating Biological Phenomena 143
7.1 Overview 143
7.2 The Verification of Simulation Experiments 144
7.3 Simulating Biomolecular Degradation 144
7.3.1 Experiment Hypothesis 144
7.3.2 Experiment Design 145

7.3.3 Experiment Setup 146
7.3.4 Experiment Results 146
7.3.5 Result Analysis 147
7.3.6 Experiment Validation 148
7.3.7 Experiment Conclusion 148
7.4 Simulating Degeneracy in a Biochemical System 149
7.4.1 Experiment Hypothesis 150
7.4.2 Experiment Design 150
7.4.3 Experiment Setup 152
7.4.4 Experiment Results 153
7.4.5 Result Analysis 154
7.4.6 Experiment Validation 158
7.4.7 Experiment Conclusion 159
7.5 Simulating Competition in Metabolic Activities 159
7.5.1 Experiment Hypothesis 160
7.5.2 Experiment Design 161
7.5.3 Experiment Setup 163
7.5.4 Experiment Results 164
7.5.5 Result Analysis 166
7.5.6 Experiment Validation 171
7.5.7 Experiment Conclusion 172
7.6 Evaluation of the Collective Intelligence Framework 172
7.6.1 Functionalities Achieved 172
7.6.2 Resources Utilised 174
7.6.3 Limitations of the Framework 174
7.7 Concluding Remarks 175
Chapter 8 Conclusion and Further Work 177
8.1 Overview 177
8.2 Review of Objectives 177
8.2.1 Characterising the major biomolecular self-organising mechanisms .. 177
8.2.2 Using a bottom-up integrative approach to model the intracellular organisational behaviour 180
8.2.3 Developing a Collective Intelligence based cell modelling and simulation environment to conduct analysis studies 182
8.3 Thesis Review 184
8.4 Contribution to Knowledge 189
8.5 Further Work 191

Appendix A Publications 194
Appendix B The Dependencies Between Metabolites and Reactions 197
Appendix C The Epigenetic Mechanisms 199
Appendix D The Features of Repast Simphony 201
Appendix E The Results of Simulation Experiments 203
Appendix F The Calculation of Experimental Parameters 210
Bibliography 215

LIST OF FIGURES

Figure 1.1: The organisational space of life 3
Figure 1.2: The U model 14
Figure 2.1: The autocatalytic cycles that traverse across gene space to metabolite space 21
Figure 2.2: The outcome of the law of sufficiency 30
Figure 2.3: The nature of a biomolecule and the biological cell.. 31
Figure 2.4: The formation of robustness and its associated biological equilibrium. 33
Figure 2.5: The role of metabolism in cellular homeostasis 36
Figure 2.6: The objective hierarchy forming nested organisation in biological systems. 39
Figure 4.1: Relationship between spatial and temporal issues in situated systems. 92
Figure 4.2: Complexity barrier between macroscopic cellular and intercellular processes that emerge from basic biomolecular activities. 96
Figure 4.3: Class diagram of biomolecules 98
Figure 4.4: Typical functional states of a biomolecule 99
Figure 4.5: Typical event durations of some actions in biological systems 101
Figure 4.6: Typical energy content of some molecules and bonds in biological systems 101
Figure 4.7: Fraction of particles that will react at any given temperature depends upon how many particles have at least the energy of Level e in the presence of an enzyme or Level t in the absence of an enzyme. ... 107
Figure 4.8: An elementary enzyme catalysed reaction. 110
Figure 5.1: The core of the SwarmCell model representing the physical space 114
Figure 5.2: Analysis of performances 122
Figure 6.1: Repast S components utilised for the SwarmCell model. 132
Figure 6.2: Structure of the SwarmCell environment. 134
Figure 6.3: Templates for native biomolecules and resources as present in model.score file. 135
Figure 6.4: Physical Space during execution. 136

Figure 6.5: Pathway Layer during execution. 136
Figure 6.6: Cell Layer during execution. 137
Figure 6.7: Scalar field space during execution. 138
Figure 6.8: SwarmCell Interface. 139
Figure 6.9: Experimental setup. 140
Figure 7.1: The results of the half-life experiment 147
Figure 7.2: Time require to attain equilibrium with respect to enzyme abundance, processing time and the initial equilibrium phase of the chemical system. 153
Figure 7.3: The global equilibrium constants attained by the different combinations of factor levels. 154
Figure 7.4: The effects of change in factor levels with respect to response time 155
Figure 7.5: The analysis of sensitivity of factor level changes, in reaching chemical equilibrium 156
Figure 7.6: The interaction effects between the abundance of enzyme (EA) vs turnover cycle of the enzyme (ET) and the initial equilibrium phase of the chemical system (EP). 156
Figure 7.7: The interaction effects between the initial equilibrium phase of the chemical system. (EP) vs the abundance of enzyme (EA) and turnover cycle of the enzyme (ET). 157
Figure 7.8: The interaction effects between the turnover cycle of the enzyme (ET) vs the initial equilibrium phase of the chemical system (EP) and the abundance of enzyme (EA). 157
Figure 7.9: Pathway for the biosynthesis of aromatic amino acids in *E. coli* 160
Figure 7.10: The performance of Chorismate Mutase with respect to abundance and processing time of enzyme Chorismate Mutase and Anthrinilate Synthase. 165
Figure 7.11: The consumption of Chorismate by the two competing tasks involving Chorismate Mutase and Anthrinilate Synthase 166
Figure 7.12: The effects of change in factor levels with respect to the performance of Chorismate Mutase 167
Figure 7.13: The analysis of sensitivity of the performance of Chorismate mutase to the above factors 167

Figure 7.14: The interaction effects between the abundance of Chorismate Mutase (CMA) vs abundance of Anthrinilate Synthase (ASA), Turnover Cycle of Chorismate Mutase (CMT) and Turnover Cycle of Anthrinilate Synthase (AST) 168
Figure 7.15: The interaction effects between the abundance of Anthrinilate Synthase (ASA) vs abundance of Chorismate Mutase (CMA), Turnover Cycle of Chorismate Mutase (CMT) and Turnover Cycle of Anthrinilate Synthase (AST) 169
Figure 7.16: The interaction effects between the Turnover Cycle of Chorismate Mutase (CMT) vs abundance of Chorismate Mutase (CMA), abundance of Anthrinilate Synthase (ASA) and Turnover Cycle of Anthrinilate Synthase (AST) 169
Figure 7.17: The interaction effects between the Turnover Cycle of Anthrinilate Synthase (AST) vs abundance of Chorismate Mutase (CMA), abundance of Anthrinilate Synthase (ASA) and Turnover Cycle of Chorismate Mutase (CMT) 170

LIST OF TABLES

Table 2.1: The amplification and reduction of activity via feedback mechanisms 29
Table 2.2: Half-lives of cytosolic proteins which depend on the nature of their amino-terminal residue 35
Table 2.3: The primary levels of biological organisation, the units at each level, and the properties of the units 43
Table 3.1: The comparison between empirical and mechanistic model development process types 54
Table 3.2: The comparison between mathematical and computational model 57
Table 3.3: The distinctions between mass action kinetic models and the stochastic kinetic models 59
Table 3.4: Comparison of modelling formalisms based on model requirements 71
Table 3.5: Comparison between objects and agents 72
Table 3.6: Comparison of open source toolkits 81
Table 4.1: Distinguishing features between existing swarm intelligence techniques and collective behaviour of biomolecules 91
Table 4.2: Abundance of intracellular compartments and biomolecules in cells 101
Table 4.3: Lengths of cellular components 102
Table 4.4: Macroscopic and microscopic views in cellular thermodynamics 108
Table 5.1: State variables and types of biomolecules 116
Table 5.2: Diffusion Coefficients in aqueous solutions 124
Table 5.3: Distance travelled during various time steps based on diffusion constant 125
Table 5.4: Time required travelling a particular distance based on diffusion constant 125
Table 5.5: Maximum turnover number and the calculated turnover cycle of some enzymes 127
Table 5.6: Half-life, average lifetime and degradation constants for proteins containing various amino-terminal residues. 128
Table 5.7: Probabilities of forming a protein of various lengths with no errors 130

Table 6.1: Relationship between components described in the model specification, Repast S and model implementation 133
Table 7.1: Design specification of the experiment showing the design matrix for simulating biomolecular degradation 145
Table 7.2: The constants and their values used for simulating biomolecular degradation 146
Table 7.3: Gibbs free energy of formation of metabolites involved in the reaction 1 of the TCA cycle 149
Table 7.4: Variables and their value ranges used for simulating the biochemical reaction 151
Table 7.5: Design specification of the experiment showing the design matrix for simulating biochemical reaction 152
Table 7.6: Variables and their value ranges used for simulating biochemical reaction 153
Table 7.7: The variables and their value ranges used for simulation competition and cooperation in metabolic pathway 162
Table 7.8: Design specification of the experiment showing the design matrix for simulating competition and cooperation in metabolic pathway 163
Table 7.9: Variables and their value ranges used for simulation competition and cooperation in metabolic pathway 164

"The problem of biology is not to stand aghast at the complexity but to conquer it."

- Sydney Brenner

Chapter 1

Introduction

"For systems biology to mature into a solid scientific discipline, there must be a solid theoretical and methodological foundation"

- Hiroaki Kitano

1.1 Overview

This chapter describes the research project's outline. The chapter begins with the background and motivation behind the research and analyses the work required. The section on relevant work defines the scope of the project and looks at some corresponding paradigms. The aim, objectives, research questions and the hypothesis are formulated based on the scope of the project. A brief justification of the adopted approach and methodology used is described in Sections 1.7 and 1.8. Section 1.9 lists the publications that emerged as result of the project. The scope and organisation of the thesis is described in the final section.

1.2 Background

The transition of molecular biology into systems biology was facilitated by the development of various high-throughput technologies, representing the 'biology' root to systems biology and formal analysis methodologies, representing the 'systems' root to systems biology (Westerhoff and Palsson 2004). The huge volumes of data generated mainly by reductionist approaches, led to a rapid growth in the field of bioinformatics. Bioinformatics developed the computational tools to provide solutions for research problems that biologists encounter. Although the scope of the application was mostly based on pattern recognition approaches, it was realised that a more formal and mechanistic framework was required for the systematic analysis of

multiple 'omics' data types. This led to the development of genome-scale *in silico* modelling to analyse the systemic properties of cellular functions (Westerhoff and Palsson 2004).

While biomolecular science studies individual biomolecules with the aim of revealing how molecules function, systems biology, aims to predict the consequences of the particular molecular mechanism on the whole organism. However, molecular sciences have become one of the most effective branches of science, by utilising reductionist approaches to characterise the molecular basis of life for a diverse number of organisms. However understanding the molecular constituents is necessary but not sufficient for system level understanding (Bork 2005; Dubitzky 2006), and a quantitative reconstruction of the system with its constituents, is required. Systems biology utilise reconstruction approaches to study system wide phenomena. One of the aims of systems biology is to understand biological phenomena, which emerge from the complex interactions that occur within and between the levels of biological organisation strata. Hence, by determining how functions arise due to the dynamic interactions of constituents, systems biology addresses the missing links between molecules and physiology (Bruggeman and Westerhoff 2007). Considering this ambitious goal, systems biology is still in its infancy, and the success of this new discipline will depend on delivering meaningful results by integrating methods and approaches developed in other disciplines.

Systems biology studies are mainly conducted in two forms. Studies based on incompletely characterised cellular systems mostly take the form of a top-down approach, to identify the correlations between the various variables of the systems. Although this approach emphasises inductive discovery science, this seldom leads to molecular knowledge. (Bruggeman and Westerhoff 2007) consider that these studies must either transform into or associate with more mechanism based approaches adopted by bottom-up systems biology studies. However the aim of this approach should not simply be to deduce functional phenomenon based on certain underlying molecular interactions, rather it is also important to demonstrate that the phenomenon really occurs. Hence, a significant effort has to be placed on experimental determination of the actual interaction parameters and a precise modelling of the phenomenon. However, bottom-up systems biology cannot tolerate unknown factors and, thus, will need to integrate with top-down, genome-wide 'omics' systems

biology approaches to ensure completeness. Since systems biology is a science (Westerhoff and Alberghina 2005), it should also aim to discover general principles, which relate to all aspects of cellular organisation. This approach to systems biology could lead to substantial fundamental insights into the principles that underlie biology (Bruggeman and Westerhoff 2007).

1.3 Motivation

One of the key challenges in systems biology is the modelling of cellular systems. The president of the International Society for Systems Biology Dr. Hiroaki Kitano has recognised the importance of developing these models as a fundamental intermediate step to achieving *in silico* biological simulations (Kitano 2002c; Kitano 2002a; Kitano 2002b; Kitano 2004a; Kitano 2004b; Kitano 2005; Kitano 2006; Kitano 2007). These advanced models should incorporate molecular information, facilitating the prediction of cellular behaviour under various conditions, which are fundamental to revealing the cellular level characteristics and principles of cellular functions (Kitano 2007). Some of the key cellular characteristics include robustness, adaptability and efficacy. Kitano (Kitano 2007) has identified the importance of mechanistic principles and constraints in biological adaptation (illustrated in Figure 1.1), which are in line with the author's direction of investigation.

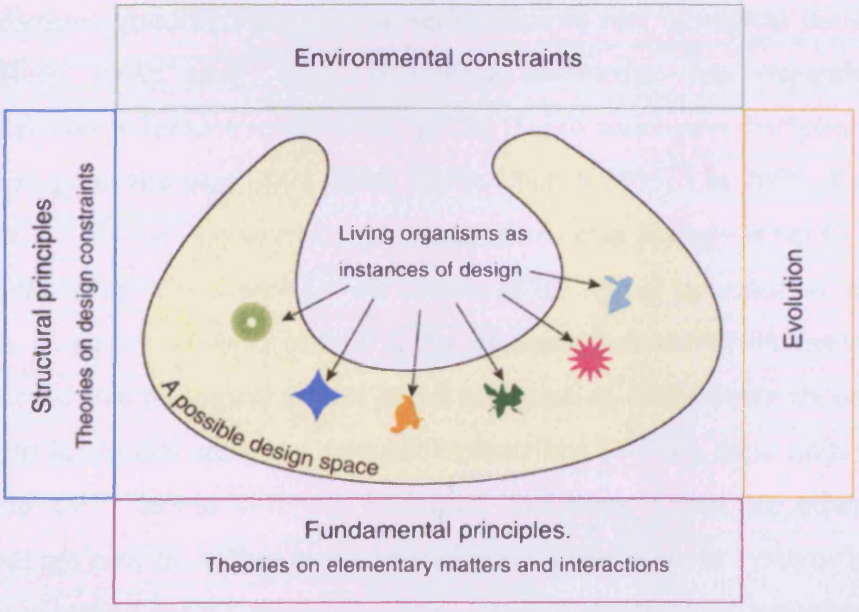


Figure 1.1: The organisational space of life (adapted from Kitano 2007)

Biological systems have been and continue to be organised by evolutionary principles. While fundamental (i.e. laws relevant to physical systems) and organisational (i.e. laws relevant to biological systems) principles act as internal constraints in biological systems, evolution acts to guide towards an organisation of biological systems based on environmental constraints or pressures. Feasible organisations are only producible within the constraints of fundamental and organisational principles. It has been acknowledged (Kitano 2007) that the success of systems biology depends not only on studies based on a specific instance of life, but also on studies based on the principles governing the entire possible diversity within observed viable biological systems.

The current data driven modelling approaches focus on specific instances or examples of life, whereas the field of artificial life is engaged in unravelling the principles governing the entire organisational space of life (Goldstein, Husbands et al. 2010). Based on the classical definition, "*artificial life is a field of study devoted to understanding life by attempting to abstract the fundamental dynamical principles underlying biological phenomena, and recreating these dynamics in other physical media - such as computers - making them accessible to new kinds of experimental manipulation and testing*" (Langton 1992), it was postulated that artificial life and complex systems research would be the driving force for understanding life via theoretical methods (Langton 1988). However, initial endeavours had little or no impact on the biological community, because these theories were unable to provide useful predictions, guiding principles or verification of real biological issues (Kitano 2002b). Since those early days, biological knowledge has expanded at an extraordinary rate in these research areas, giving rise to entire new disciplines, such as systems biology in the year 2005 (Bork 2005; Church 2005; Liu 2005; Kahlem and Birney 2006). Although a consensus definition of systems biology is yet to emerge, it has been defined as "*the search for the syntax of biological information, that is, the study of the dynamic networks of interacting biological elements*" (Aebersold 2005). A major part of this biological syntax is the organisation of elements encoded by the genome, into functional units and dynamic interactions between these units to control and perform their various complex biological functions. There are other areas of research that are now providing new and exciting perspectives for systems biology by introducing a varied set of principles. For example the field of complex adaptive systems studies the adaptability of systems, the field of cybernetics studies control

and communication in systems (Heylighen and Joslyn 2001), the field of natural computing aims to understand nature from the perspective of information processing (Lila and Grzegorz 2008). The development of a general theoretical framework and its "*integration into biological research, thus represents an exciting branch of systems biology*" (Aebersold 2005).

These definitions imply that systems biology and artificial life share a common objective: "*A principled and comprehensive understanding of living systems*" (Kim and Eils 2008). Both interdisciplinary fields utilise formalisms to model and analyse biological systems. Although the use of models is complimentary, systems biology focuses on analysing and understanding experimental data using fairly generic modelling techniques, artificial life "*considers rather elaborate and specific computational and other formal models as objects of experimentation, aiming to understand general biological features that are not necessarily represented by quantitative data*" (Kim and Eils 2008). Hence, the modelling philosophy of artificial life differs considerably from systems biology, as it studies not only "life as we know it" but also "life as it might be".

As the models of biological systems increase in complexity in the future, the two fields are expected to considerably overlap as the methodologies to model the biological organisation strata are combined with advanced techniques for empirical model inference. Further, it is also predicted that as model complexity increases the relative amount of molecular biology data and associated knowledge to validate model inference methods will decrease. Hence models that incorporate multiple levels of biological organisation will become increasingly important as a source of realistic synthetic test data (Kim and Eils 2008).

One of the exciting fields of artificial life is developing biologically inspired approaches by observing biological phenomena. Collective Intelligence is a social phenomenon and is defined as the group of individuals doing things collectively that seem intelligent (Malone 2006). It is a shared intelligence that emerges from cooperation, competition and coordination of many individuals. Swarm Intelligence refers to the phenomena of a system of spatially distributed entities coordinating their actions in a decentralised and self-organising manner, so as to exhibit intelligent collective behaviour in local interactions. The concepts of self-organisation and

emergence underlie swarming, and these systems are inherently adaptive, robust, flexible, stochastic and concurrent.

1.4 Research Required

One of the grand challenges in biology is the formulation of a unified fundamental theory governing biological systems (Kitano 2007) which may lead to the formulation of laws for life. To achieve this goal it is essential to resolve the gap between the level of description used in thermodynamics and other basic physical sciences and the abstraction levels (i.e. interactions within biological organisation strata) used to define concepts such as robustness, adaptability, efficacy and other system level properties in biological systems.

Kitano frequently emphasises robustness, which is a fundamental characteristic of biological systems (Kitano 2004a). This issue is encompassed by "*for systems biology to mature into a solid scientific discipline there must be a solid theoretical and methodological foundation*" (Kitano 2007). Systems biology is widely accepted as encompassing both computational modelling and simulation (Jones 2008). Computer simulation is an effective way of visualising complex dynamics, intrinsic to biological systems and exploring the validity of assumptions that form the foundations of our understanding of cellular processes. Kitano (Kitano 2007) stated that "*The scientific goal of systems biology is not merely to create precision models of cells and organs, but also to discover fundamental and structural principles behind biological systems that define the possible design space of life*". He expanded on this point by identifying the importance of understanding the fundamental and organisational theories that provide deeper insights into the governing principles that underpin complex evolvable systems. Of the numerous challenges that need to be overcome, a key issue is how to represent system level properties such as robustness, adaptability and efficacy, so that we can study the effects of perturbations at molecular level on the performances of a biological system. A theory of biological robustness should be extended to deal with organisational level adaptation. This will require defining the parameters that govern biological organisation and development of a comprehensive set of innovative computational methods to model such characteristics. This work may bring abstract theory to practical use by identifying specific constraints governing the organisation

of biological systems. Kitano stated that as theoretical research progresses, the ability to predict and reverse engineer the organisation of biological systems will advance and finally, the theory will have to be integrated with thermodynamics (Kitano 2007). Further, fields such as nonequilibrium dissipative systems, nonlinear dynamics and chaos theory, are yet to be extended to the principles of living systems. Nevertheless, these theories still do not consider the diversity and organised nature of biological systems, nor do they consider the major challenge, when attempting to bridge this gap of selection through evolution (Kitano 2007).

Self-organisation is considered to be one of the mechanisms of biological evolution (Kauffman 1993). These concepts that were initially developed in chemistry and physics are now beginning to be applied to the organisation of the living cell. Studying self-organisation processes in cell biology enforces a focus on principles and collective behaviours of the biomolecules that underlies the emergence of coherent dynamic cell shapes and functions (Karsenti 2008). Eric Karsenti, Head of the Cell Biology and Biophysics Unit, EMBL stated that a major difficulty facing biology, concerns the origin of structures and their associated functions. This has been an ongoing question in developmental biology, and related questions such as the origin of intracellular structures and their associated purpose must also be addressed at the cellular level (Karsenti 2008).

It has been declared that "whole cell simulation is a grand challenge of the 21st century" (Tomita 2001). To this end there are numerous groups attempting to build cell simulation environments with the aim of addressing various intracellular activities at different modelling resolutions. These resolutions range from the level of atoms to molecules and cells, which represent the micro, meso and macro modelling approaches. Macro modelling approaches are population based models and assume the cell is a well mixed environment of biomolecules, this is in direct opposition to current thinking, which has moved beyond the simple concept of a cell as an unstructured mixture (Andersen 2004). Mesoscopic approaches model cells at a molecular resolution, whilst microscopic approaches, such as molecular dynamics model molecular behaviour at the atomic level. A detailed discussion of these approaches is presented in Chapter 3.

1.5 Relevant Work

The author's direction of investigation was confined to the principles governing the functional organisation of biological cells, especially focusing on the adaptive dynamics between the levels of molecular resolution and cellular resolution. Moreover, the investigation is interconnected to the fields of Systems Biology, Complex Systems and Natural Computing. The model is focused on molecular resolution, especially modelling biomolecular activities in space and time. There are various levels of simulation aiming to model intracellular activities.

1.5.1 Corresponding Paradigms

Many cell simulation environments are emerging, aiming to model the whole or parts of the cell, based on various mechanistic principles. The population based mass action kinetic modelling approach is adopted by simulation environments such as the E-Cell (E-Cell Project 2009) and the Virtual Cell (NRCAM 2009) being developed by the National Resource for Cell Analysis and Modelling. The population based stochastic kinetic modelling approach is adopted by Agent Cell (Emonet, Macal et al. 2005) being developed collaboratively by the Institute for Biophysical Dynamics, the James Frank Institute and the Centre for Complex Adaptive Agent Systems Simulation, Argon National Laboratory, SmartCell was developed in the Serrano Laboratory of the Heidelberg Laboratory of EMBL, and MesoRD was developed by Uppsala University, Sweden. A particle based stochastic modelling approach is adopted by simulation environments, such as M-Cell (Stiles and Bartol 2001) being developed by the Salk Institute for Biological Studies, Cell++ being developed by Parkinsons Laboratory of Computational Systems Biology, CyberCell (Sundararaj, Guo et al. 2004) is under development at the Institute for Biomolecular Sciences, University of Alberta, ChemCell (Plimpton and Slepoy 2005) is being developed at Sandia National Laboratories, Smoldyn is being developed by the Molecular Science Institute, Berkley and GridCell (Boulianne, Al Assaad et al. 2008) is being developed by the Integrated Microsystems Laboratory, McGill University. Moreover, the WebCell (Lee, Park et al. 2007) and Silicon Cell (Snoep, Bruggeman et al. 2006) are basically collaborative environments that share cell simulation resources, such as tools and data for modelling cells. Microsoft Research, Cambridge Laboratory is venturing into developing a fully programmable *in silico* cell, by using pure computational

formalisms rather than mathematical formalisms. A detailed review of these approaches is presented in Chapter 3.

The research closest to that presented in this dissertation, is that of particle based simulations, which focuses on molecular resolution, and represents spatial information by modelling biological entities such as functional products as individual objects. This modelling is centred on a whole molecule approach.

1.6 Aim, Objectives and Hypothesis of the Research

1.6.1 Aim of the Research

The aim of the research is to investigate systems biology approaches to representing biological complexity from molecules to cells and developing computational approaches to bring abstract theories to practical use.

1.6.2 Objectives of the Research

The objectives are:

- **Characterising the major biomolecular self-organising mechanisms**

This will require eliciting novel cellular information processing strategies at the molecular level, by focusing on information/signal dissemination and transformation. By delivering this objective we intend to address questions relating to cellular self-maintenance; drivers for self-organisation and Collective Intelligence, and the effects of limitations of molecular activities on the intracellular organisational behaviour.

- **Using a bottom-up integrative approach to model the intracellular organisational behaviour**

By delivering this objective we intend to address questions relating to representing collective behaviour of biomolecules *in silico* to model cellular level phenomena; a suitable model development process; the modelling approach that can represent intracellular organisational behaviour; to study the emergence of cellular level characteristics such as adaptation, robustness and efficacy; functionally uniting the

activities of functional products to form collectives and the criteria for identifying functional units to represent intracellular tasks/objectives.

- **Developing a Collective Intelligence based cell modelling and simulation environment to conduct analysis studies**

The purpose of analysis studies is to learn about and get a better understanding of cellular phenomena. By delivering this objective, we intend to address questions relating to the required molecular level information to model biomolecules and their interactions; approaches to analyse the dynamics of biomolecular interaction; measuring and controlling organisational behaviour within a biological cell, and measurement of cellular performance.

- **Developing biomolecular inspired adaptive algorithms to conduct design studies**

The purpose of design studies is problem solving, or seeking solutions to problems found in biological cells, namely remedies for pathological phases, or finding solutions, which engineer biological systems with new requirements. In delivering this objective we intend to address questions relating to engineering a biological cell as an *in silico* swarming system; the construction and deconstruction of tasks between basic molecular activities to complex cellular activities; the representation of communication barriers amongst biomolecules; the representation of the extremely concurrent nature of biomolecular interactions; incorporation of forms of positive and negative feedback and modelling the amplification of fluctuations that give rise to solutions.

1.6.3 Hypothesis

The research hypothesis is:
A Collective Intelligence based cell modelling framework, which is able to adapt to multiple task/objectives concurrently, in the face of perturbation and uncertainty, would mechanistically represent the diverse intracellular performances/functions, and capture the adaptive dynamics of a biological cell.

This hypothesis is based on the following, that:

- Swarm systems are able to pursue multiple lower level tasks concurrently which interact to create system level functions
- Swarm systems are able to adapt their behaviour and maintain performance/functions to meet their goals in the face of perturbation and uncertainty
- Cellular adaptation in physiological timescale occurs due to self-organisation which is the underlying principle of swarm systems
- Coordination, cooperation and competition are some of the hallmarks of swarm systems which can also be observed in the collective behaviour of biomolecules
- Intelligent behaviour emerges out of the activities of entities in space and time.
- Biomolecular activities are transformed into performances, which manifest as cellular level functions
- Functional units which define the tasks, are subjected to adaptive pressure

1.7 Research Direction

This section justifies the direction of research in terms of adapted systems biology approach used for the study, modelling methodologies, mechanistic principles, and the formalism and framework used for the development of the approach. These are more fully discussed in Chapter 3.

Biological adaptations occur in physiological, developmental and evolutionary timescales. However the scope of research has been in modelling multi-scale adaptive dynamics from molecules to cell at the physiological timescale, where biomolecular interactions significantly contribute to this process. A mechanistic model development approach is adopted to model the diverse behaviour of biomolecular species and to provide mechanistic explanations of cellular phenomena. Although different mechanistic principles have been developed to describe different aspects of observed natural phenomena, they have limitations in their applicability to represent biological phenomena. A combination of mechanistic principles governing the organisation of biomolecular activities will be required to describe biological phenomena. A detailed discussion of these principles is presented in Chapter 3. The mechanistic principles

are primarily based on systems theory (especially complex adaptive systems), where principles of self-organisation are implemented, using the logic of Collective Intelligence to model adaptive dynamics from the bottom-up in a physiological timescale. A detailed discussion of this mechanism is presented in Chapter 4. The aim of this approach is to demonstrate how principles and properties of Collective Intelligence can address how biological cells dynamically adapt to multiple objectives concurrently, facilitated by constituent biomolecular activities. The objectives of biological systems are constantly evolving, due to the ever changing demands of their environment. Biological systems meet these demands by pursuing these objectives, aided by their constituents, giving rise to biological functions. Tasks emerge, when pursuing these objectives, which are concurrent and mutually dependent. The main contributors to concurrency in biochemical activities are the specialised activities performed by redundant members of diverse biomolecular species. Complex global tasks of the cell are formed from diverse basic tasks of intracellular groups of biomolecular species. Cellular functions are quantified in terms of the performances of solutions, which are constructed/deconstructed in terms of the objectives/tasks of the cell. Categorising the collective behaviour of functional products in terms of objectives/tasks can deconstruct the global objectives/tasks of a cell into basic tasks required to pursue them. A detailed discussion of this process of simplifying cellular complexity due to diverse biomolecular interactions, is presented in Chapter 2.

A cellular environment represents both biomolecules and their activities which contribute to the self * properties of the cell. These activities cause direct and indirect influences amongst various species of native biomolecules, which facilitate in self regulation of cellular processes. Agent based formalism is used in the wider framework of Collective Intelligence to model self-organisation and the emergence that occurs due to diverse biomolecular activities. Further, this approach facilitates analysis of global effects of changes in behavioural rules imposed on diverse biomolecular species, where the effects of rules are amplified due to redundant members of biomolecular species. The representation of agent based formalism at the level of molecular resolution also addresses the heterogeneous nature of the cellular environment and the existence of very low numbers of some functional products. Since the organisational behaviour within a cell cannot be directly observed or empirically measured, it requires a simulation framework to be built that can represent

native biomolecules, capture the results of their activities and provide a means to evaluate these results. The analysis of cellular behaviour should be based on the chemical activities of molecules rather than their abundance, since activities provide an accurate description of a chemical system, in which performances of functional products are analysed based on their level of activities.

1.8 Research Methodology

A bottom-up systems biology study is conducted, which adopts a mechanism based approach to deductive reasoning. The methodology adopted for this research is the U-model approach (see Figure 1.2), which has been used for significant number of studies on Collective Intelligence (Schut 2007). In line with the project's aim, a suitable theory or mechanistic principle was formulated to explain the multi-scale adaptive dynamics from molecules to cell. The objectives and research questions were formulated to meet the aim of the project. A working hypothesis was formulated to test the mechanistic principle. Further observations were collected from the literature to address the hypothesis. The structure of the model was developed from the proposed mechanistic principle. Appropriate model parameters, which are consequential to the observations, were identified. Appropriate data sources were identified to represent the model parameters. These included independent variables, that were altered for different series and variables that remained constant throughout a simulation experiment. The dependent variables represent the data points for the simulation, which were to be used for analysis of the experimental results.

The model requirements were listed to develop a mechanist model. Preparation for the experiments began with the model specification. The specification describes the assumptions made in advance and design choices made progressively as the research proceeded. These are described in detail in Chapter 4 and Chapter 5. The implementation was based on the model specification and the identification of a suitable simulation package for the simulation study. A number of experiments were conducted iteratively. This includes the stages of experimental design and setup, performing the experiment, obtaining results and analysing the results.

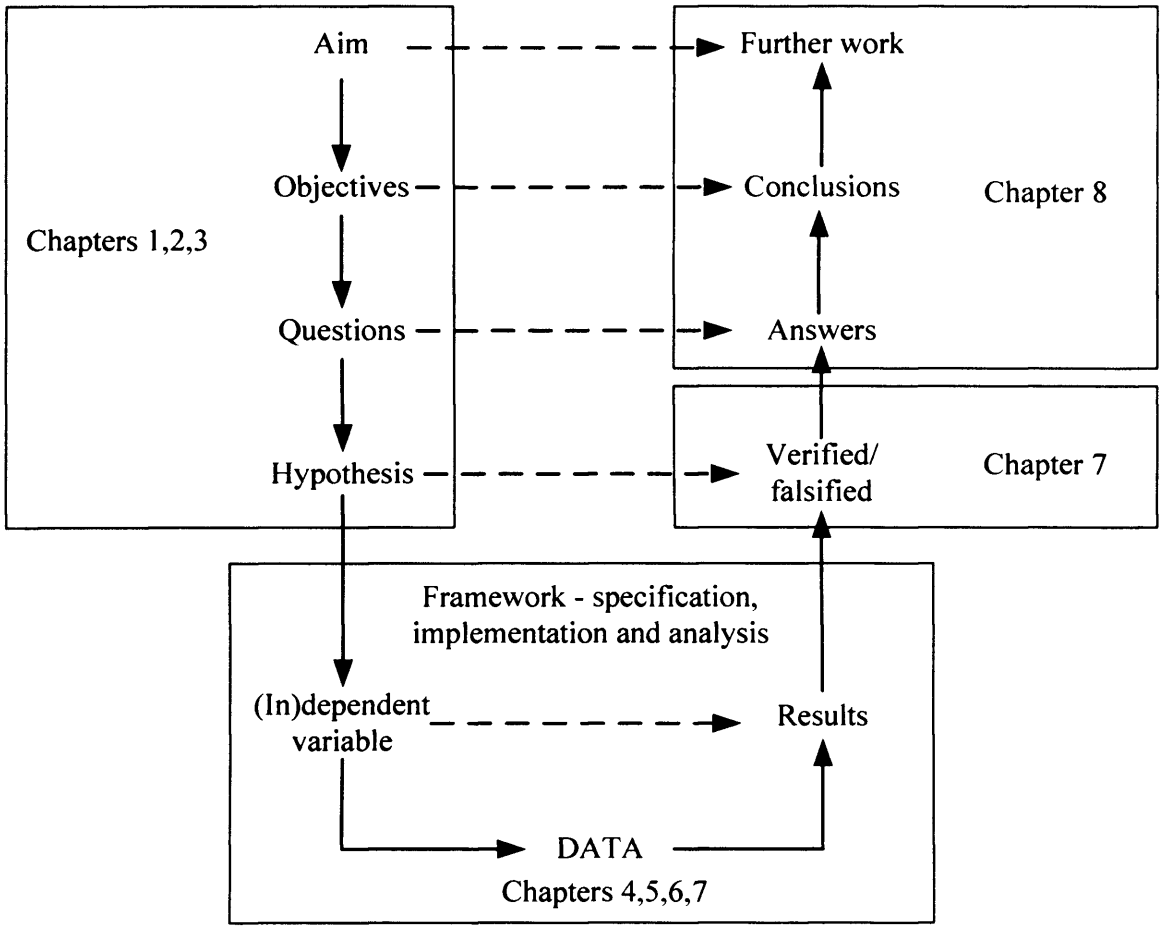


Figure 1.2: The U model

1.9 Research Contributions

The research contribution is based on determining the fundamental and organisational principles behind biological systems that define a possible design space of biological cells, and applying these principles to build mechanistic models of biological phenomena. The novelty of the thesis and its major contribution to knowledge is based on defining cellular functions in the context of a multi-objective topology and implementing this principle, as an *in silico* model, to study a performances of intracellular functions by measuring activities of diverse species of functional products. Further, this approach represents biological adaptation at the biochemical level – a feature that network topology is unable to represent. The major contribution to computing is identifying a novel Collective Intelligence approach based on

information processing strategies of biomolecules and utilising it for modelling intracellular activities. The contributions include:

1. Use of an agent based formalism in the wider framework of Collective Intelligence, which considers principles and properties of self-organising processes to determine fundamental and organisational principles of a biological cell (see Chapters 2 and 4).
2. Showing the significance of analysing the biomolecular activities rather than their abundance, as this provides an accurate description of a biochemical system. The biomolecular organisational behaviour is analysed by quantifying cellular functions in terms of measuring performance of objectives/tasks formed by the activities of diverse functional products (see Sections 3.5.2.3 and 4.4).
3. Providing an environment to analyse organisational behaviour within a cell, that cannot be directly observed or empirically measured. This is achieved by using a simulation framework to represent native biomolecules, capturing results of their activities and providing a way to evaluate these results (Periyasamy, Gray et al. 2008a; Periyasamy, Kille et al. 2008) (see Chapters 5 and 6).
4. Showing that cells have adopted a unique strategy to continuously realise their objectives/tasks or adaptive requirements (self-awareness) by eliminating obsolete information and generating new information in their internal organisation. The tendency for biomolecular degradation by means of random or regulated process and collective autocatalysis provides an ideal reinforcement adaptive mechanism for a cell (Periyasamy, Gray et al. 2008b) (see Section 2.5).
5. Implementing a novel system-theoretic approach to molecular systems biology by utilising biomolecular inspired multi-objective strategies from a Collective Intelligence perspective to capture higher level performances of a cell (Periyasamy, Gray et al. 2009) (see Section 7.5).
6. Using novel criteria for modularising interactions among functional products, which are based on performance interactions, which emerge from competition and cooperation among the functional products (Periyasamy, Gray et al.

2009). Direct and inverse performance interactions can reveal organisation of basic objectives/tasks into complex global tasks, in order to construct/deconstruct tasks between molecular resolution and cellular resolution (see Chapters 2 and 5).

Titles and abstracts of presented and published work are listed in Appendix A.

1.10 The Scope and Organisation of this Thesis

The scope of the thesis is defined from different perspectives. As stated in Section 1.4, the thesis is confined to principles governing organisational space of biological cells. The novel framework proposed offers great potential for modelling multi-scale adaptive dynamics from molecules to cell in a physiological timescale. The purpose of the framework is to model how a biological cell is organised to adapt and use this to understand more about the transitions between healthy and pathological phases of biological systems. The thesis is focused on modelling biochemical systems, based on principles governing an organisational space rather than developing a data driven modelling approach. The aim of the thesis is to investigate approaches, representing biological complexity from molecules to cells, and developing computational approaches to bring abstract theories to practical use. This thesis argues that a system theoretic approach based on Collective Intelligence, where the underlying concepts of self-organisation and emergence, underlie the approach is well suited to representing the multi-scale and multi-objective/task nature of a biological cell from the bottom-up. It proposes a Collective Intelligence based cell modelling and simulation environment, which can be used to conduct analysis studies on the collective behaviour of biomolecules driven by their activities. These activities are organised into a hierarchy of tasks, where basic tasks contribute to the formation of complex and mutually dependent global tasks of a cell, which ultimately represent cellular functionalities.

The thesis is organised according to the progressive development of the Collective Intelligence framework, a novel approach to modelling and simulating a biological cell using the principles of Collective Intelligence. Its chapters are:

Chapter 2: Multi-Scale Adaptive Dynamics from Molecules to Cell. This defines the problem of modelling the adaptive dynamics of biological cells and evaluates the

requirements to address these problems. The heart of the problem is to understand how biological systems are organised, to realise the transitions between healthy and pathological phases, and adapt accordingly. The thesis is focussed on adaptations that occur within physiological timescales, where biomolecular activities contributing to functional organisation, play a key role within a cell's lifecycle.

Chapter 3: Modelling Biological Phenomena. This specifies the functional and non-functional requirements needed to build mechanistic models based on the first principles of addressing the problems articulated in Chapter 2. A critical review is conducted of related work, with respect to modelling from a molecular resolution to cellular resolution and addressing the feasibility of achieving the specified requirements based on available resources and technology.

Chapter 4: Representation of Biomolecules and their Activities within an *In silico* Environment. This describes how swarming can address issues raised in Chapter 2. The principles and properties of Collective Intelligence are addressed in the context of collective behaviour of biomolecules. *In silico* representations of native biomolecules and their activities, which constitute a cellular environment are discussed.

Chapter 5: A Collective Intelligence Approach to Modelling Intelligent Cellular Organisation. This provides a model specification based on the problem definition and model requirements discussed in Chapters 2 and 3, respectively. The model specification provides an overview of the model's focus, resolution and complexity. The design concepts describe the general concepts underlying design of the model.

Chapter 6: Swarm Based Cell Modelling and Simulation Environment. This describes the general implementation of the model specification, which is used to setup and run various simulation experiments based on specific scenarios.

Chapter 7: Model Evaluation by Simulating Biological Phenomena. This evaluates the Collective Intelligence framework by conducting a series of simulation experiments. These experiments model the physical and biological constraints, involved in the organisational behaviour within biological cells, which affect their adaptive processes. Based on the results of the experiments, the validity of the simulation experiment and framework is justified.

Chapter 8: Conclusion and Further Work. This provides answers to the investigated research questions, based on the objectives and draws conclusions based on the findings of the research.

1.11 Concluding Remarks

The scope of systems biology consists of top-down systems biology studies, bottom-up systems biology studies and discovering general principles of biological systems. It has been acknowledged that the success of systems biology depends not only on studies based on a specific instance of life, but also on studies based on principles governing the entire organisational space of life. Hence, modelling of adaptive dynamics is identified as an essential requirement to understand the organisational space of biological systems. This requires the development of advanced models with molecular information that facilitates the prediction of cellular behaviour under various conditions. This is needed to reveal the cellular level characteristics and the underlying principles of cellular functions. The aim of the thesis is to investigate systems biology approaches to representing biological complexity from molecules to cells and developing computational approaches to bring abstract theories to practical use. The conclusion is to adapt a bottom-up systems biology approach and utilise a mechanistic model development process to develop a computational model, using an agent based formalism in the wider framework of Collective Intelligence to represent the intracellular behavioural/functional organisation. The research contribution is determining the fundamental and organisational principles behind biological systems that define the possible design space of biological cells and applying these principles to build mechanistic models of biological phenomena.

The next chapter investigates the characteristics and properties of biological systems and the challenges in modelling the adaptive dynamics of biological cells by describing biological complexity from molecules to cell.

Chapter 2

Multi-Scale Adaptive Dynamics from Molecules to Cell

"For systems biology to be truly successful, not only studies on specific instances of life, but also studies on principles governing the entire design space are required."

- Hiroaki Kitano

2.1 Overview

The aim of this chapter is to characterise mechanisms and factors that are consequential to adaptive dynamics of biological cells and determine requirements to address these problems. Section 2.2 describes biomolecules and their activities, constituting a cellular environment. The significance of the problem is to gain an understanding of how biological systems are organised to realise transitions between healthy and pathological phases, and adapt accordingly. Although biological adaptations occur in physiological, developmental and evolutionary timescales, the thesis is focussed on adaptations that occur within physiological timescales, where biomolecular activities contributing to functional organisation, play a key role within a cell's lifecycle. Cellular activities are hierarchically organised into various basic tasks, which merge to form the complex and greater tasks of a cell. Section 2.3 evaluates the mechanisms of biological adaptation and specifies two categories of goals/objectives, which define these tasks driving the adaptive process. Section 2.4 describes the multi-objective nature of biological systems and the constraints involved in pursuing these objectives. The formation of intelligent cellular organisation from the collective behaviour of biomolecules is discussed. Section 2.5 describes the hierarchical nature of biological systems by simplifying cellular complexity via the

construction/deconstruction of basic objectives/tasks into mutually dependent complex global tasks.

2.2 Cellular Organisation, Adaptation and Complexity from the Bottom-Up

Modelling and simulating multi-level dynamics of biological systems are one of the most complex endeavours in computational systems biology, due to the fact that biological processes consist of multi-level spatial and temporal scales (Bassingthwaighte, Chizeck et al. 2006; Schnell, Grima et al. 2007; Noble 2008). Living systems are the most complex systems known in nature. This is due to the multiple levels of constraints associated with them. Living systems are constrained by physical laws, like non-living systems and also have additional levels of constraints associated with complex biological processes. These two levels constitute the fundamental and organisational principles, which are required to model the complexity of biological cells from the bottom-up. When considering the relationship between individual biomolecules and the cells to which they contribute, we can identify their resemblance to complex, dynamic, self-organising, adaptive, concurrent, robust, reactive and proactive systems (Michener, Baerwald et al. 2001). Some of typical properties of complex systems include dynamics, emergent behaviour, nonlinearity, bi-stability, nested organisation, feedbacks (i.e. horizontal and vertical) and scale freeness (Dubitzky 2006). Biomolecular activities occurring within the gene, transcript, protein and metabolite space contribute to the organisation of a biological cell. These activities form various causalities (i.e. causal links amongst events), which form the organisational closure of a cell (see Figure 2.1). This closure is different from thermodynamic closure, which is observed in isolated systems. Although biological systems are organisationally closed, they are thermodynamically open systems that exist far from thermodynamic equilibrium by exchanging matter and energy with their environment (Van Regenmortel 2007). For example, at the organisational level various resources (e.g. metabolites) are consumed and produced by various enzyme mediated reactions, and if this is visualised by comparing every resource against every reaction, complex dependencies between enzyme mediated reactions at a thermodynamic level can be observed. Appendix B provides an example

of these dependencies and relevant free energy constraints, which determine the thermodynamic equilibrium of respective reactions. As a physical system the laws of thermodynamics govern cellular metabolism and as living system adaptability, robustness and efficacy ensures the persistence of a system.

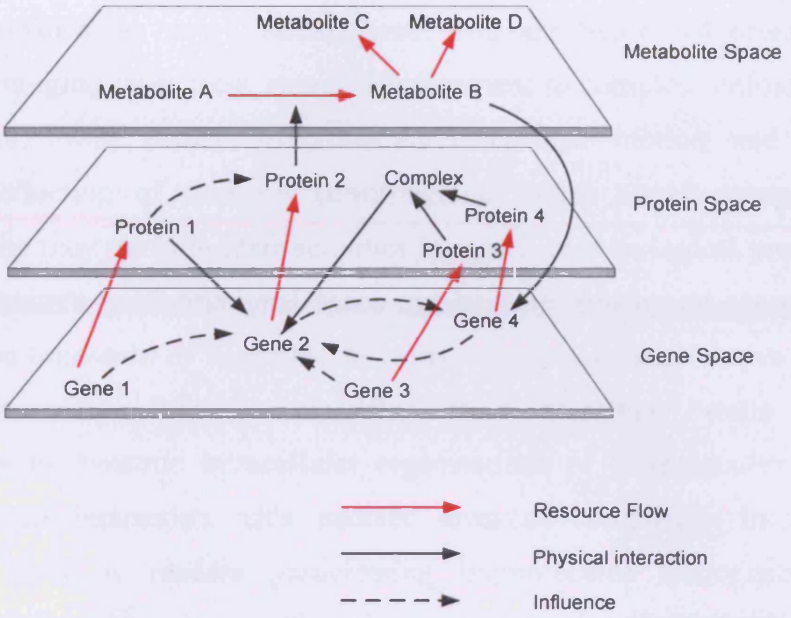


Figure 2.1: The autocatalytic cycles that traverse across gene space to metabolite space

An appropriate systems biology approach (Bruggeman and Westerhoff 2007) will have to be adopted to model the self-organisation of biomolecular activities in order to study emergence of an intracellular behavioural organisation. Since it requires a mechanism based explanation, it has to be mechanistically modelled using a bottom-up approach and integrating molecular level information. Modelling at the level of molecular resolution will require representing molecular properties, together with spatial and temporal constraints of the cellular environment.

2.2.1 Characteristics of Biomolecules and Cells

There are two kinds of properties which characterise biomolecules. Intrinsic properties are completely determined by a biomolecule's primary structure (i.e. the mass and sequences of DNA, RNA and proteins). While the primary structures of DNA and RNA (i.e. mRNA) contribute to the biological activities by harbouring and disseminating sequential information, the three dimensional structures of proteins and RNA (i.e. tRNA and rRNA) contribute to biological activity by functioning as

messengers, transporters, mechanical entities and enzymes (Stryer 1988). Biomolecules physically interact through activities to form collective autocatalysis, so that they recursively depend on each other in generation and realisation of various biological processes. The three-dimensional shape adopted by these molecular sequences in water is crucial to their biological activity. These tertiary structures are largely maintained by non-covalent forces and are hence subjected to thermal fluctuations ranging from local atomic displacement to complete unfolding (Brooks, Karplus et al. 1988). Protein structures are in constant motion, and they tend to sample a collection of different confirmations, which could change interaction patterns, while they perform their activities in a particular biological process (Vinson 2009). A protein's conformational space can be described by an energy landscape. Based on the timescale of biomolecular interactions and their relative mobility the cellular organisation can be classified from seemingly static intracellular organisations to dynamic intracellular organisations of biomolecules. Further the duration of an interaction adds another level of complexity in biomolecular interaction, since it renders participating biomolecules inaccessible to other biomolecules. The interactions of biomolecules, that represent a dynamic organisation are brief and produce complex biochemical tasks. These include the covalent and non-covalent interactions of biomolecules that produce biochemical activities, such as gene-regulatory, signalling and metabolic activities of the cell. These dynamic activities are highly adaptive in the context of a cell's physiological timescale. The interactions of biomolecules, that represent a seemingly static organisation are lengthy and produce complex spatial structures, via non-covalent biomolecular interactions within a cell's physiological timescale.

While biomolecular adaptation is crucial in altering characteristics of biomolecular behaviour, biomolecular activities (i.e. their performance) are crucial to the adaptation of a cell. A cell, as a living organisation, has managed to perform its biological activities by regulating the physical activities of its biomolecules. A cell is composed of physical entities such as macromolecules, small molecules, ions and water, which are constantly in flux. These physical entities are constrained by the laws of thermodynamics and become part of a living system, when their contributions have an effect on the living system. The extent to which these entities are self produced by a cell determines the degree of cellular autonomy. However, scale and nature (i.e.

positive or negative) of these entities' contribution to cellular activities can differ. Further the distribution, interaction and migration of these entities can influence the global dynamics of a cell. The internal organisation (i.e. functional and structural) of a cell depends on the self-organising interplay of non-covalent and covalent interactions. Reversible interactions of biomolecules are mediated by three kinds of non-covalent bonds, namely electrostatic bonds, hydrogen bonds and van der Waals bonds (Stryer 1988). These weak non-covalent forces are at the heart of all mechanistic activities of biomolecules and influence the dynamics of a cell. A key attribute that emerges out of these forces is the affinity for interactions, which also introduces competition and cooperation amongst constituent biomolecules in a shared environment. In an environment, that contains about 70 percent (i.e. by weight) of water, depletion forces (i.e. hydrophobic attractions) and diffusion play a non-specific role in biomolecular migration and distribution, while directed transport are specific to molecular species (Dogterom 2001). Molecules constituting the cells diffuse at a very slow rate due to molecular crowding. This slow rate of biochemical transformation and migration, which affects the rate of information/signal propagation, has caused cells to adopt a distributed strategy to control and coordinate cellular activities during the course of evolution.

2.3 Biological Adaptation from the Fundamental and Organisational Perspectives

Biological adaptations occur within the physiological, developmental and evolutionary timescales. Although the problem is focused on physiological timescales, it is useful to understand how biological systems are organised to adapt across these timescales, i.e. how the information regarding the performance between biological systems and the environment are exchanged across these scales. Biological systems dynamically adapt to multiple objectives concurrently, facilitated by their constituents. The objectives of biological systems are constantly evolving due to ever changing demands of its environment. These objectives are imposed by the environment which consists of physical and biological elements of individual biological systems. Biological systems meet these demands by pursuing objectives aided by their constituents, giving rise to biological processes which are perceived as

biological functions. Biological tasks emerge when pursuing these objectives which are concurrent and mutually dependent. These objectives on which the selective pressure is imposed are eventually organised into a hierarchy forming the biological organisation strata, where the amount of time required in pursuing the objectives increases, when moving up the hierarchy.

The neo-Darwinian view of evolution is built on three main observations of natural selection (Crespi 2001). First is heredity, where the composition of traits is determined by parents. Second is variation, where random alterations expand the search space of individuals, providing desirable attributes of diversity. Third is differential reproduction, where fitter individuals have a higher probability of surviving or reproducing to the next generation. However, according to modern research on evolution, there are two fundamental limitations of the existing theory (Eberhart and Shi 2007). The first is that the origin of life by chance or alteration is highly improbable in the earth's historical time frame. The second is that evolution of complex life forms solely through alterations is also highly improbable. This leads to a new view of evolution, in which self-organisation plays an important role in biological adaptation (Kauffman 1993). Complex systems can appear over a relatively short time frame compared to Darwinian evolution. In this new perception of evolution, it appears that natural selection and self-organisation are intertwined and operate together to facilitate biological adaptation. The following sections will focus on self-organisation, while Section 2.5.2 will briefly discuss natural selection in the context of multi-level biological organisation.

$$\text{biological adaptation} = \text{natural selection} + \text{self-organisation}$$

Based on principles of biological adaptation it is important to understand natural goals, which act as drivers and the constraints involved in guiding the organisational behaviour of a cell. These goals come in two forms, objectives which are universal to every biological system, and objectives that are specific to species. Species specific objectives will have to be pursued, whilst concurrently complying with universal objectives of living systems. In the context of intracellular adaptive dynamics, these goals/objectives and constraints represent biochemical activities in a cell. However, a gradual increase in diversity of biochemical activities, means there is now a great deal of complexity due to billions of years of evolution (Corning 1995). Initial biochemical

tasks were much simpler, compared to the complex tasks that exist today. Hence the best way to understand the principles behind construction of biochemical tasks, is to look at simple biochemical activities that existed in proto-cells, which brings our attention to the theory of the origins of life, and enable common objectives of living systems to be abstracted. Before striving to answer, *how life emerged?* It is essential to answer two vital questions. *What constitutes a minimal life?* And *why did it emerge?* Stuart Kauffman proposed five physical conditions for minimal life (Kauffman and Clayton 2006). To answer the second question, it is important to understand objectives of molecules in terms of physical and chemical laws. However a question that needs addressing is: *Did the molecules have the deliberation to create minimal life* or *were they guided by the constraints of the physical environment to create life as a spin-off, whilst maintaining their original objectives?* The notion of objective may differ at different levels in biological organisation strata. At an atomic level, the atoms stabilise by having a propensity to reach a noble gas configuration. However, they are constrained by the different affinities of different atoms, this leads to attaining a specific molecular configuration, whilst maintaining their original goal. At the molecular level, molecules stabilise by having a propensity to lower the internal energy (i.e. electronic, vibrational and rotational energy) state. This is achieved by a conformational change to reach the lowest possible energy state, based on its immediate environment (van Gunsteren, Bakowies et al. 2006). Further these molecules tend to cluster to reach stability. This propensity of physical interaction between molecules to reach their goal may answer the important question, *why molecules interact?* The laws of thermodynamics play an important role in this process (Wolfe 2002).

Although there are no standard models for the origin of life, it is thought the first biological systems (i.e. a protocell) emerged from simple organic molecules, that were capable of self-maintaining, self-replicating and evolving (Solé, Munteanu et al. 2007). There are two broad classes of theory, as to how life first originated from non-living matter. The *replicator first theory* states that large molecules capable of replicating (such as RNA) formed by chance, whereas the *metabolism first theory* states that small molecules formed an evolving network of reactions driven by an energy source (Shapiro 2007). A minimal life can be viewed from a physical (i.e. thermodynamics), chemical or biological perspective. The thermodynamic definition

of life states that a localised region that increases order (i.e. decreases in entropy) through cycles driven by energy flow, would be considered alive. Shapiro states five requirements for the metabolism first theory which are useful to abstract the universal objectives pursued by biological systems:

1) A boundary is needed to separate life from non-life,
2) An energy source is needed to drive the organisation process,
3) A coupling mechanism must link the release of energy to the organisation process that produces and sustains life,
4) A chemical network must be formed to permit adaptation and evolution, and
5) The network must grow and reproduce

A boundary need not be a physical barrier, such as a membrane bounded system. It can also be an organisational closure, formed due to collective autocatalysis. For a chemical network to adapt and evolve, it should be goal/objective oriented. These drive the adaptive process. Understanding reproductive strategies is important, in identifying the units of selection involved in biological adaptation. This is discussed in Section 2.5.2. There are various replication strategies established in a biological hierarchy. The replication of biological entities (atoms, biomolecules, cells and organisms) can occur independently or dependently. Totally independent entities (e.g. unicellular organisms and some multi-cellular organisms) can self-replicate whereas dependant entities have to rely on other entities of the same species (e.g. sexual reproduction), or entities from different species (e.g. biomolecules and virus) or external synthesis machineries (e.g. atoms). Atoms can neither self-replicate, nor influence other atoms for reproduction. They have to depend on external nucleosynthesis machineries (such as the Big Bang nucleosynthesis, Stellar nucleosynthesis, Supernova nucleosynthesis, or Cosmic ray spallation) (Clayton 1983) and can be classified as allopoietic systems. Reproduction within biomolecular species depends on complex interactions with other biomolecular species. This forms the synthetic machinery known as autocatalytic sets or collective autocatalysis. The biological cell as a whole is considered to be an autopoietic system, where numerous autocatalytic sets interact via control loops to self govern the cell. The transition from allopoietic to autopoietic status is one of the hallmarks of the protocell. An autocatalytic set is a collection of molecular entities, each of which can be created

catalytically by other molecular entities within the set, such that as a whole, the set is able to catalyse its own production (Kauffman 1993). Further, these sets have the ability to replicate themselves even if they are split apart into two physically separate spaces (Shenhav, Oz et al. 2007). Biochemical activities are complex autocatalytic sets and the reproductive information for these sets can be stored as sequential and compositional information (Segré , Ben-Eli et al. 2000; Segré and Lancet 2000). Although sequential information storage mechanisms dominate (DNA and RNA sequences) organisation of biological cells, the idea of a "compositional genome", which can accumulate and reproduce collectives of biomolecules (i.e. chemical information), is being proposed as an alternate theory for the formation of protocells.

Objectives that are universal to every living system and are specific to organisms, have to be pursued concurrently for persistence of biological systems. The two major tasks found in biological systems are anabolism and catabolism, which self-regulate the distribution of matter and energy in biological systems. Matter in various forms is produced to perform diverse activities within a cell. These can be universal to every biological system or specific to organisms. The tasks of catabolic activities are to release energy and basic building blocks for the production of complex biomolecules.

2.4 Organisational Space of Biological Cells

Biomolecules give rise to living entities by arranging themselves into coordinated biochemical activities, whose ultimate outcome is the production of life. The multi-dimensional problem that needs to be resolved, incorporates balancing a myriad of biological activities at various levels of biological organisation to result in a viable living system. However, a suitable resolution must exist within the "organisational space" defined by the constraints of each constituent biomolecule and their activities. At a cellular level, solutions to the adaptive requirement emerge from the simultaneous optimisation of multiple and mostly conflicting (due to competition amongst biochemical objectives) objectives via various critiquing mechanisms (forms of feedback and reinforcement mechanisms which facilitate self-organisation and selection), and function as regulators in space and time. A critic has a perception at its system level, that one outcome is qualitatively better than another at this level, but cannot determine whether this will be true at higher levels and thus cannot determine,

if it will lead to an absolute fitness specific to the requirement. Moreover, a critic doesn't inherently know what is the optimum solution, or even if one exist (Eberhart and Shi 2007). Although each of the objectives will not have an optimal solution, the solutions observed will ultimately satisfy the requirements in a sustained biological equilibrium. However challenges to this equilibrium which exceed the capacity of a specific system to compensate will create a pathological process, resulting in a multi-objective re-optimisation manifested as biological adaptation. Further pathological processes have become an integral part of biological adaptation due to failure in achieving the objectives caused by unanticipated constraints. Moreover there will be multiple biological solutions, which represent different "trade-offs" among objectives and constraints, associated with a biological system. The preferred solution will vary depending on changing requirements (i.e. criteria) exerted by the organisation's dynamic environment.

Feedback mechanisms are noticeably different from reinforcement mechanisms. In feedback, molecular switches directly interact with a signalling molecule (input signal) to alter the response without changing the basic responsive behaviour of a system to future occurrences. Positive feedback includes replication of functional products and activating them. Negative feedback includes the degradation of functional products, the inhibition of a functional product's activities, competition for resources, exhaustion for resources and the saturation of biomolecular activities. Feedback is short-lived, being limited by the duration of interactions. In contrast, reinforcement occurs when an event following a response causes an alteration in the probability of that response occurring in future. Reinforcement changes the basic responsive behaviour of a system to future occurrences independent of the signalling molecule (e.g. alterations in processing time of enzymes or their abundance). A permanent change in the responsive behaviour of a system to future occurrences will occur with reinforcement (Wikipedia Contributors 2009a) .

At molecular resolution level, two types of molecular switch exist. One type remains active by default, and is deactivated by a signalling molecule (negative feedback) and another type remains inactive by default and can be activated by a signalling molecule (positive feedback). However at cellular resolution level, when we look at these redundant events (molecular switches as redundant counterparts in two different states due to activation or deactivation) as a statistical process, two distinct patterns emerge.

These feedbacks include positive feedback that amplifies a desired outcome and negative feedback that reduces an undesired outcome.

Table 2.1: The amplification and reduction of activity via feedback mechanisms

Switches	Increase in signalling molecules	Decrease in signalling molecules
Active by default	Decrease in Active switches	Increase in Active switches
Inactive by default	Increase in Active switches	Decrease in Active switches

2.4.1 Cellular Level Properties

2.4.1.1 Adaptability

Intelligence is often associated with learning, which is an adaptive process. The most appropriate definition for intelligence that covers all computational intelligence approaches is described as "the capability of a system to adapt its behaviour to meet its goals in a range of environment" (Fogel 2006). The ability to learn or adapt is one of the hallmarks of intelligent systems. This can also be witnessed in biological cells, where cellular intelligence emerges as an organisational/system level property. The mechanism, that drives this intelligent behaviour is reinforcement adaptation, which is ubiquitous to biological systems. Reinforcement adaptation is facilitated via a critic, which follows a general principle that serves to guide the adaptive process. Biological systems can be assumed to follow the *law of sufficiency,* which states that if a solution is good enough, fast enough, and cheap enough, it is sufficient (Eberhart and Shi 2007). Hence the suitability of a solution (i.e. fitness) is not an absolute measure, rather it is a relative measure (i.e. how good the solution is relative to other solutions). Figure 2.2 shows the outcome of the law of sufficiency, which causes diversity in outcomes, which could give rise to diverse solutions as observed in nature. If perfection is the norm, there will be no room for deviation or defects in the outcomes eventually leading to uniformity in solutions.

The proactive nature of cellular behaviour is a result of the collective organisation of biomolecules and their interactions in space and time. Each biomolecule is simply reacting in a determinate way to stimuli and in-turn responding by stimulating other biomolecules to regulate activities amongst them. Various activities are required to

provide system wide responses to perturbations. However these activities have their limitations, and have to be regulated in terms of when, where and what activities should occur to provide timely responses to perturbations in a constrained environment. As a result, various stages of regulation have evolved in anticipation of perturbations, which facilitate transformation of reactive activities of native biomolecules to a collectively proactive organisation. The presence of higher stages of regulation such as translational and post-translational regulation, facilitate anticipation of recurring perturbations, which also improve the performance of a cell.

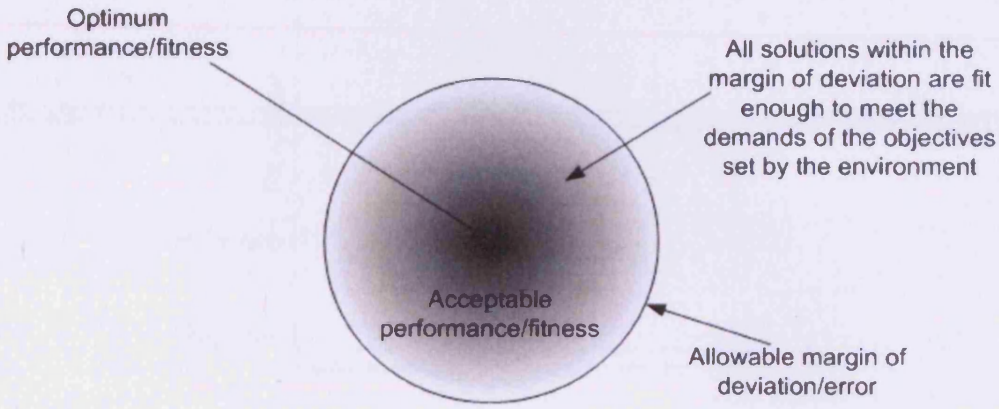


Figure 2.2: The outcome of the law of sufficiency is diversity in solutions, where the fittest solutions converge into an attractor basin

From a reductionist perspective, the organisational properties evident at cellular level such as efficiency, robustness and adaptability, cannot be perceived by characterising biomolecules. In the context of reductionism, a cell is perceived tangibly as its constituent biomolecules migrating, physically interacting and causing the density of biomolecular populations to fluctuate in space and time. However, this perception is misleading, since the cell is a collective of autonomous biomolecules exhibiting cohesiveness only at a holistic level. Moreover the intra-organisational behaviour of a cell cannot be directly observed or empirically measured, because this requires analysis of the performances of biomolecular species via their activities, analysing the contributions of basic tasks to the complex global tasks of the cell and tracing causalities via causal links amongst biomolecular activities. At an organisational level, the cellular behaviour can only be probabilistically determined, since causalities occur due to concurrent biomolecular activities. Figure 2.3 shows the determinist and

reactive nature of a biomolecule giving rise to a cellular organisation, which is probabilistic and proactive in nature. The deterministic nature of biomolecular behaviour can produce coordinated behaviour amongst biomolecules, causing reproducible or rhythmic intracellular organisational behaviour, in the face of perturbation and uncertainty.

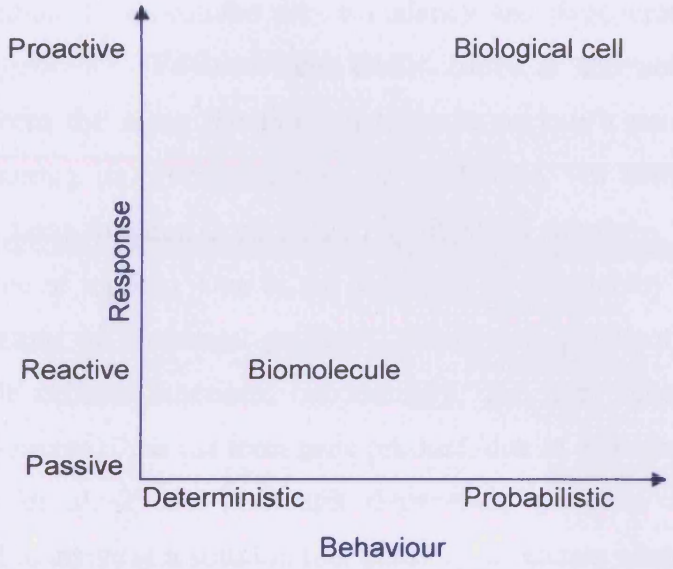


Figure 2.3: The nature of a biomolecule and the biological cell. From the reactive and deterministic nature of biomolecular activities, the complex, nondeterministic and proactive cellular behaviour will emerge.

2.4.1.2 Robustness

Robustness is an organisational/system level property, which is defined as "the ability to maintain performance in the face of perturbation and uncertainty"(Stelling, Sauer et al. 2004). However comprehension of how robustness is accomplished at the cellular or molecular level is still limited (Hartman, Garvik et al. 2001), due to its intimate link with the complexity of cellular systems (Stelling, Sauer et al. 2006). An important realisation is, that robustness is concerned with preserving the functions of a system rather than system states. This distinguishes robustness from stability or homeostasis (Kitano 2007). *Homeostasis* is a process, that preserves the state of the system rather than its function. Robustness determines the boundaries (see Figure 2.4: The formation of robustness and its associated biological equilibrium.) of the multi-dimensional problem (i.e. perturbation and uncertainty) and the function (i.e.

performance) space, in which biological equilibrium can exist. *Perturbation* defines the extrinsic (environmental) stimulus and intrinsic (programmed) stimulus. *Uncertainty* defines the stochastic nature of the constraints, such as the intervals between biomolecular activities and the availability of resources, which a cell cannot produce. In the context of biological adaptation, function is defined as progression along some causality, to a goal or successful outcome (Dusenbery 1992). Some of the factors that contribute to robustness are redundancy and degeneracy, plasticity and concurrency. *Degeneracy* (Edelman and Gally 2001) is the ability of different solutions to perform the same function, such as an enzyme's performance can be maintained by altering its processing time or abundance. In contrast *Redundancy* occurs, when the same function is performed by identical solutions. Also redundancy refers to the degree of replica. One of the outcomes of degeneracy is the pleiotropic and polygenic nature of functional products, where they positively and negatively influence multiple cellular functions, concurrently. The term functional product is currently more favoured, than the term gene product, due to changing views of genes (Gerstein, Bruce et al. 2007). Although degeneracy provides flexibility (many options) for a cell to arrive at a solution (i.e. possibly accelerate adaptation), it adds to complexity in recognising contributions and compensatory adjustments made by different options to the solution (this phenomena is demonstrated in Section 7.4). *Plasticity* is the ability of a system to readily adapt to new, different, or changing requirements (Garnier, Gautrais et al. 2007). *Concurrency* manifests with the existence of redundant and specialised biological entities, such as diverse biomolecular species and cell types. The effects of robustness are sensitiveness (fluctuation of performance to perturbations) and adaptability. Robustness facilitates adaptability by accumulating variations whilst maintaining a functional phenotype, such as silent or neutral mutations in the genome.

Further, cellular organisation has the ability to efficiently adapt within the bounds of biological equilibrium and gracefully degrade its performance, when functional/performance requirements, perturbation or uncertainty levels demand more than the capacity of robustness. Hence, not only does biological cell maintain performance, which is constrained by its genome, within the capacity of its robustness, but it also has the ability to reconfigure the responsiveness at the genome level to meet performance demands of the dynamically changing capacity of

robustness. The fitness of solution in an internal cellular organisation is constantly being evaluated and it is the measure of performance with respect to an objective. That is how well an intended task is being fulfilled. Although every functional product has a purpose (intended activity), ultimately their contribution to the overall performance of a cellular organisation, which in turn contributes to its reproductive success, is essential to an understanding of their impact from the bottom-up. These functional products will have positive contributions to sustaining biological equilibrium, when their activities are performed when required. However, when their activities are silenced or performed when not required, it can have a negative contribution to sustaining biological equilibrium. Biomolecular activities are directional/vectorial in terms of their causality (cause and effect), which contributes to the transformation of the cellular organisation's equilibrium state, either towards or away from equilibrium, depending on an organisation's state. Hence the purpose of a functional product in the context of its higher organisation (cell) depends on the circumstance, in which the activities are performed. In a normal system various feedback mechanisms, formed by regulatory switches which span from transcriptional level to post translational level, ensure the activities occur in an appropriate circumstance to sustain cellular functions.

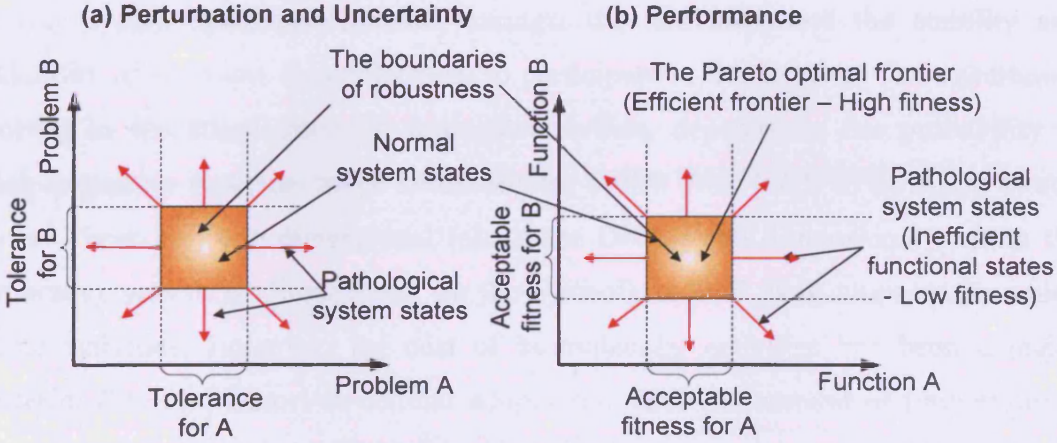


Figure 2.4: The formation of robustness and its associated biological equilibrium. (a) Perturbation and Uncertainty- The existence of the normal system phases (biological equilibrium within the bounds of robustness) boundaries in 2-dimensional problem space. (b) Performance - The existence of Pareto optimal frontier (The region of high fitness) boundaries in 2-dimensional function space.

2.4.2 Uncertainty in a Cellular Environment

The process of biological adaptation involves self-organisation and selection, which contributes to the optimisation of biological systems. These two mechanisms, which are facilitated by feedback and reinforcement mechanisms, should occur with acceptable fidelity to ensure persistent behaviour in biological systems. A cell's ability to organise implies, that it has the ability to optimise cellular activities under various perturbations and uncertainty. The existence of uncertainty in a cellular environment for which the genome has no control, is due to the presence of faulty activities, unpredictability of causal activities inherent due to concurrency, and the downstream amplification of activities. For example, during the course of evolution an error frequency of about 10^{-4} per amino acid residue, has been selected to produce the greatest number of functional proteins in the shortest time (Stryer 1988). The ability to organise depends on the predictability of biomolecular activities, which have to significantly dominate uncertain activities. Due to the uncertain nature of the cellular environment, cellular adaptation is driven by the most probable molecular activities that occur, based on the constraints in their local environment. Constraints reduce uncertainty by guiding the system. The main constraints for molecular activities include, cost of the activity in terms of time and energy (i.e. enzyme turnover cycle), spacetime interval amongst the activities, and the stability and availability of reactants (biomolecules) to participate in the activity. The uncertainty involved in spacetime intervals amongst activities, depends on the probability at which respective reactants meet. Biomolecules utilise three kinds of diffusion search spaces. These are, one dimensional (along the DNA), two dimensional (within the membrane), and three dimensional (in the cytosol), to find their counterparts which initiate activities. However, the cost of biomolecular activities has been a major constraint (limiting factor) in cellular adaptation, since the amount of time required for various biomolecular activities, significantly dominates the time requirements for diffusion mediated encounters.

The stability of native biomolecules also plays a major role in the self-organising process of a cell, because it determines the functional ability of these molecules. The main factors, which affect stability of molecules are temperature, pH and vulnerability to destruction. Proteins are the molecular machines of a cell and they have evolved to be the major contributors to the organisational dynamics of the cell. Proteins exist in

various stages of the lifecycle (Belle, Tanay et al. 2006) and differ noticeably in their half-lives. While some are destructed very rapidly (typically enzymes), others are very stable (mechanical proteins). In Proteins the half-life is determined to a large extent by its amino-terminal residue (see Table 2.2), which acts as a signal for stability and has been retained over the course of evolution (Stryer 1988). There is a complex interplay between protein degradation, its regulation and other determinants of protein metabolism (Saric and Goldberg 2006). The cellular organisation has adopted this susceptibility of native biomolecular degradation as nonspecific negative feedbacks, which contribute to the internal organisation of a cell.

Table 2.2: Half-lives of cytosolic proteins which depend on the nature of their amino-terminal residue (Adapted from (Stryer 1988)

Amino-terminal residue	Half-life
Stabilizing	
Methionine	
Glycine	
Alanine	>20 hours
Serine	
Threonine	
Valine	
Destabilizing	
Isoleucine	~30 minutes
Glutamate	~30 minutes
Tyrosin	~10 minutes
Glutamine	~10 minutes
Proline	~7 minutes
Highly destabilizing	
Leucine	~3 minutes
Phenylalanine	~3 minutes
Aspartate	~3 minutes
Lysin	~3 minutes
Arginine	~2 minutes

2.4.3 The Impact of Time and Energy in Biological Adaptation

The role of energy in biological adaptation has been emphasised in "thermoeconomics", as productivity, efficiency and profitability of various mechanisms for capturing and utilising available energy to build biomass and do work (Corning 2002). In metabolism there is a net energy gain in catabolic activities, and a net energy loss in anabolic activities. In biochemical systems, energy released by catabolism is utilised to drive the synthesis of ATP (known as currency of energy), which in turn is used for anabolism. Since ATP is released to a common pool and used as a currency, cells have the flexibility to utilise it for any activity that requires it. To facilitate this enzymes play a crucial role in metabolism, because they drive biologically desirable but thermodynamically unfavourable reactions by coupling them to favourable ones. The self-organisation processes in cells are non-spontaneous, because energy is required to produce various functional products to maintain order in cells. Various steady states of biological systems, which have emerged to maintain biological equilibrium far from thermodynamic equilibrium, attract non-spontaneous processes to increase order, whereas thermodynamic equilibrium attracts spontaneous processes to decrease order. The trajectory between these two biochemical system phases is controlled by metabolism, where anabolism is dominated by non-spontaneous processes, and catabolism is dominated by spontaneous processes (see Figure 2.5).

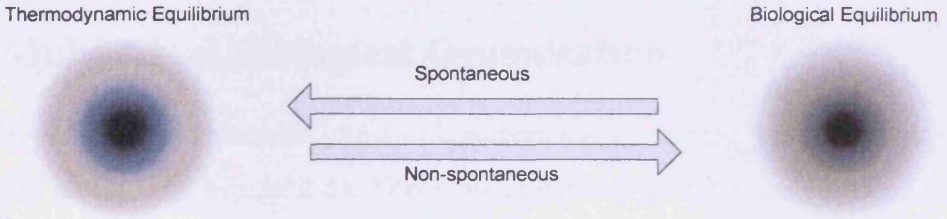


Figure 2.5: The role of metabolism in cellular homeostasis

This specificity has constrained and guided self-organisation in biochemical systems. Constant energy flux (energy dissipation) between spontaneous and non-spontaneous processes provides instability, which is required for the self-organisation process. If the metabolic phase of a biological system reaches thermodynamic equilibrium, it will no longer be considered as a living system. The frequency of reproduction of cells

will depend on the amount of energy utilised for reproduction. Energy utilised for other mundane activities of a cell can reduce the frequency of reproduction.

Various regulatory switches have evolved to self-organise a cellular environment. While some switches utilise little or no energy (e.g. binding of signalling molecules), others require chemical modifications using high energy bonds (e.g. chemical modifications mainly by phosphate groups and other groups such as acetyl, methyl and adenyl). Activities of functional products are orchestrated via various regulatory mechanisms which range from transcriptional regulation (genetic level), through post-transcriptional regulation, translational regulation (transcript level) and post-translational regulation (protein level). While transcriptional regulation provides slow and globalised cellular responses, post-translational regulation provides rapid and localised cellular responses. Transcriptional response is the most time and energy consuming process, since genetic information has to be transcribed and mostly translated to produce a functional product. In contrast, post-translational response is the least time and energy consuming process, since functional product is simply switched between an active and inactive state. Further transcriptional regulating is relatively stationary, while the remaining regulatory mechanisms are mobile and provide rapid and localised regulation within a cellular environment. Regulations facilitate in the timing of a functional product's activities. Appropriate timing of activities is essential, because its impact depends on the phenotypic state of a cell.

2.5 Multi-Level Biological Organisation

A biological cell is organised into an objective/task hierarchy, which contains various cohesive levels (see Figure 2.6). These objectives range from the level of molecular species, where they are atomic and independent of one another, to the basic tasks and finally to cellular level, where objectives become global, mutually dependent and biological. When more than one biomolecular species is involved in the formation of a basic task, mutual dependency will exist amongst the biomolecular species. Hence there is a gradual transition from objectives being independent at the molecular level to mutual dependency of objectives at the cellular level. The objectives between levels of the hierarchy are semantically different. The tasks/objectives range from being physical to chemical and biological, when traversing from molecular resolution

to cellular resolution. At molecular resolution, the tasks are physical. At the biomolecular species level, the objective is represented by their ensemble activity. At the cooperative level where basic functional units emerge, objectives are involved in completing chemical tasks. However, at the cellular level objectives have the characteristics that are fundamental to living systems. That is efficient use of energy, timely responses to perturbation, persistence and other biological characteristics. Further these system level tasks/objectives are not communicated directly to constituent biomolecules, rather they are self-maintained in a concurrent manner. Nature is inherently concurrent and biological systems are no exceptions. Since cellular objectives are not maintained centrally, cells have adopted a unique strategy to continuously realise their objectives by eliminating obsolete information from their organisation. The propensity of biomolecular degradation by means of random or regulated processes and collective autocatalysis provides an ideal reinforcement adaptive mechanism for a cell. The process of biomolecular degradation can eliminate obsolete biomolecular activities and so keep cellular activities up to date, and recycle resources to maintain cellular activities in a resource constrained and dynamic environment. These mechanisms are ubiquitous cellular processes and are pivotal for adaptive dynamics and evolution of an intelligent cellular organisation (Periyasamy, Gray et al. 2008b).

Cellular level objectives are constrained by lower level objectives, many of which are in conflict, so various regulatory mechanisms facilitate in managing these conflicts. The higher level objectives enforce adaptive requirements for the lower level objectives. Measuring performances of objectives within a hierarchy would facilitate understanding of the functional organisation of a cell. Multi-objective topology provides a concurrent and hierarchical view of cellular dynamics. A typical multi-objective optimisation scenario will generate a set of dominant solutions, which forms the Pareto optimal frontier (the efficient frontier) (Wikipedia Contributors 2010a). Optimisation uses a controlled trial and error process, where a cellular system is steered along a path of increasing organisation. Pareto optimality is an economic concept, which can be used to study system efficiency and the distribution of component activities. A Pareto efficient frontier is one, in which any change to enhance the performance of an objective is impossible without making the performance of another objective inferior. This is often the case, when there are

conflicts among mutually dependent objectives. A mathematically oriented (quantitative) definition for self-organised behaviour has been articulated as (Fleischer 2005) *"Self-organized behaviour in a complex system involving multiple performance measures is a sequence of system states corresponding to movement along a Pareto optimal frontier"*. This defines the best global solution that can emerge based on the constraints.

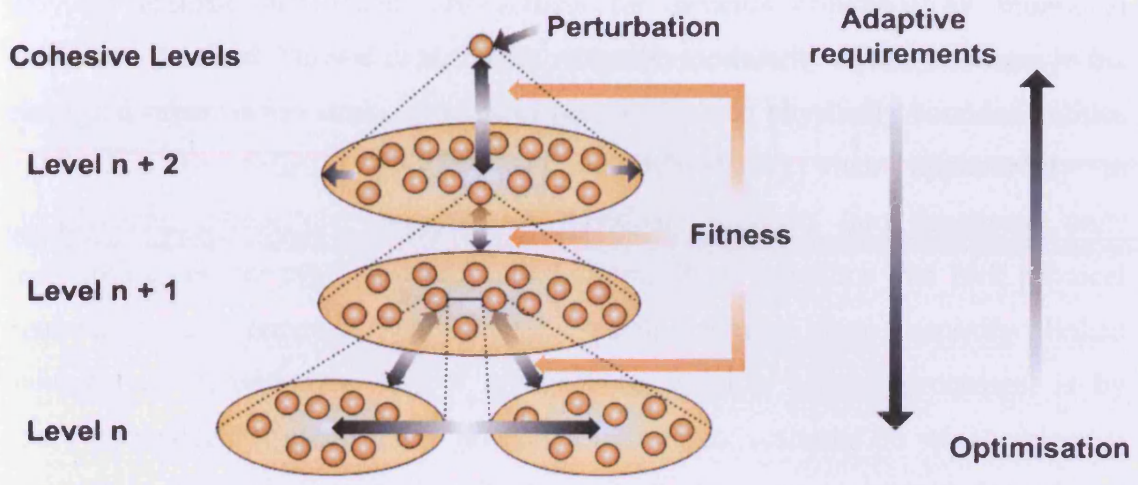


Figure 2.6: The objective hierarchy forming nested organisation in biological systems.

For example, aerobic and anaerobic respirations are dynamic solutions, which have emerged to fulfil the objective of liberating energy in the presence and absence of oxygen respectively. In the presence of oxygen, biomolecular activities pertaining to aerobic respiration will dominate, and in the absence of oxygen, biomolecular activities pertaining to anaerobic respiration tend to dominate. Hence, these two solutions, although they appear redundant with respect to a cellular objective of releasing energy, are really complementary (i.e. degenerate) with respect to the problem of oxygen content (Rosenfeld 2002). These adaptive strategies, which are a result of collaborative efforts of biomolecules, provide complimentary solutions for cells. The critiquing mechanisms of evolution are destined to select appropriate anatomical or physiological solutions (Regenmortel 2004).

2.5.1 Task Formation and Integration in Cells

Modularity is a way of simplifying complex systems into a set of simple systems using functional abstractions. To this end various criteria for simplifying complex biochemical activities of life have been proposed, using modularity to encapsulate

biological complexity. One such modularity is based on a cellular component, biological processes and molecular function, which do not represent the nested hierarchy of biological organisation existing today (i.e. how various basic tasks have been evolving to form various complex cellular level tasks), that has evolved from proto-cells to complex multi-cellular organisms. Moreover the long standing question is to what extent the concept of modularity introduced for engineered systems, provides realistic and useful abstractions for systems organised by biological adaptation (Szallasi, Periwal et al. 2006). Although modularity can be observed in the biological organisation strata in terms of perceivable and physically bounded entities (molecules, organelles, cells, organs and individuals), their applicability in modularising intracellular activities of functional products into functional units constituting cellular processes is doubtful. Intracellular functions that lack physical boundaries are temporal phenomena, which emerge from causally linked biomolecular activities. A logical approach to simplify cellular processes, is by constructing/deconstructing these processes into objectives/tasks on which selective pressure is imposed. Further modularity is concealed, due to mutual dependency amongst higher level tasks. The effects of mutual dependency amongst the objectives/tasks, which occurs due to the presence of degenerate and redundant factors, and the convergence and divergence of causal effects of biomolecular activities, adds to the complexity of modularising biochemical activities. Mutual dependencies complicate the process of identifying the degree of orthogonality (i.e. independence), which facilitates the modularisation from molecular resolution to cellular resolution via deconstruction of objectives into the basic and atomic tasks required to pursue them. The emergence of global cellular behaviour is a result of functional products, which are specialised to pursue their intended tasks. Further acts of cooperation, competition and coordination emerge from the collective behaviour of functional products. These actions are not mutually exclusive, rather they contribute concurrently to the pursuit of various collective tasks of cell and higher multi-cellular organisations. The criteria used to modularise interactions among functional products, are based on performance/fitness interactions, which emerge out of competition and cooperation among functional products. This is the mechanism by which evolution formed and evolved collaborative groups, containing one or more species of functional product. These functional products within a group cooperate with each other for a common objective/task. Competitive and cooperative adaptation among

various biomolecular species is ubiquitous amongst their activities. While inverse performance/fitness interaction exists between competing biomolecular species, positive performance/fitness interaction will exist among cooperating biomolecular species. Direct and inverse fitness interactions can reveal the organisation of an objective hierarchy in order to construct/deconstruct tasks between molecular resolution and cellular resolution. Further this relationship is appropriate to model impact amongst various species of a biomolecule's activity on the intracellular and the cellular level tasks, as a whole.

Hypercycles are formed due to convergence and divergence of causalities. The interaction between a common transcription factor and various *cis* regulatory sites, is an indication of divergence in causality. The presence of divergence points in biochemical networks is an indication of competition for a common substrate and from these, conflicts among higher level cellular tasks/objectives will arise. Shared resources are a major cause of conflicts in intracellular organisation. A basic task or a cooperative module in biochemical activities is defined as a group of one or more species of functional product collaborating for a common objective. These modules will have the characteristic, that every functional product's performance will have a beneficial effect on the other and the whole group's performance. The absence of any one member species of a group, will have no value for the existence of the remaining member species of the group (all or nothing phenomena). In molecular complexes the participating biomolecular species form cooperative groups. In the context of metabolic networks, this is a pathway which exists between two junction points. This will be the smallest module of objective function, from which higher levels of objective function will have to be assembled. Fitness at a functional product level is a function of its efficiency and stability. Efficiency depends on a product's affinity for interaction, and the time and energy requirements for its activity. An improved performance for one competing group implies a decreased performance for the other group. Hence they have an inhibitory effect on other competing groups. Further, biomolecules are forced to sacrifice their efficiency for betterment of a cellular organisation. This inverse performance between two levels can only occur in the presence of conflicting objectives. These conflicting groups will impose immense selection pressure on their regulatory mechanism.

Multi-scale interactions deal with associating molecular level activities to cellular level processes. These include representing spatial, temporal and energy constraints, and analysing efficiency, robustness and adaptability from molecular resolution to cellular resolution. Biomolecular activities differ in timescales, which can range from microseconds, as observed in some of the most efficient enzymes, to minutes as observed in transcription and translation of functional products. Although these differences may not appear significant superficially, it has a significant impact on the self-organisation of cellular processes. While a scoring mechanism is essential to measure performance, a ranking mechanism facilitates by guiding molecular level interactions to a desired system level behaviour. Further, posing questions at a cellular resolution and seeking answers at a molecular resolution, and vice versa, is one of the requirements of multi-scale interaction. Scoring and ranking biomolecular activities will enable a traverse between these two resolutions. Every biomolecular activity is susceptible to critiquing mechanisms (which act as regulators) of adaptation, which occur horizontally and vertically in the biological organisation strata. Further these critiquing mechanisms are exercised at physiological, developmental and evolutionary timescales.

Managing integrity in multi-cellular organisms requires an additional set of gene products to regulate extrinsic control mechanisms. Existing multi-cellular organisms show two distinct types of control mechanisms to maintain multi-cellular integrity. They are hormonal control mechanisms and neuronal control mechanisms. Hormonal control is a broadcasting mechanism used by specialised cell groups to communicate with other cells types. These decentralised control mechanisms have the ability to target specific cell types without any directional constraints. With the emergence of neurons in complex multi-cellular organisms, biological systems have evolved to incorporate centralised control strategies, which are mostly reliable but fragile, have a high rate of signal/information propagation and high specificity. Emergence of this cell type expanded the multi cellular organism's phenotypic space by providing more options, which led to the production of better solutions to meet the adaptive requirements of higher biological organisation.

2.5.2 Units and Levels of Selection

Recent work on major evolutionary transitions (from one level of organisation (the cell) to another (multi-cellular individuals)) emphasises the point that the general theory of evolution by natural selection must be hierarchical (Brandon 2001). Selection is one of the causes of biological adaption, which is defined as the *"differential reproduction of biological units due to difference in form or character between these units"*(Crespi 2001). Various biological units are subjected to the effects of selection at different levels of biological organisation. Units of selection are defined as the *"units whose frequencies are adjusted by natural selection across generations"*(Crespi 2001). Levels of selection are defined, as the *"levels of biological organisation where natural selection occurs, within generations"*(Crespi 2001). Biological units are arranged in a hierarchy (see Table 2.3), with lower level units nested into higher ones. Units at different levels exhibit diverse properties with respect to how they reproduce and the mechanisms by which they interact with units at different levels and aspects of the environment.

Table 2.3: The primary levels of biological organisation, the units at each level, and the properties of the units (adapted from (Crespi 2001))

Levels	Amount of variation	Turnover rate	Inheritance fidelity	Expression of traits	Nature of units
Genes	High	High	Very high	No	Replicator
Chromosomes	High	High	Medium	No	Replicator
Genotypes	High	High	Low	No	Replicator
Gene products	High	High	N/A	Yes	Interactor
Cells	Variable	High	N/A	Yes	Interactor
Individuals	High	High	Low	Yes	Interactor
Groups	Variable	Variable	Variable	Yes	Interactor
Species	Variable	Very low	Variable	Yes	Interactor
Communities	Variable	Low	Variable	Yes	Interactor

2.5.2.1 Units of selection

Individuals occupy a special place in the biological organisation strata, because each contains genes, chromosomes, genotypes, functional products and cell(s). Individuals

typically live, reproduce and die as units. Further they represent the constituents which combine in various ways to form the levels above them. The extent to which units at different levels are units of selection depends on Darwin's three conditions. First *"the units must exhibit variation among themselves in their effects"* (Crespi 2001). Second *"the units must have some rate of differential reproduction, which determines the frequency of selective episodes, and this turnover must be causally linked to variation among units"* (Crespi 2001). Third *"the units must persist as unique, replicating variant units for a sufficient number of selective episodes to have their frequencies adjusted by natural selection"* (Crespi 2001).

Based on the above criteria genes, chromosomes or genotypes can be the only units of selection. The "gene is the primary unit of selection because it is the only unit exhibiting high variation, high turnover rate and the ability to replicate or reproduce with extremely high fidelity" (Crespi 2001). Chromosomes become units of selection only when the rates of recombination are very low or zero, and genotypes become units of selection only in asexual organisms, where the absence of recombination and meiosis results in the inheritance of an entire genome unaltered. Further, being a replicator is vital to being a unit of selection.

2.5.2.2 Levels of selection

This describes, where selection exerts its pressure in the biological hierarchy. "Selection requires the expression of trait variation at some level, and interaction of that trait variation with the environment so that the units at that level and the lower level differentially reproduce" (Crespi 2001). Expressed traits include functional products or effects and phenotypes of individuals, groups or communities. Of these, individuals usually represent the most important level of selection.

2.5.3 Timescales of Biological Adaptation

Biological adaptations occur in physiological, developmental and evolutionary timescales. The information for this adaptive process is mainly stored as genetic information in the genome. This information exists not only in a gene's coding sequences but also in its regulatory sequences (Hopi and Jerry 2007; Prud'homme, Gompel et al. 2007; Wray 2007; David and Virginie 2008). While genetic adaptations contribute to biological adaptations at evolutionary timescales, epigenetic adaptations

contribute to biological adaptations at developmental timescale (cellular differentiation). Since genetic adaptations mostly occur during the reproduction of organisms, the evolutionary timescales will differ with the rate of reproduction. For example, the reproductive rate of bacteria is much higher than multi-cellular organisms. The rate of genetic adaptation tends to be much higher in organisms with a higher reproductive rate. Hence the evolutionary timescales will differ with organisms.

2.5.3.1 Epigenetic adaptations

Epigenetics has several meanings all with independent roots in biology. Although the working definition has become narrower, the term epigenetics was introduced and defined by Conrad H Waddington as the study of genotype giving rise to phenotype (Bird 2007). This represents the most extreme case of epigenetics where the position of each molecule is accounted for by the phenotypic state of the cell. Robin Holliday (Robin 1990) has defined epigenetics as the mechanism for spatial and temporal control of gene activity, during the development of complex organisms. It implies changes in phenotype, that is changes influencing the development of an organism, are due to mechanisms other than changes in the DNA sequences. There are various epigenetic mechanisms (Allis, Jenuwein et al. 2006; Tost 2008) listed in Appendix C and most of them are trans-generational mechanisms.

Many geneticists now believe that the behaviour of our genes can be altered by experience and can be passed on to future generations. This could transform our understanding of biological adaption (Hunter 2008). Hence the outcome of a phenotype is influenced by environmental factors, and epigenetic processes mediate genotype-to-phenotype relationships within the limits of a genotype, and respond to environmental perturbations to produce a phenotype. Only a subset of the genome is expressed at any given moment during physiological and developmental activities of an organism, and this is controlled by genetic as well as epigenetic mechanisms (Turner 2007). While genetic adaptation is a slow process, epigenetic adaptation is comparatively a quicker process (Rando and Verstrepen 2007). Although these two processes seem to evolve independently, they both contribute to the final phenotypic outcome. Hence the success of the phenotype not only depends on the genotype, but

also on the epigenotype. While phenotypic features cannot directly influence genotypic information they can influence epigenotypic information.

2.6 Concluding Remarks

This chapter has defined the adaptive dynamics of biological cells by utilising a multi-objective topology. This differs from a conventional network topology based description of intracellular dynamics. The chapter has also exemplified biological complexity from molecules to cell by deciphering the functional organisation of biological cells via multi-objective representation of intracellular adaptive dynamics. This chapter has characterised the crucial factors involved in biological adaptation such as adaptability, robustness and efficacy in the context of multi-objective topology which provides a hierarchical and concurrent view of the intracellular dynamics. An appropriate systems biology approach will have to be utilised to model the self-organisation of biomolecular activities in order to study the emergence of intracellular behavioural organisation. Since this requires a mechanism based explanation, it has to be mechanistically modelled using a bottom-up approach, which integrates molecular level information. Modelling at the level of molecular resolution requires representation of both the molecular properties, and the spatial and temporal constraints of the cellular environment. One of the challenges is that the organisational behaviour of a cell, is not something that can be directly observed or empirically measured. Instead it needs a group of actors to represent the functional products, a set of cellular resources utilised by these functional products, a way to capture the results of the functional products' activities, and a method to evaluate these results. The cellular activities, which correspond to a functional organisation are hierarchically organised into various basic tasks, which merge to form the complex and greater tasks of a cell. The next chapter specifies the functional and non-functional requirements needed to address the problems described in this chapter. It critically reviews related work with respect to modelling intracellular dynamics and evaluates suitable methodologies and platforms to address the research aims.

Chapter 3

Modelling Biological Phenomena

"The data are accumulating and the computers are humming, what we are lacking are the words, the grammar and the syntax of a new language..."

- Denis Bray

3.1 Overview

The aim of this chapter is to specify modelling requirements for addressing problems articulated in Chapter 2 and critically review related work with respect to modelling from molecular resolution to cellular resolution. Further, it addresses the feasibility of achieving the specified requirements with available resources. Section 3.2 specifies the major requirements needed to build a model to embrace the inherent principles governing self-organisation, adaptability, robustness and efficacy in biological cells. The scope of systems biology is reviewed in Section 3.3 to identify appropriate systems biology approach for the study. In Section 3.4, model development processes and hierarchical modelling approaches are reviewed to identify appropriate modelling methodologies. Section 3.5 reviews the scopes, strengths and limitations of various types of biological modelling formalisms. Section 3.6 evaluates modelling methodologies and modelling formalisms against model requirements to identify an appropriate model development process, hierarchical modelling approach and modelling formalism. Further, an agent based formalism in the wider framework of Collective Intelligence is identified as the approach for modelling and simulation of multi-scale adaptive dynamics from molecules to cell. Section 3.7 evaluates the feasibility of implementing and developing the chosen modelling approach in terms of available biological data sources, programming environments, platforms and computational advances.

3.2 Requirements for Modelling the Adaptive Dynamics from Molecules to Cell

3.2.1 Model Requirements

Modelling multi-scale adaptive dynamics from molecules to cell level, will require *in silico* representations at molecular resolution to model biomolecules and their activities in space and time. The activities which cause direct and indirect interactions amongst the biomolecules contribute to self-organising and emergent behaviour in biological cells. However these emergent behaviours must be analysable to produce practical models. One of the challenges is that organisational behaviour of a cell is not something that can be directly observed or empirically measured. Instead it needs a group of actors to represent the functional products, represent a set of cellular resources utilised by these functional products, capture the results of the functional products' activities and a method to evaluate these results. Self-organisation is the main principle required to build a mechanistic model. Since biomolecules do not possess any cognitive ability, they are unaware of the global state of a cell. The state of a cell has no direct influence on the behaviour of biomolecules, rather they simply react to the immediate environment in which they exist. To represent these phenomena, biomolecules must be represented as reactive entities without any deliberation with respect to their behaviour based on the global state of cell.

The following requirements are desirable for modelling the multi-scale adaptive dynamics from molecules to cell.

- The ability to model between molecular (the lowest level of abstraction) and cellular resolution (the highest level of abstraction). This requires representing multiple scales (from molecular to cellular) simultaneously to analyse performances within the objective/task hierarchy, analysing the timescales of molecular activities and the timescales at which their contributions can be realised, analysing energy requirements for molecular activities and energy production and consumption at the cellular resolution, analysing the efficiency of the functional products' activities and the efficacy of the diverse objectives/tasks to which they contribute, analysing stability of the functional products and their robustness at cellular resolution, and analysing adaptability

in the physiological timescales of functional products and at cellular resolution to which they contribute.

- The ability to represent concurrency, which is endemic in biological cells
- The ability to mechanistically explain emergent properties from molecules to cell
- The ability to measure organisational behaviour within a biological cell

3.2.2 Data Requirements

Many macroscopic descriptions of cellular phenomena are only an approximation, idealisation and generalisation of real molecular processes. Due to insufficient description/information at this level, they often rely on probabilistic or statistical concepts. In contrast, microscopic descriptions of molecular activities are associated with detailed descriptions. However, many molecular details are insignificant, irrelevant and inconsequential to specific macroscopic phenomena (Fromm 2005). Hence, every detail at molecular resolution will not be required to represent the intracellular organisational behaviour. The level of detail required to represent phenomena will increase when moving from population, to molecular and atomic levels. Further the information used at each level is semantically different. The significant, relevant and prominent properties for activities and interactions that are consequential to intracellular organisational behaviour will have to be identified. Two types of constraint which represent organisational and physical constraints will have to be represented, to model their effects on collective behaviour. This will require molecular level information, such as their diffusion constants, time and energy requirements for their activities, their localisation and abundance in a cell. Biomolecular activities are transformed into events, when they occur in a stipulated space and time. Modelling these events will require information at molecular resolution, such as time and energy requirements to represent the respective events. Modelling event intervals will require the diffusion constants of biomolecules, the distance amongst biomolecules and the affinities for interaction. Further biochemical thermodynamic information is required to model the physical constraints of biomolecular activities. Gibbs free energy (Stryer 1988) is mainly used as a thermodynamic property in biochemistry to provide quantitative answers to the

probable direction of chemical reactions. Free energy is also used to represent affinities for interaction amongst biomolecules. Information such as biomolecular degradation and error frequencies for transcription, translation and replication can also be beneficial when building a comprehensive model of a cell.

3.2.3 Technical Requirements

Modelling biomolecules will require a formalism that can model the behaviour of biomolecules and the intracellular environment. A cellular environment is spatially heterogeneous, because it is crowded, granular and inhomogeneous (Ridgway, Broderick et al. 2006). Stochasticity refers to the inherent uncertainty to movement, interaction and activity of every biomolecule. Further, the low copy numbers of some vital bimolecular species and their ensemble activities will fluctuate in a way that can only be described in terms of probability (Ridgway, Broderick et al. 2006). Under these circumstances, the low of mass action for reaction kinetics is no longer applicable. Hence modelling at molecular resolution will require a stochastic approach. Further at the level of molecular resolution, molecular activities are discrete, thus requiring formalisms capable of simulating discrete time steps. Hence, a discrete event simulator is required to model discrete biomolecular activities in space and time. Further the formalism should also be able to simulate at the individual molecular level to represent molecular behaviour and intracellular spatial heterogeneity (Ridgway, Broderick et al. 2006). Moreover, at an individual level the rules are qualitative, however a quantitative approach is required to analyse ensemble activities of biomolecules across space and time. Since a multi-level approach is required to model the adaptive dynamics from molecules to cell, the formalism should be able to structure hierarchically and be compositional into functional units from the bottom-up. A suitable framework will be required to capture and analyse the emergent behaviour of biomolecular interactions.

3.2.4 Non-Functional Requirements

The framework should be *extensible* to allow consideration of future expansions of biomolecular representations and their rules. The model should be *scalable* in order to meet computational demands, when implementing large scale models of biological cells containing billions of biomolecules. The model should be *interoperable* with

existing modelling tools, as this will facilitate exchanging model descriptions and testing results.

3.3 The Scope of Systems Biology

While biomolecular science studies individual biomolecules with the aim of revealing how molecules function, systems science aims to predict the consequences of a particular molecular mechanism on the whole organism. However, molecular sciences have become one of the most successful branches of science, by characterising the molecular basis of life for a diverse number of organisms. Molecular bioscience uses reductionist approaches, which initially give prominence to the overall behaviour of systems, and progressively identifies and explores the constituents via decomposition, to characterise the underlying functions of the constituent biomolecules. However, understanding the constituent is necessary but not sufficient for system-level understanding, and a quantitative reconstruction of a system with its constituents, is required. Systems science utilises reconstruction approach to study system wide phenomena. One of the aims of systems biology is to understand biological phenomena, which emerge from complex interactions that occur within and between the levels of the biological organisation strata. Hence, by determining how a function arises, due to dynamic interactions of constituents, systems biology addresses the missing links between molecules and physiology (Bruggeman and Westerhoff 2007). The systems biology approach utilised for studying the biological organisation strata will determine appropriate methodologies for multi-level representation of biological phenomena.

3.3.1 Top-Down Systems Biology

A top-down approach to systems biology identifies "*molecular interaction networks on the basis of correlated molecular behaviour observed in genome-wide 'omics' studies*" (Bruggeman and Westerhoff 2007). It gives insights via inductive reasoning, reasoning from detailed facts to general principles. The models based on this approach are phenomenological. They are not based on mechanisms and, mostly, do not integrate knowledge about relationships between molecules. The top-down approach, which has emerged as a new and dominant method for systems biology,

identifies patterns in molecular interactions, which underlie system behaviour. Moreover, this approach is utilised for cellular systems which have not yet been characterised to a considerable mechanistic detail, and in which much remains to be discovered (Bruggeman and Westerhoff 2007). The major strengths of this approach is that it is genome wide and incorporates most 'omics' technologies.

3.3.2 Bottom-Up Systems Biology

A bottom-up approach to systems biology gives prominence to functional units and studies, the mechanisms through which functional properties arise in interactions of constituents (Bruggeman and Westerhoff 2007). It gives insights via deductive reasoning, reasoning from general principles to a particular observation. This can reveal functional properties, which emerge from the lower levels of biological or cellular organisation, which have been characterised to a high level of mechanistic detail (Bruggeman and Westerhoff 2007). The main goal of this approach is to combine biochemical process models into a global scale representation of biological systems. Models based on this approach are mechanism-based. Although all bottom-up systems biology studies have a common goal to obtain mechanism-based descriptions from lower levels to higher levels of biological organisation, the resources required for modelling differ with the mechanistic principles used. The problem with a bottom-up approach is computability and scope of its application. Computability depends on what level of abstractions the reconstruction process begins and terminates. The scope is the validation of fundamental molecular processes in living systems, as well as non-living systems and emphasising that at this level no other processes are required (Noble 2008).

3.3.3 Discovering General Principles of Biological System Behaviour

Since systems biology is a science (Westerhoff and Alberghina 2005), it should also aim to discover general principles, which relate to all aspects of cellular organisation. This effort in biology is driven by the fact that different species have many systemic properties and molecular mechanisms in common. "*Such interspecies commonalities lead to general principles that offer predictive power and a fundamental understanding of living systems that transcend single species*" (Bruggeman and

Westerhoff 2007). This approach to systems biology can lead to substantial fundamental insights into the principles, which underlie biology.

In combination with experimentation and theory, modelling (Szallasi, Stelling et al. 2006) remains an integral and important part of systems biology studies. The next section describes available biological modelling approaches.

3.4 Biological Modelling Methodologies

Identifying an ideal modelling approach based on model requirements, still causes a lot of confusion and disagreement amongst modellers (Nestorov, Hadjitodorov et al. 1999). Further a model is only as good as the data available to develop and test it. Although new models are constantly being proposed by modellers, claiming the new model extends the knowledge of a phenomenon or process, these models appear to be unable to cover the full complexity of the real world. However, this has not stopped modellers from developing new models at varying levels of complexity, generality and validity. The increasing success rate of modelling technologies in providing solutions in all areas of modern life is sufficient to justify this progressive development (Nestorov, Hadjitodorov et al. 1999).

3.4.1 Model Development Processes

There is a growing demand for models of biological systems to better reflect biological phenomena. The model development process depends on whether a top-down or bottom-up systems biology approach is adopted. There are two major types of model development processes (Tham 1998 - 2000), driven by the two extremes of feasibility and reality. The first is based on empirically generated data (empirical or data driven models), which facilitates top-down systems biology studies. The second is based on underlying principles governing behaviour of phenomenon or process (mechanistic models), which facilitate bottom-up systems biology studies. Table 3.1 summarises the main differences between these types of development processes.

Table 3.1: The comparison between empirical and mechanistic model development process types (This table is developed from (Tham 1998 - 2000))

	Empirical	Mechanistic
Procedure	1. Collect data from the process 2. Specify the correlation structure between variables 3. Use a numerical technique to find parameters for the structure, such that correlation between the data is maximised 4. Validate model against an 'unseen' data set 5. If model is not satisfactory, go to step 2.	1. Use fundamental knowledge of interactions between process variables to define the model structure 2. Perform experiments to determine parameters of the model 3. Collect data from process to validate the model 4. If model is not satisfactory, go to step 1 and re-examine process knowledge
Advantages & Disadvantages	▪ Depends on availability of representative data for model building and validation ▪ Apart from cause and effect between variables, little else is required in terms of process knowledge ▪ A trial and error approach is adopted ▪ Are feasible in delivering some form of working model ▪ The parameters of data driven models are just numbers encapsulating combined effects, thus it is difficult to attach physical or biological meaning to them.	▪ Does not require much data for model development, and hence is not subject to idiosyncrasy in data ▪ Requires a fundamental understanding of principles governing the process ▪ Can be very time consuming ▪ Provides more realistic predictions ▪ Can conduct more analysis studies ▪ It provides an opportunity to associate meaningful elements.

Empirical model development approaches build predictive models based on 'omics' datasets, which currently lack a necessary comprehensiveness and accuracy in measurements to build realistic models. However they exist due to a need to develop quantitative techniques to make use of these datasets and consider their associated uncertainties (Lee, Gianchandani et al. 2006). These models are problem specific and their applicability is limited to empirical conditions, in which the cause (input) and effect (output) relations were obtained. In contrast mechanistic model development approaches are ideal to design new processes, to troubleshoot pathological behaviours

in systems, or to guide towards fundamental improvements in process operability. Based on model requirements, a mechanistic model development approach is chosen to model the principles governing the biomolecular activities within biological cell. Mechanistic principles are used to model different aspects of reality, which include laws of nature, such as chemical and physical laws, and principles from economic theory, information theory, systems theory, organisational theory and theory of computation. Although these principles emerged to describe different aspects of natural phenomena, they have limitations in their applicability to represent biological phenomena. A combination of these principles will be required to describe biological phenomena, since biology has yet to transform itself into a theory rich science (Wingreen and Botstein 2006) to formulate its own laws. The applicability of a model will depend on the scope of the principles. The more universal the principle is, the wider its applicability will be. Existing modelling approaches are evaluated to identify ideal mechanistic principles, from which a model will be developed.

3.4.2 Hierarchical Modelling

Multi-level modelling is an important part of modelling biological phenomena due to the hierarchical nature of biological organisation. There are various hierarchical modelling methodologies for representing biological phenomena within the extremes of top-down and bottom-up methodologies. Top-down modelling is based on analytic thinking, whereas bottom-up modelling is based on synthetic thinking. All these methodologies have strengths and weaknesses.

A bottom-up modelling methodology is more suitable when extensibility of the model is required. Bottom-up developed models have increased compositionality at the lower level and a greater independence from certain higher level requirements, which constitute the most volatile part of the model (Markus and Andreas 2004). Top-down modelling approaches are ideal for modelling specific biological problems. Since construction of a model starts with a specific biological problem or phenomena, it is more likely to produce a workable model for the scenario being used. However, a model cannot be tested until it is completed. Although a decision to use a top-down or bottom-up modelling methodology is not primarily based on extensibility, it will give added value to a model. Extensibility of a model depends on compositionality and flexibility.

Compositionality depends on an ability to decompose or compose models. While top-down modelling tries to study specific biological phenomena, bottom-up modelling forces the modeller to think in terms of orthogonal and extensible components, because a constructed model must be flexible enough to study unexplainable biological phenomena. Bottom-up modelling of diverse biological phenomena is facilitated by addition of components and reconfiguring their interactions. Bottom-up modelling contributes to separation of concerns, which favours orthogonal extensibility of biological function. Separation of concerns can also reduce complexity in modelling. Top-down modelling contributes to the separation of scenarios (Markus and Andreas 2004). However biological adaptability is dependent on concerns which manifest as functional units.

Flexibility is an important feature of the model since it determines modifiability to meet new future requirements. Since bottom-up modelling is initiated by representation of the constituents and their interactions, its stability depends on the consistency of lower level details. These details, which consist of molecular information, have been well characterised in molecular sciences. Since bottom-up models are not globally controlled, they can virtually be represented at any size without major modifications to underlying architecture. Moreover the basic architecture will remain the same, which gives design flexibility when addressing other applications. In contrast, top-down modelling, based on scenarios, will have to start from scratch to represent new scenarios.

3.5 Existing Biological Modelling Formalisms

Based on the scope of an application, existing modelling formalisms are characterised as deterministic or stochastic models, as discrete or continuous models, as macrosopic, mesoscopic or microscopic models, as quantitative, semi-quantitative or qualitative models, predictive or explorative models, homogeneous or heterogeneous spatio-temporal models (Kell and Knowles 2006). Modelling formalisms can also have a combination of above mentioned characteristics and the chosen selection will depend on the modelling requirements.

Further, there are two major dynamic modelling formalisms based on *in silico* representations of biological phenomena. They are the mathematical formalisms,

which are based on denotational semantics, and computational formalisms which are based on operational semantics.

Table 3.2: The comparison between mathematical and computational model

	Mathematical	Computational
Language	Equations	Algorithms
Semantics	Denotational: meaning of model is represented in equations	Operational: meaning of model is represented as a set of instructions
Basic entity	Transfer function	State machines
Relations	Relates different variables to each other	Relates states to each other
Meaning	Models rate of change	Models cause and effect
Dynamism	The steps performed by executor are abstracted. This hides causal, spatial and temporal relationships between those steps	All relationships between steps are exposed.
System transition	Changes occur in variables' values when the system changes state	Highlights why and how system transitions occur
System behaviour	Abstracts overall system behaviour through equations. Describes the average system's behaviour with continuous state spaces	Precisely describes a system's behaviour with discrete state spaces
Origins	Mathematics	Computing theories

Table 3.2 summarises the main differences between mathematical and computational formalisms (Fisher and Henzinger 2007; Hunt, Ropella et al. 2008). The current modelling tools for systems biology are dominated by mathematical approaches (Coveney and Fowler 2005). Mathematical simulations are solved by using computing resources as a service. Equations are sequential tools which attempt to

model a system, whose behaviour is completely determined by input/output relationships. However, input/output relationships are not suitable for characterising the behaviour of concurrent systems such as biological systems (Corrado 2009). A network topology will be the outcome of equation based models, which provide a sequential representation of the system's dynamics. Further it cultivates sequential thinking. Equation based models with inherent sequential assumptions impact the notion of causality which corresponds to a temporal ordering of events. However in the context of parallelism, causality is a function of concurrency (Corrado 2009). Intracellular biochemical activities are highly concurrent. Network topologies will also have to deal with combinatorial explosions arising from different states of biomolecules, their relations and interactions. Further analysing distribution of causalities at points of convergence and divergence, is an inherent problem in equation based models. It should be realised, that emergence of continuous phase spaces at the aggregation level, which mathematical formalisms use, are actually emergent properties of discrete state spaces at an individual level, which can be represented in computational formalisms.

Various modelling formalisms and tools have been reported in (Gilbert, Fuss et al. 2006; Grima and Schnell 2008; Pahle 2009; Walker and Southgate 2009).

3.5.1 Mathematical Modelling Formalisms

Currently there are many different types of models, using various mathematical formalisms to represent biological systems. However the scope of this discussion is based on modelling strategies, which study intracellular reactions. Based on spatial representation and predictability of intracellular reactions, the models are classified as non-spatial-deterministic, spatial-deterministic, non-spatial-stochastic, and spatial-stochastic models. Population based kinetic models are one type of the widely used mathematical models that treat reacting components as population pools. These include mass action kinetic models which are deterministic and can either represent spatial or non-spatial scenarios, and stochastic kinetic models, which are stochastic and can either represent spatial or non-spatial scenarios. The distinctions between these two approaches are tabulated in Table 3.3. However in individual based stochastic models, resolution of spatial representation is high (single molecular

resolution), and predictability of reactions is stochastic. Further, they assume heterogeneity in molecular distribution and discreteness among molecules.

Table 3.3: The distinctions between mass action kinetic models and the stochastic kinetic models

	Mass Action Kinetic Models	Stochastic Kinetic Models
Model type	Macroscopic	Macroscopic
Focus	Population dynamics	Population dynamics
Population pools	Continuous concentrations	Discrete population
Entity	Indistinguishable mass of identical elements	Indistinguishable mass of identical elements
Spatial representation	Homogeneous distribution No reaction diffusion	Homogeneous distribution No reaction diffusion
Spatial resolution	Low	Low
State of the model	Defined by a concentration of elements	Defined by population number of all species involved
Predictability	Deterministic, - same starting condition same result	Stochastic
Transformation Function	Reaction rate	Probability
Equation	Rate/Differential equation	Chemical master equation
Chemical Kinetics	Rate constant	Reaction constant
Assumptions	Homogeneity, continuum	Homogeneity, discrete

3.5.1.1 Mass action kinetic models

Kinetic models can be created using various methodologies to simulate biological systems. The most commonly used methodology is mass action kinetics, which is based on the law of mass action (Wanner, Finney et al. 2005). These models represent molecular information, as aggregated variables for every distinct molecular population. The state of a model is defined by population of its molecules at any particular time. The main assumptions of this model are that reactants are well mixed,

homogeneously distributed in space and large numbers of chemical species are present, which can be represented as concentrations varying on a continuous scale. However these assumptions do not comply with cellular systems, since their characteristics are far from continuous and deterministic. Dynamic behaviour of a cell is mostly expressed using ordinary differential equations (ODE), which are implicitly non-spatial and deterministic. When adding spatial dimensions to these approaches, ODEs are transformed into corresponding partial differential equations (PDE). **E-Cell** (E-Cell Project 2009) is an international research project, which aims to model and reconstruct biological phenomena *in silico.* This model utilises ODEs, in the form of rate equations (RE) to represent intracellular dynamics. In contrast, **Virtual Cell** (NRCAM 2009) utilises PDEs with a finite volume method in the form of reaction-diffusion equations (RDE), to represent reaction and diffusion rates of molecules, in its spatial simulation framework. A cell's spatial structure is depicted as compartments in this framework. These compartments are further subdivided into sub-volumes via a mesh-generator. Although a finer time step and sub-volume size can produce more accurate solutions, it will require more computational resources. Although PDEs are known to be one of the most computationally scalable spatial simulation algorithms, their deterministic nature cannot accurately represent intracellular noise (Takahashi, Arjunan et al. 2005). These models are deterministic, continuous, macroscopic, quantitative and can represent spatial homogeneity or heterogeneity. However they are not hierarchical or compositional.

3.5.1.2 Stochastic kinetic models

These models also represent molecular information as aggregated variables for every distinct molecular population. Further these models treat molecular population pools, as discrete populations and map mass action reaction rates onto probabilities to generate a stochastic formulation of chemical kinetics (i.e. rate equations) in the form of a chemical master equation (CME) (Gillespie 2007). CME is computationally simulated, using the Gillespie algorithm (GA), which is also known as the Stochastic Simulation Algorithm (SSA). This algorithm uses probabilities, called reaction constants, which are derived from chemical kinetics rate constants, to determine whether a reaction occurs. The algorithm is initiated by specifying molecular population numbers, and reaction constants for possible reactions of respective molecules. Random numbers will determine duration of elapsed time, and what

reaction occurred within that interval. Finally, molecular population numbers are adjusted along with dependent probabilities. This cycle is continued during simulation. Although these models take account of the discrete and random nature of chemical reactions, they still do not consider distinction between individual molecules. Reacting species are still represented as population pools and are spatially homogeneously distributed. Cellular systems exhibit complex spatial heterogeneity and the relative positioning of biomolecules with respect to one another is fundamental to the organisation of biological cells. Algorithms have been developed from SSA to address the problem of spatial heterogeneity in reaction diffusion systems. They utilise reaction-diffusion master equations (RDME), to represent reaction and diffusion probabilities of molecules in their formalism. Simulation software such as **SmartCell** (Ander, Beltrao et al. 2004) and **MesoRD** (Hattne, Fange et al. 2005) has adopted this strategy, so it can tackle some issues of spatial heterogeneity. However they are incapable of simulating at molecular resolution. These models are deterministic, continuous, macroscopic, quantitative and can represent spatial homogeneity or heterogeneity. However they are not hierarchical or compositional.

Intracellular reaction kinetics can be modelled using RE, RDE, CME and RDME. However an appropriate choice of formalism for a particular study depends on concentrations, distance travelled by molecular species during their lifetime, size of the intracellular space, in which a reaction occurs and the extent of macromolecular crowding in the stipulated region (Grima and Schnell 2008). Although these macroscopic approaches represent molecular information at an aggregation level, they are unable to represent spatially heterogeneous populations at molecular resolution and have an inherent problem of combinatorial complexity, caused by biomolecules, which can assume multiple distinct states (post-translational modifications, ligand occupation or conformational states). Also three dimensional structures of biomolecules are represented as conformational states, which in turn will determine their chemical activity. In addition molecules can aggregate to form a complex, which is also considered as different states of the participating biomolecular species. These issues can only be addressed by models capable of modelling at the level of molecular resolution, where the concept of concentration is not applicable and effects of stochastic activities dominate the system behaviour.

3.5.1.3 System Dynamics

System Dynamics (SD) is a 'whole system' modelling approach, which is used to understand the overall dynamic behaviour of complex systems over discrete time steps (North and Macal 2007). This involves identification of the main state variables, which define behaviour of the system and relating these variables through difference equations or differential equations. SD is a graphical representation of the population (aggregation) based mathematical model, where complex systems are modelled using feedback loops, stocks (state variables) and flows (time delays) to describe the nonlinear behaviour of the whole system (Wikipedia Contributors 2009c). SD is a top-down approach and has been applied in population, ecological and economic systems. Although they are ideal to model horizontal causality of complex systems, they lack an ability to model self-organisation, emergence, and upward and downward causal levels that are prevalent in complex hierarchical systems. These models are deterministic, continuous, macroscopic, quantitative, and can represent spatial homogeneity. However they do not support composition or hierarchical structuring.

3.5.2 Computational Modelling Formalisms

Approaches based on computing introduce systems, hierarchical and concurrent thinking to the study of biological phenomena. Various formalisms from computer science are contributing towards gaining a deeper understanding of biological function (Brent and Bruck 2006). However their applicability is mostly limited to qualitative modelling of biological phenomena, and they are effective when modelling with incomplete quantitative data. These models are formal models, primarily based on operational semantics. Numerous terms have been assigned to this category of models, including executable biology, programming biology and algorithmic systems biology (Laursen 2009).

These formalisms can have a combination of characteristics. They can represent concurrency in the form of synchronous state changes, where state machines change state simultaneously, and/or in the form of asynchronous state changes, where some state machines change state independently. Further, they can be deterministic due to synchronous state transitions, be non-deterministic due to asynchronous state transitions, or stochastic due to probabilistic state transitions. In the context of structuring, they can be compositional when the behaviour of the system is specified

by interacting modules, and/or hierarchical if modules can be utilised as reusable building blocks to represent higher levels of the model (Fisher and Henzinger 2007).

3.5.2.1 Boolean networks

Boolean networks are used in qualitative modelling of biological networks to represent individual biomolecules as in an active or inactive state, while intermediate states are neglected. Hence a node can only have two states. The activation states of all biomolecules at a specific step, will determine activation states for the next step. A Boolean network can simulate causal and temporal relationships amongst the activation of different biomolecules (Fisher and Henzinger 2007). Since it uses a network topology, it models sequential dynamics in biomolecular networks. These models are deterministic, use discrete time steps, can represent individual biomolecules (mesoscopic), and qualitative. The main limitation of Boolean networks is that they do not support composition and hierarchical structuring of models, and are not designed to represent intracellular space.

3.5.2.2 Petri nets

One of the strengths of Petri nets is modelling of concurrency in biological systems. It is an established technique for modelling distributed systems (Fisher and Henzinger 2007). Petri nets are graphs, which contain two kinds of nodes - *place nodes*, which represent the resources of system, and *transition nodes, which* represent events that change resource states. Nodes can be in many states. Petri nets have been used for qualitative modelling of concurrent behaviour in biochemical networks, such as metabolic pathways and protein synthesis. While Boolean networks are deterministic, Petri nets can also be nondeterministic, stochastic or both. However, stochasticity is imposed rather than arising from underlying interactions. Further, Petri nets are discrete, mesoscopic and qualitative. The main limitation of Petri nets is that they do not support composition of larger models from smaller ones and are not designed to represent intracellular space. Intracellular space can only be represented explicitly, when different place nodes contain, the same biomolecular species, representing different compartments. However, this is analogous to extending an ODE formalism to a PDE formalism and makes extensibility a daunting task. Moreover, since Petri nets are based on network topology, they suffer from combinatorial explosion. Petri nets do not support composition or hierarchical structuring. However Petri net

formalism has been extended with Hierarchical Petri nets to support composition of more complex models (Materi and Wishart 2007).

3.5.2.3 Particle based formalism

Particle based formalisms utilise biological entities, as individual passive objects using a centrally controlled thread of execution. They have been used for spatial stochastic simulation at molecular resolution and in Molecular dynamics to model at atomic resolution.

In **particle based stochastic approaches** every reacting molecule is represented individually and reactions between molecules occur in a probabilistic manner. A mesoscopic model treats biological entities as individual objects. The state of the model is defined by aggregated states of all particles in the model. These models are used for reaction diffusion systems, such as metabolic pathways and signal transduction systems. Most tools that are based on this approach use Brownian dynamic algorithms to simulate the Brownian motion, and use Monte Carlo algorithms to simulate reaction events (Tolle and Le Novere 2006). Monte Carlo algorithms compute an outcome by generating random numbers, which are compared to a probability calculated from reaction rates. This framework is adopted with variations in representing space by **MCell** (Stiles and Bartol 2001), **Smoldyn** (Andrews and Bray 2004), **CyberCell** (Broderick, Ru'aini et al. 2005), **ChemCell** (Plimpton and Slepoy 2005), **Cell++** (Sanford, Yip et al. 2006), **GridCell** (Boulianne, Al Assaad et al. 2008) and **StochSim** (Le Novere and Shimizu 2001). These particle based stochastic approaches were developed, when spatial heterogeneity and stochasticity itself became an objective of the research (Pahle 2009). Factors that contribute to spatial heterogeneity and stochasticity are low molecular numbers, macromolecular crowding, spatial constraints and intracellular noise. The aim of these approaches is to explicitly model intracellular kinetics in the presence of factors, which contribute to stochastic behaviour.

Molecular dynamics (MD) is a particle based approach and plays an important role in modelling intra-dynamics of molecules by giving insights into molecular motion on an atomic scale (van Gunsteren, Bakowies et al. 2006). MD is a multidisciplinary method and is a specialised discipline of molecular modelling based on statistical mechanics. Fundamental physical rules regulate the motions of all molecules constituting a cell. The computational requirement of MD simulation increases with

the number of interacting atoms. Although MD is the most accurate and fundamental approach in the context of representing every physical parameter applicable at the level of the atoms, it cannot be used to simulate whole cell systems, due to the presence of a large number of atoms among biomolecules constituting a cell. A range of molecular dynamics software is listed in (Wikipedia Contributors 2009b).

Particle based models are stochastic, discrete, mesoscopic, quantitative and can represent intracellular spatial heterogeneity. Further they do not support composition or hierarchical structuring.

3.5.2.4 State charts

State chart formalism is naturally suited to object specification, which have well characterised internal behaviour. In conjunction with object model diagrams, they provide a graphical representation of the dynamics of objects, using states, transitions, events and conditions (Avital, Jasmin et al. 2008). Many biological phenomena have been modelled using state chart formalisms (Cohen and Harel 2007). The state chart formalism can be nondeterministic, stochastic or both, use discrete time steps, can technically represent biological entities, such as atoms, molecules or cells, can be qualitative or quantitative and can represent intracellular spatial heterogeneity. State charts models can be structured hierarchically and compositionally.

3.5.2.5 Cellular automata

Cellular Automata (CA) formalism has been used to model natural phenomena, which include physical systems and biological systems, such as molecular, bacterial, cellular and ecological models (Walker and Southgate 2009). CA can be used to model both temporal and spatio-temporal processes, using discrete time and/or spatial steps (Materi and Wishart 2007). Further, an extended formalism based on Dynamic Cellular Automata (DCA) has incorporated stochasticity by permitting random motion to represent molecules in a cell. **SimCell** has adopted the DCA formalism (Wishart, Yang et al. 2005). Basic CA formalism is deterministic, discrete, mesoscopic, can be qualitative or quantitative, and can represent intracellular spatial heterogeneity. CA models can be structured compositionally to support hierarchical structuring.

3.5.2.6 Agents based formalism

An Agent based formalism is similar in concept and design to Dynamic Cellular Automata. Since agents exist in an environment, they are allowed to interact with each other and their environment in space and time, based on pre-defined rules. The rules define the behaviour of entities by the diverse states that an entity can be in during its life time. The motion can be directed or random. The rules can be simple or highly complex. In contrast to CA models, agent based formalisms do not require spatial grids or synchronised time steps (Materi and Wishart 2007). Agent based formalisms are widely used to model complex systems in areas such as sociology (Epstein 2009), business (North and Macal 2007), economics (Buchanan 2009; Farmer and Foley 2009) and ecology (Grimm, Berger et al. 2006). In contrast, use of agent based formalism to model biological complexity, when a range in scale is needed from molecules to organisms, is still in its infancy. Currently agent based formalisms are emerging as solutions for systems biology (Webb and White 2006; Thorne, Bailey et al. 2007). Most agent based formalisms are represented as Multi-Agent Systems (MAS) (Walker, Southgate et al.; Cannata, Corradini et al. 2005; Merelli, Armano et al. 2006; Catholijn and Jan 2007; Sutterlin, Huber et al. 2009). **Agent Cell** (Emonet, Macal et al. 2005) utilises an agent based formalism at the cellular resolution level and stochastic approaches to model the intracellular dynamics. It has used a top-down approach to model cellular behaviour.

Agent based formalisms are stochastic, use discrete time steps, can technically represent biological entities such as atoms, molecules or cells, can be qualitative or quantitative, and can represent intracellular spatial heterogeneity. Agent based models can be structured hierarchically and compositionally.

3.5.2.7 Process calculi

Process calculi approaches emphasise significance of interactions amongst biomolecules, as the driving force for biochemical processes (Kwiatkowska and Heath 2009). Here execution of the model is defined via a sequence of events. There are many variants of process calculi, that have been used as a modelling language for molecular interactions, such as π-calculus, ambient calculus and brane calculus (Fisher and Henzinger 2007). SPiM is a modelling and simulation tool based on stochastic π-calculus, which uses the Gillespie algorithm (Gillespie 2007) as a basis of

its computational engine to simulate biochemical systems (Gilbert, Fuss et al. 2006). Currently process calculi models are being developed by a programming biology group at Microsoft Research, Cambridge Labs (Cardelli 2005). There is work being undertaken on bioware languages for systems biology. This will represent the structure and function of biological systems via formal languages (Cardelli 2005). Further Microsoft Research – University of Trento Centre for Computational and Systems Biology have engaged in an approach known as Algorithmic Systems Biology (Corrado 2009), which aims at devising proper abstractions of living systems, in order to capture their intrinsic concurrency, causality and probabilistic nature into algorithmic descriptions that can be executed, analysed and simulated in computers.

The process calculi formalism can be deterministic or stochastic, use continuous time, can technically represent biomolecular processes, is qualitative and cannot represent intracellular space. Process calculi models can be structured compositely, but are unable to represent hierarchical structuring.

3.5.2.8 Scenario based formalisms

These include Live Sequence Charts (LSCs), which are interobject in nature and are appropriate for describing behavioural requirements (Avital, Jasmin et al. 2008). LSC is a visual formalism for specifying sequences of events and message passing between objects. Behaviours are specified as scenarios of events and actions, with diverse possibilities. LSC uses two types of charts, namely, universal and existential. Universal charts are used to specify restrictions by constraining certain behaviours. Existential charts specify sample interaction between a system and its environment. The scenario based formalism can be deterministic, use discrete time steps, be phenomenological (represent cellular level behaviour), qualitative and cannot represent intracellular space. These models cannot be structured hierarchically and compositionally.

3.5.3 Hybrid Modelling Formalisms

These models combine mathematical and computational formalisms. These frameworks integrate variables, which span discrete and continuous domains. The discrete component of a model utilises a computational formalism, and the continuous component of the model utilises a mathematical formalism. Discrete variables are

controlled by the changes in discrete states, which can depend on continuous variables. Further, a change in continuous variables is governed by transformation equations, such as differential equations, which depend on discrete states. By merging mathematical formalisms and computational formalisms, hybrid models tend to bridge the gap by incorporating characteristics unique to mathematical and computational formalisms. These models are appropriate for modelling relationships between substances that change overtime. **CompuCell** (Izaguirre, Chaturvedi et al. 2004) utilises a hybrid approach, where the cell is modelled as objects and the intracellular and intercellular behaviour is incorporated with differential equations. Moreover the basic Petri net formalism has been extended to deal with continuous variables, thus giving rise to Hybrid Petri nets (Materi and Wishart 2007).

3.6 Evaluation of Modelling Methodologies and Formalisms against Model Requirements

Since modelling multi-scale adaptive dynamics from molecules to cell requires a mechanism based description of functional properties that emerge as a result of molecular interactions, the study follows a bottom-up systems biology approach. This approach utilises a mechanistic model development process, where the structure of the model depends on the mechanistic principle adopted. Further a hierarchical representation of the intended study is based on a bottom-up methodology. This is because the aim of the study is to understand how biological cells dynamically adapt to multiple objectives concurrently, facilitated by constituent biomolecular activities, which require traversing from lower level molecular resolution to higher level cellular resolution. Multi-objective topology provides a concurrent and hierarchical view of biological systems, whereas network topology provides a sequential and horizontal view of biological systems. However, mathematical models, which use network topology, are designed to model at population/aggregation level and are unable to model at level of molecular resolution.

Moreover the state of the system in mass action kinetic models, stochastic kinetic models and particle based stochastic models is assessed based on the abundance of reactants. They analyse population dynamics in intracellular reactions. However the actual state of a cell should be represented by levels of native biomolecular activities

rather than abundance of biomolecules. Native biomolecules that are performing (i.e. engaged in activities rather than being idle or inactive) represent an effective population which is mostly a subset of the available population in a cellular environment. Although the states of chemical systems are described in terms of concentrations of molecules, it is the chemical activities of the molecules, which provide the most accurate description of a system. In Chemistry the use of concentration as an approximation to chemical activity is based on the assumption, that the difference between concentration (actual population) and chemical activity (effective population) is insignificant, due to the presence of high populations of molecules and a negligible percentage of inactive molecules found in conventional chemical systems. However in biological cells, where functional products are complex molecules and only certain states out of all possible states will have an ability to perform the intended activity, there will be a significant deviation between actual population and effective population. Due to very low populations this disparity is amplified further, which means the state of a cell is misjudged. Activities of various biomolecular species will contribute to the internal organisation of a cell. Apart from contributing to molecular crowding, biomolecules that merely occupy a cellular environment will have minimal contributions to the performance of a cell. Since particle based stochastic models use Monte Carlo algorithms to compute an outcome by generating random numbers, which are compared to a probability, calculated from reaction rates, they do not distinguish between processing time requirements for different activities at the level of molecular resolution. This is a property of reactants, especially the enzymes, which perform most of the chemical transformation in a cell. Although reaction rates inherently, consider the rate of association, the rate of dissociation and the rate of catalysis into a single expression, this distinction must be explicit, when modelling at the level of molecular resolution. Reaction rates represent ensemble averages of reactants of the chemical system. Hence, deliberately altering the interactions of reactants based on reaction rates at the level of molecular resolution will basically reproduce an observed behaviour. Further, the inherent constraints of biomolecular activities, such as processing time, energy requirements and thermodynamic constraints that effect the internal organisation of the cell are not considered in kinetic models, because they model rate of change in chemical systems.

The choice of modelling formalism is based on the ability to meet model requirements. For hierarchical modelling between molecular resolution and cellular resolution, models can be built at the abstraction level of individual biomolecules, at population/aggregation level and at cellular level representing a biological cell as whole, Mathematical formalisms work at population level, by identifying key system level aggregate variables, that define the behaviour of a system. They can either represent a population of molecules, cells, or organisms at a time. Simultaneous integration of equations across levels is not feasible. Although mathematical formalisms are ideal to model sequential dynamics and horizontal causality of complex systems, they lack the ability to model self-organisation, emergence, and upward and downward causal levels that are prevalent in complex hierarchical systems. Individual based models used by most computational formalisms, can represent biomolecules, cells or organisms as atomic entities. At cellular level they represent a cell as a whole and decompose it into components representing sub-systems. At the molecular resolution they can represent the causal links amongst biomolecular events. Table 3.4 shows that an individual based approach that is able to represent space is suitable for representing biomolecules and meets the model requirements listed in Section 3.2.1. These formalisms include state charts, CA, agent based formalism and the particle based formalism. These representations are intra-object in nature and utilise an object modelling approach that facilitates reusability. Particle based formalisms use passive software objects to represent entities. In contrast agent based formalisms use active objects to represent entities. The important differences between objects as represented in particle based models and agents are listed in Table 3.5 (Odell 2002).

Table 3.4: Comparison of modelling formalisms based on model requirements

Modelling Formalisms	Molecular resolution	Multi-scale	Emergence	Stochastic	Autonomy	Self-organisation	Concurrency	Mechanistic Principles
Mass action kinetic models	No	No	No	No	No	No	No	Law of mss action
Stochastic kinetic models	No	No	No	Yes	No	No	No	Law of mss action Probability
System dynamics	No	No	No	No	No	No	No	Rate Law
Particle based stochastic models	Yes	Yes	No	Yes	No	No	Yes	Probability Diffusion
Molecular Dynamics	Yes	Yes	No	Yes	No	No	Yes	Statistical mechanics
Boolean networks	Yes	No	No	No	No	No	No	Boolean Logic
Petri nets	Yes	No	No	Yes	No	No	Yes	Logic
State Chart formalism	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Automata theory
Cellular Automata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Automata theory
Agent based formalism	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Automata theory
Process calculi	Yes	No	No	Yes	No	No	Yes	Automata theory

Table 3.5: Comparison between objects and agents

Characteristics	Particles (Passive Objects)	Agents (Active Objects)
Degree of autonomy	▪ Can exhibit control over its states ▪ Cannot exhibit control over its behaviour ▪ The decision whether to execute a method lies within the object that invokes the method	▪ Can exhibit control over its states ▪ Can exhibit control over its behaviour ▪ The decision lies within the agent that receives the request
Flexibility	Have no flexibility in addressing its behaviour	Can flexibly address its behaviour (i.e. reactive, proactive and social)
Thread of control	Have a centrally controlled thread	Have their own thread of control

To model uncertainty, concurrency, self-organisation and emergence, which are prevalent in intracellular biochemical activities, a formalism based on active objects is required, since they have a high degree of autonomy. Although the rules, which remain constant at the physiological timescale, define what a particular species of biomolecule can perform, the uncertainty involved in, when and where these rules are executed by redundant members of a species cause emergent behaviours in a cell. Hence, this requires autonomy at an individual level. Formalisms that support autonomy are state charts, agents and CA. State charts have been used for top-down reconstruction of biological systems. However, when considering the flexibility of representing directed and random motion of biomolecules, ability to represent grid or continuous space, ability to represent synchronous and asynchronous time steps, ability to represent interaction between molecules and their environment, ability to dynamically schedule molecular events and duration of those events, agent based formalism provides a more appropriate solution than Cellular Automata formalism. Further, general purpose agent based tool kits are widely available compared to the situation for the other two formalisms.

3.6.1 Agent Frameworks Applicable for Modelling Cellular Phenomena

There are various agent oriented conceptual frameworks for modelling biological phenomena. Most agent based formalisms are represented as Multi-Agent Systems (MAS) (Walker, Southgate et al.; Cannata, Corradini et al. 2005; Merelli, Armano et al. 2006; Catholijn and Jan 2007; Sutterlin, Huber et al. 2009). However agent based formalisms are also represented as Complex Adaptive Systems (CAS). Agents are basic building blocks for both CAS and MAS frameworks. The distinguishing feature between a CAS and MAS is that CAS focuses on system level properties and features, while MAS focuses on individual level features. In MAS the system is composed of multiple interacting adaptive agents, whereas in CAS the agents, as well as the systems, are adaptive. CAS has a high degree of adaptive capacity, which gives resilience in the face of perturbation.

MAS are utilised for top-down approaches to resemble the macro models, where a whole population of entities are divided into homogeneous sub-populations (Cannata, Corradini et al. 2005). This approach attempts to simulate changes in the average characteristics of a whole population. By contrast in bottom-up approaches, which resemble meso models, the spatially distributed entity population is heterogeneous and consists of distinct agents with unique state and interaction behaviour that evolve with time. Further, agents can be combined to create a society of agents that would facilitate in creating an environment for artificial life (Kyung-Joong Kim 2006). Bottom-up approaches are widely used for simulations of self-organisation and emergent phenomena (Schut 2007). Swarm/Collective Intelligence (CI) which inherently considers self-organisation and emergence uses biologically inspired approaches to study collective behaviour in self-organising systems (Kennedy and Eberhart 2001). Swarm systems are typically made up of a population of simple agents interacting locally with each other and with their environment. There is no centralised control structure dictating how individual agents should behave. Local interactions between such agents mostly lead to emergence of global behaviour. Further, in the CI framework, agents as well the system have an ability to adapt.

In Aspect oriented approach, MAS representation is used to capture the emergent behaviour that arise from diverse interactions of multiple agents (Palmer,

Kirschenbaum et al. 2005; Linda, Daniel et al. 2006), because it is complex to encapsulate emergence in a conventional programming language. An object oriented approach is capable of encapsulating attributes and behaviour of a single agent within a class instance. However emergent properties do not exist within single agent behaviour, but emerge from spatio-temporal interactions of many agents. Hence, to encapsulate emergence, an approach that can represent modularity across a set of objects and a set of object interactions is required. Further, it requires encapsulation of a set of classes and class method invocations (Palmer, Kirschenbaum et al. 2005). An object oriented approach can not support such modularity. An Aspect oriented approach can support encapsulation of concerns that cross the object oriented class boundary.

Further, there are middle-out approaches that use both of these approaches, which can be demarcated at a particular level in the hierarchy. A holonic approach also uses top-down and bottom-up approaches in a totally fused manner and captures the benefits of both approaches (Rodriguez, Hilaire et al. 2007).

A framework, in which agent based formalism can be implemented is needed. This will depend on the mechanistic principle used to model the biological cell.

3.6.2 Selection and Implementation of an Appropriate Approach

A cellular environment represents both biomolecules and their activities, which contribute to the self * properties, such as self-regulating and self-awareness of a cell. The activities cause direct and indirect influences amongst various species of native biomolecules, which facilitate in the self regulation of cellular processes. Agent based formalism is used in the wider framework of Collective Intelligence to model self-organisation and emergence that occurs due to diverse biomolecular activities. This approach facilitates analysis of global effects of changes in behavioural rules imposed on diverse biomolecular species, where the effects of rules are amplified due to redundant members of a biomolecular species. Representation of agent based formalism at the level of molecular resolution also addresses the heterogeneous nature of a cellular environment and the existence of very low numbers of some functional products. The core of a model is driven by the principles of Swarm/Collective Intelligence, which capture the inherent characteristics of a cell such as adaptability, robustness and efficacy with no external supervision (Schut 2007). Modelling and

simulating these characteristics is essential to truly understand the mechanism by which intracellular solutions emerge via various biomolecular activities to meet the adaptive requirements of cells. This insight is essential to understand the transformation between normal and pathological processes in cellular systems. Some of the noteworthy properties of Collective Intelligence systems are adaptivity, emergence, global-local order, interaction, rules, redundancy, robustness and randomness (Schut 2007).

Since organisational behaviour within a cell cannot be directly observed or empirically measured, it requires a simulation framework, such as CI, which can represent native biomolecules, capture the results of their activities and provide a means to evaluate these results. Analysis of cellular behaviour should be based on chemical activities of molecules rather than their abundance, since activities provide an accurate description of a chemical system, where performances of the functional products are analysed based on their level of activities.

The challenge of implementing agent based formalisms lies in specifying how agents behave, and in choosing the rules they use to make decisions. This is mostly done by common sense and guesswork, which means it is only sometimes sufficient to model real behaviour. Further, an attempt to model all the detail of a realistic problem, can quickly lead to a complicated simulation, where it is difficult to determine causalities amongst the behaviour. For agent based modelling to be useful, the development of a model must proceed systematically and avoid random assumptions, have careful grounding, test each part of the model against reality, and introduce additional complexity, only when it is required. If properly done an agent based approach can provide an exceptional understanding of the emergent properties of the interacting parts in complex phenomena, where intuition fails to give an understanding (Farmer and Foley 2009).

3.7 Feasibility of Meeting Model Requirements

The aim of this section is to evaluate the feasibility of implementing and developing a Collective Intelligence based cell modelling environment with available biological data sources, programming environments, platforms and computational advances.

3.7.1 Resource Requirements and Availability

3.7.1.1 Biological data sources

Models developed by application of any empirical or mechanistic model development process, will only be as good as the data available to develop and test them. Hence availability of a collection of quantitative, high-quality and validated datasets that reflect the organisation of functional products, into functional units; with dynamic interactions between these units, which control and perform their diverse and complex biological functions (i.e. syntax of biological information), is critical to the success of systems biology (Aebersold 2005). However current techniques that have successfully accumulated large amounts of detailed information for established genetic approaches are not sufficient for systems biology. "*Systems biologist will have to develop new quantitative technologies that are capable of systematically measuring the dynamics and ordering of, as well as relationships and interactions between the molecules that constitute biological systems*" (Aebersold 2005).

Biological data sources, that have been accumulated over the last few decades range from molecular biology data, 'omics' data sets (Joyce and Palsson 2006; Katherine Hollywood 2006), to biochemical and biophysical data. Identifying appropriate and reliable data sources is a challenge, due to abundance of inconsistent information found globally. There are currently 1170 molecular biology databases distributed globally (Galperin and Cochrane 2009). They are broadly categorised as Nucleotide sequence databases, RNA sequence and structure, Protein sequence databases, Structure databases, Genomics databases (non-human), Metabolic enzymes and pathways, Human other vertebrate genomes, Human genes and diseases, Microarray data and other gene expression databases, Proteomics resources, other molecular biology databases, Organelle databases, Plant databases, and Immunological databases. However the required data source will depend on the systems biology approach used. A review on data sources required for integrative systems biology has been conducted (Ng, Bursteinas et al. 2006), which identifies data for top-down systems biology studies. However for the selected approach, information compatible at the level of molecular resolution will be required.

Modelling biomolecular activities and the stochastic nature of intervals between activities will require information at a molecular resolution. This process needs

information about pre-conditions and post-conditions for the activities, and the time and energy requirements of the activities, information about the stochastic nature of spacetime intervals between activities, such as abundance and diffusion constants of biomolecules. Although there is limited information on concentration and the abundance of molecules inside a cell, various experimental evidence suggests that many intracellular biochemical reactions involve reactant molecules at nanomolar (nM) concentration scales (Grima and Schnell 2008). Based on absolute numbers, a concentration of 1nM roughly corresponds to 2500 molecules in a sphere with a diameter of 20 micrometers, which is equivalent to an average mammalian cell. Moreover 100 nM corresponds to 100 molecules in a typical E.coli bacterial cell (a cylinder - 2 micrometers in length and 1 micrometer in diameter). The relationship between concentration and absolute molecular numbers in specific volume is tabulated in Appendix F-1. In terms of average intermolecular distance, 1nM corresponds to an average distance equal to 1.47 micrometer, while 100 nM corresponds to a distance approximately equal to 0.32 micrometer. The relationship between concentration and average molecular distance is tabulated in Appendix F-2. Although most metabolites have concentrations in the nM range there are some glycolytic metabolites having concentration in the order of millimoles (mM) (Grima and Schnell 2008).

The widely used thermodynamic property in biochemistry is Gibbs free energy (which is a measure of the potential of a chemical system to do work, and can be used to model physical constraints. Thermodynamic constraints will determine the probability of molecular activities occurring. This information is available at (Alberty 2005; Alberty 2009). Thermodynamic information on biochemical reactions is stored as the Standard Gibbs free energy of formation ($\Delta_f G^\circ$) and Standard Enthalpy of formation ($\Delta_f H^\circ$) of the molecular species involved in biochemical reactions. These sources are standard and reliable, and are the most efficient way to store thermodynamic information. Further there are tools to calculate the above values at any desired pH, ionic strength and temperature, and to calculate Standard Gibbs free energy of reactions ($\Delta_r G^\circ$), the Standard Enthalpy of reactions ($\Delta_r H^\circ$), and the equilibrium constants for reactions at any desired temperature. Population based chemical kinetic data such as the k_{cat}, rate of transcription and translation, can be transformed into compatible molecular level information. Duration of molecular

activities such as an enzyme's processing time/ turnover cycle, which are μs – ms events, should be derived from enzyme turnover numbers which can be obtained from an enzyme database such as BRENDA (Chang A., Scheer M. et al. 2009). Time required for translation of various proteins can be calculated from the peptide length of proteins and rate of translation, e.g. in Prokaryotes it is approximately 15 amino acids per second. This information can be obtained from protein databases, such as UniProt (Galperin and Cochrane 2009). The time required for transcription of RNA can be calculated from the lengths of the genes, and the rate of transcription, e.g. rate of transcription in Prokaryotes is approximately 45-50 nucleotides per second and rate of DNA polymerisation is approximately 833 nucleotides per second. This information can be obtained from nucleotide sequence databases (Galperin and Cochrane 2009).

From a biological perspective, energy requirements are characterised in terms of Adenosine Tri Phosphate (ATP), which is known as the currency of energy in a cell. Energy requirements for various molecular activities can be obtained from standard biochemistry text, such as (Stryer 1988) (namely, the energy cost for synthesis of a protein with N number of amino acids is 4N ATPs). The average life span of functional products can be calculated from half lives of functional products. A typical half-life of mRNA is 2 to 3 minutes in Prokaryotes and hours to days in Eukaryotes (Stryer 1988). Diffusion constants for various biomolecules can be calculated from their molecular mass or obtained via empirical observations. For typical size proteins in Prokaryote the diffusion constant is $5m^2s\text{-}^1$. Error frequencies for various biomolecular activities are also obtainable from standard biochemistry text such as (Stryer 1988). For example, the error frequency for Protein synthesis is 10^{-4} per amino acid residue.

3.7.1.2 Programming environments for managing modelling resources

Programming environments differ for bioinformatics, computational biology and systems biology. Bioinformatics focuses on data analysis and management, whereas computational and systems biology focuses on systems modelling and simulation. Currently a suitable language for bioinformatics does not exist. Hence toolkits in several different languages have been written to provide bioinformaticians with options to choose the best language for a specific job. Libraries of routines and data

objects for bioinformatics exist in C (EMBOSS's Ajax Library), C++ (NCBI's C++ toolkit, TIGR), Java (BioJava, caBIO), Perl (BioPerl), Python (BioPython), R (BioConductor), Ruby (BioRuby) and Lisp (BioBike). The BioSQL project presents a generic relational model for representing biological sequence objects and features, their annotations and ontologies are independent of an actual data source. The BioMart project (Haider, Ballester et al. 2009) is focused to provide easy and fast data mining access to large datasets of mammalian genomes.

The issue of model integration arises, when building Meta-models or multi-level model integration. Integrating micro-models with macro-models, model synchronisation, upward causation and downward causation are other issues that must be considered, when integrating models at different abstraction levels. The current tools for modelling and simulation are dominated by mathematical modelling approaches (Coveney and Fowler 2005), which are not interoperable with informatics tools and are biologist unfriendly. An Agent Based Modelling and Simulation (ABMS) approach will not only provide solutions for both biologists and biology, but also be interoperable with existing informatics tools, biologist friendly and most importantly, when agents are represented at molecular resolution, facilitate interoperability between structural biology and systems biology - two important yet disconnected research domains.

The issue of model interoperability arises when using heterogeneous models across applications. The possible ways to integrate models are to use standard data exchange formats or produce reusable modules for model construction. Systems Biology Markup Language (SBML Contributors 2010) is a file format that has become the standard for computational biologists to exchange kinetic models between SBML compatible tools. CellML is a similar file format used to store and exchange *in silico* mathematical models (CellML Project Community 2001-2010). In contrast Little b is a modelling language, that considers a modular approach to modelling by taking individual modelling components, plugging them together and getting a comprehensive view of the emerging behaviour (Krieger 2006). Little b is a modularised reusable package, that can be used to assemble a model cell. The chosen Collective Intelligence approach focuses on using object modelling techniques and agent oriented programming. SB-UML is an object oriented modelling language that is customised for systems biology (Magali Roux-Rouquié and Soto 2005). The aim of

agent oriented programming in systems biology is integrating diverse heterogeneous models to achieve model interoperability. In agent oriented software engineering, complex systems are organised as autonomous software entities called agents which are situated in an environment and communicate via high-level languages and protocols. This approach supports biologist querying a system, which is very close to their mental model(Luck and Merelli 2005; Merelli, Armano et al. 2006).

There are many languages for agent oriented programming (Dastani and Gomez-Sanz 2005). Most agent programming languages, such as 3APL, Jasen, Jadex and Jack are designed to implement reactive agents. Agents can be implemented in object-oriented languages, declarative languages or a combination of both. Agent oriented programming languages and tools, such as Jade, Jack and Jadex are extensions of Java, or implemented in Java (Jason, 3APL). Agent Programming languages, that are implemented using declarative languages are ConGolog, MetateM, DyLog, Flux, DALI, MINERVA, ALIAS and Agent-0. 3APL, PROSOCS and PROVA (Kozlenkov, Penaloza et al. 2006) are implemented in combination of imperative (Java) and declarative (Prolog) programming languages (Dastani and Gomez-Sanz 2005).

Rather than developing an agent based simulation environment from scratch by using one of the above programming languages, it is feasible to use an agent based simulation package to implement a model. The next section evaluates some widely available platforms for simulating Collective Intelligence.

3.7.1.3 Platforms for simulating complex adaptive systems

Currently available platforms for simulating Collective Intelligence can be located at (SwarmWiki Contributors 2009). There are open source, freeware and preparatory packages for agent based modelling and simulation. Some widely used open source packages are SWARM, Repast Simphony (Repast S), MASON and Ascape. Some widely used freeware packages are NetLogo and StarLogo. Available proprietary packages are AgentSheets, AnyLogic and iGEN. Some reviews have been conducted recently on widely used platforms (Gilbert and Bankes 2002; Castle and Crooks 2006; Railsback, Lytinen et al. 2006). Based on a comparison of open source packages (see Table 3.6), Repast Simphony was identified as the most suitable tool to model and simulate the collective behaviour of biomolecules, mainly due to Repast Simphony's

open source, rapid progress, versatility, support and its expanding user community (Macal and North 2008; Repast Development Team 2008).

Table 3.6: Comparison of open source toolkits

	SWARM	RePast Simphony	MASON	Ascape
Current Developers	Santa Fe Institute / SWARM Development Group, USA	Argonne National Laboratory, USA	Center for Social Complexity & evolutionary computation lab, George Mason University, USA	NuTech Solutions, Inc. BiosGroup, Inc. Metascape, LLC
Date of Inception	1996	2000	2003	1997
Current version	2.2	Repast S 1.2	14	5.5
Website	http://www.swarm.org	http://repast.sourceforge.net	http://cs.gmu.edu/~eclab/projects/mason	http://ascape.sourceforge.net/
Implementation Language	Objective-C / Java	Java / Python / Microsoft.Net	Java	Java
Operating System	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX
Flexibility	User specified algorithms	User specified algorithms	User specified algorithms	User specified algorithms
Speed	Runs well on screen/ Has batch mode	Runs well on screen/ Has batch mode	Runs well on screen/ Has batch mode	Runs well on screen/ Has batch mode
Facilities	Extensible (Not Built in) result logging and graphing	Built in and Extensible result logging and graphing	Extensible result logging and graphing (graphing not Built in)	Built in and Extensible result logging and graphing
Analysis	Support basic statistical methods	Support advance statistical methods	Support basic statistical methods	Support basic statistical methods
Adaptation	Merely reactive	Evolution of Agent algorithms including learning	Evolution of Agent algorithms including learning	Evolution of Agent attributes
Self-organisation	Multi-level feedback	Multi-level feedback, Feedback between agents and their environment	Feedback between agents and their environment	Feedback between agents and their environment
Causality	Vertical and horizontal	Vertical and horizontal	Horizontal	Horizontal
Exascale computing	Millions of agents	Billions of agents - Trillions of agents	Millions of agents	Millions of agents
Process	Single processes discrete event simulator	fully concurrent multithreaded discrete event scheduler	Single processes discrete event simulator	Single processes discrete event simulator
Required programming experience	Strong	Strong	Strong	Strong

External tool integration	Yes (e.g. R and S-plus statistical packages)	the R statistics environment, *ORA and Pajek network analysis plugins, VisAD scientific visualization package, the Weka data mining platform, many popular spreadsheets, the MATLAB computational mathematics environment, and the iReport visual report designer;)	JFreeChart, iText, Java Media Framework, and Quaqua	No
Distributed Computing support	No	Yes (via Terracotta)	No	No
Availability of demonstration models	Yes	Yes	Yes	Yes
Source code of demonstration models	Yes	Yes	Yes	Yes
Tutorials / How-to Documentation	Yes	Yes	Yes	Yes
Additional information	(Minar, Burkhart et al. 1996)	Agent Analyst Extension (http://www.institute.redlands.edu/agentanalyst) Useful weblog: http://www.gisagents.blogspot.com	(Luke, Cioffi-Revilla et al. 2004)	(Inchiosa and Parker 2002)

3.7.2 Technological Feasibility

Simulations using Swarm/Collective intelligence typically involve many agents and are generally classified as huge simulations (Schut 2007). A technological feasibility was undertaken to assess it in terms of software and hardware requirements to extend the cell model to represent a minimal cell to realise the full potential of the Collective Intelligence based cell modelling environment. Current agent based modelling software are extensible and support billions of agents. The Global Scale Agent Model to simulate epidemics includes 6.5 billion distinct agents (Epstein 2009). However, a typical *E.coli* prokaryotic cell contains an estimated 50 million molecules, which includes all the macromolecules, metabolites, cofactors and ions (Broderick, Ru'aini et al. 2005). Further, a typical *Mycoplasma* bacterium only contains an estimated 1 million molecules (Broderick, Ru'aini et al. 2005). However, representing a biological cell, which contains a large molecular population will require a parallel or distributed

simulation system, which involves scaling the simulation over multiple processors. Hence a logical solution to scale up a Collective Intelligence based cell modelling and simulation environment will be to use a high performance computing architecture. The Repast community has attempted to use a distributed simulation architecture, which could facilitate enormous computing resources and data sets. The High Level Architecture (HLA) integrated with Repast (HLA_RePast) is capable of harnessing the computational power of a distributed simulation infrastructure (Minson and Theodoropoulos 2008). Moreover HLA_GRID_RePast (Theodoropoulos, Zhang et al. 2006), which integrates HLA_Repast and HLA_GRID, acts as a middleware for executing distributed, large scale simulations of agent based systems over Grids. However, the Repast development team has recently integrated the current version of Repast Simphony with Terracotta, which is an open source scalability platform. Terracotta seems to provide a better solution than previous approaches, since it is easy to scale java applications to multiple computers (Terracotta 2009). An alternate solution is to use supercomputers. Based on the top 500 supercomputer list (Meuer, Strohmaier et al. 2009), which lists the world's most powerful supercomputers, that are competing for the top spots biannually. These computers have reached a performance capacity of the order of petaflops/s (quadrillion calculations per second) with nearly a quarter of a million cores. "Personal Super Computers" are now emerging as an alternative to conventional supercomputers. These personal workstations, have around 960 cores and up to 250 times the computing performance of a PC, and are sold at a fraction of the cost of conventional supercomputers (Nvidia 2009). Personal super computers now promise teraflops on a desktop, which is equivalent to the world's fastest supercomputer in 1997. Moreover, personal computers are rapidly improving performances by using multi core architectures. There are prototypes by AMD and Intel with architectures combining multi processors containing multi cores (AMD 2008).

3.8 Concluding Remarks

In this chapter the modelling requirements are identified. Biological modelling methodologies are reviewed and compared with the requirements. This led to a decision to adopt with suitable adaptations a bottom-up systems biology approach and utilise a mechanistic model development process to develop a computational model,

using agent based formalism in the wider framework of Collective Intelligence to represent the intracellular behavioural/functional organisation. This is because the aim of the study is to understand how biological cells dynamically adapt to multiple objectives concurrently which are facilitated by the constituent biomolecular activities, which require traversing from lower level molecular resolution to higher level cellular resolution. Such a multi-objective topology provides a concurrent and hierarchical view of biological systems, whereas a network topology provides a sequential and horizontal view of biological systems. However, mathematical models, which use a network topology, are designed to model at the population/aggregation level and are unable to model at the molecular resolution level. The Collective Intelligence approach challenges the assumption used in classical chemistry for its applicability in cellular chemistry. This approach focuses on biomolecular activities rather than the biomolecules because when, where and what biomolecular activities are performed are crucial for adaptive dynamics in a physiological timescale. Further it can be used to analyse the causation of biomolecular activities in space and time.

This model is driven by the principles of Swarm/Collective Intelligence, which capture the inherent characteristics of a cell, such as adaptability, robustness and efficacy with no external supervision (Schut 2007). Modelling and simulating these characteristics is essential to understanding the mechanism by which intracellular solutions emerge from a situation in which many biomolecular activities are happening that meet the adaptive requirements of cells. This insight is essential to gaining an understanding of the transformation between normal and pathological processes in cellular systems. Some of the noteworthy properties of Collective Intelligence systems are adaptivity, emergence, global-local order, interaction, rules, redundancy, robustness and randomness (Schut 2007) which are the characteristics needed if we are to represent a biological cell. Out of the widely used agent based modelling and simulation toolkits, Repast Simphony was chosen mainly due to its rapid progress, versatility, support and expanding user community. The next chapter describes how swarming can address the issues raised in Chapter 2.

Chapter 4

Representation of Biomolecules and their Activities within an *In silico* Environment

"Imagination is more important than knowledge. For knowledge is limited to all we now know and understand, while imagination embraces the entire world, and all there ever will be to know and understand."

- Albert Einstein

4.1 Overview

The aim of this chapter is to describe how swarming can address issues raised in Chapter 2. It introduces the scope, principles and properties of Swarm/Collective Intelligence. The Collective Intelligence framework is based on a meta-formalism, which can be used for complex and self-organising systems. The problem framework is based on Cellular Intelligence, that is a biological cell's ability to organise and adapt to perturbation and uncertainty, which reflects on the characteristics of intelligence. *In silico* representations of native biomolecules and their activities, which constitute a cellular environment, are discussed in Sections 4.3 and 4.4, respectively. Fundamental principles utilised are self-organisation and thermodynamics to represent biological and physical constraints respectively. This guides the intracellular organisation by reducing uncertainty. Section 4.5 describes biological and physical constraints involved in self-organisation of biological cells. The dynamic framework utilises a multi-objective topology, as its model structure and describes the logic of Collective Intelligence, which can be used to construct/deconstruct tasks for the intracellular organisational behaviour of a cell in a physiological timescale. Section 4.6 describes the *in silico* representation of a cellular environment, which uses both biomolecules and their activities.

4.2 Swarming the Internal Organisation of Biological Cells

Intelligence is often associated with learning, which is an adaptive process. The most appropriate definition for intelligence, that incorporates all computational intelligence approaches is described as "the capability of a system to adapt its behaviour to meet its goals in a range of environment" (Fogel 2006). The ability to learn or adapt is one of the hallmarks of intelligent systems. This can also be witnessed in biological cells, where cellular intelligence emerges as an organisational level property of the collective behaviour of biomolecules. Cellular intelligence is defined as the ability to regulate when, where and what biomolecular activities should occur to maintain biological equilibrium in a range of environments. However, the process of adaptation is fundamentally different at a cellular level, since intelligence resides not in individual native biomolecules, but in diverse interactions/activities amongst them. At an organisational level, behaviour can be perceived as a pattern of response to perturbation. These patterns of responses are self-regulated amongst diverse biomolecular activities which include transcriptional, post-transcriptional, translational and post-translational activities. Although there are various categories of real-time adaptation, such as supervised, unsupervised and reinforcement adaptation, the mechanism that drives adaptive behaviour in biological systems is reinforcement adaptation (Eberhart and Shi 2007). While a supervised adaptive process will have a predetermined goal and external supervision to meet the goal, reinforcement adaptation does not have such goals or supervision. It relies on a critic to provide heuristic reinforcement information. Modelling collective behaviour of biomolecules will involve representing cellular adaptation in a Swarm/Collective Intelligence framework.

4.2.1 The Principles of Swarming

Evolutionary Computation (EC) paradigms are inspired by adaptive strategies utilised by biological systems. While these strategies can be found in every level of biological organisation, almost all EC techniques, which comprise the techniques of Evolutionary Algorithms (EA) to Swarm Intelligence (SI) have been inspired by organism level adaptive strategies (Eiben and Smith 2007). While EA techniques are based on trans-generational genetic adaptation of organisms (biologically inspired), SI is mainly based on intra-generational collective behavioural adaptation of organisms

(socially inspired). Natural selection forms the basis for EA techniques and there are many different variations such as Genetic Algorithms (GA), Evolutionary Programming (EP), Evolution Strategies (ES), Genetic Programming (GP) and Learning Classifier Systems (LCS). Adaptation using distributed collective problem solving strategies and self-organisation, form the basis for SI techniques (Engelbrecht 2005). The techniques of SI mainly comprise Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Differential Evolution (DE) and Cultural Algorithms (CA) (Periyasamy, Gray et al. 2008b).

Particle swarm algorithms have been inspired by the collective behaviour of social animals. They operate on the principles of collision avoidance, velocity matching and flock centering. Ant colony algorithms have been inspired by social insects. They operate on the principle of stigmergy, which is a form of indirect communication using the environment as a mediator. Two forms of stigmergy have been defined: sematectonic and sign-based. Sematectonic stigmergy refers to communication via changes in the physical characteristics of the environment (e.g. nest building, nest cleaning, and brood sorting). Sign-based stigmergy facilitates communication via a signalling mechanism implemented via chemical compounds deposited by ants (e.g. pheromone trails). A stochastic diffusion search uses direct one-to-one communication and the information is diffused via this communication, while a Gravitational Search Algorithm (GSA) is developed based on the law of gravity and notion of the mass interactions. Information is transmitted using gravitational force between different masses. Intelligent water drops (IWD) has been inspired from natural rivers. In the IWD algorithm, several artificial water drops cooperate to change their environment in such a way that the optimal path is revealed as the one with the lowest soil on its links (Wikipedia Contributors 2010c).

The study of Swarm Intelligence is providing insights into management of complex systems (Miller 2007). Swarm technologies are solving complex problems, where traditional approaches are unsuccessful (Hinchey, Sterritt et al. 2007). Swarming has been inspired by collective behaviour of social insect colonies and other societies (Bonabeau, Dorigo et al. 1999; Garnier, Gautrais et al. 2007). The proactive behaviour of swarm systems mainly results from a reactive behaviour of its constituent entities rather than an entitie's deliberative behaviour. Swarm Intelligence refers to the phenomena of a system of spatially distributed entities coordinating their

actions in a decentralised and self-organising manner, so as to exhibit complex collective behaviour from local interactions. The concepts of self-organisation and emergence underlie swarming and these systems are inherently adaptive, robust, flexible, stochastic and concurrent. The first step towards modelling intracellular organisational behaviour is, understanding the mechanisms that foster collective behaviour among biomolecules. The main features of Swarm Intelligence involve forms of limited or minimal communications and/or interactions, large numbers of interacting entities with limited reach, and some global efficient, emergent or self-organised behaviour (Fleischer 2003). Further the four basic ingredients for manifestation of self-organisation are (Bonabeau, Dorigo et al. 1999):

- Forms of positive feedback – an amplification mechanism that promotes creation of autocatalysis amongst biomolecules, which build up the activities in group. It promotes cooperation, in which mutual dependency fosters persistence of members of the group. Activities are generally amplified by replicating biomolecules and/or activating them.
- Forms of negative feedback – this compensates positive feedback and facilitates stabilisation of a group's activities. Activities are usually lessened by degradation of native biomolecules, either by competitive or non-competitive inhibition. It controls competition among groups. Negative feedback is caused by inhibition of biomolecular activities, competition for resources, saturation of biomolecular activities and exhaustion of a resource.
- Amplification of fluctuations – this gives rise to new solutions for internal organisation of cell. It includes fluctuations in when, where and what native biomolecular activities should occur, alterations caused by mutations and recombination, and variations in time and energy requirements of individual activities. Alterations at the genomic level will have a long term global effect compared to alterations during transcription and translation, which are local and short term. Although intracellular organisation sustains itself despite randomness, randomness facilitates discovery of new solutions.
- Multiple interactions of multiple entities – give rise to extremely concurrent and redundant biomolecular activities distributed in time and space within a cellular environment. Although redundant activities occur at different points in

time and space, they provide a statistical interpretation of ensemble activities, on which global effects are judged. Basically redundancy provides a medium by which the characteristics of individual native biomolecules are amplified.

The above ingredients of self-organisation will naturally give rise to organisational level properties such as dynamic, emergence, robustness, plasticity, bifurcation and multi-stability (Garnier, Gautrais et al. 2007). In biological cells, dynamic nature is observed in the form of oscillatory behaviours with respect to biochemical tasks or points of control. There are numerous points of control within the diverse cellular biochemical activities, which tend to regulate a cell. Chemical oscillation is a macroscopic (statistical interpretation at population level) process which wavers between conflicting courses of action (biomolecular activity) and exhibits periodic changes of control activities. This is a macroscopic phenomenon, which results from the ensemble behaviour of native biomolecules. A notable feature of oscillation is the existence of equilibrium and presence of restoration forces in either direction, which grow stronger the further the system deviates from equilibrium. In biochemical systems, this force is formed by a feedback couple, (positive and negative), with respect to a steady state. While certain feedbacks have specific effects on a task such as activation or inactivation of a specific species of functional product, others (e.g. production and degradation of native biomolecules) can have general effects on cellular tasks/objectives. The extent and sensitivity of deviation are two properties that can be observed in a chemical oscillatory system. The extent depends on causal distance of feedbacks. The closer feedbacks are to the point of control the smaller the deviation from equilibrium, because they provide greater flexibility by reducing uncertainty and control delays over biomolecular activities contributing to the steady state. Thus, the distance of feedbacks from the point of control contributes to delays, which affect the time required to realise the appropriate level of response required at the point of control. Sensitivity determines the speed at which deviation occurs. Further a cell consists of many chemical oscillatory systems, and this in turn produces complex control behaviour, that facilitates sustaining internal organisation of a cell.

Emergent properties arise from nonlinear interactions amongst biomolecules. Bifurcation is the appearance of a qualitative change in collective behaviour, when changes are made to the bifurcation parameters. This behaviour can produce new stable solutions. Multi-stability implies that for a given set of parameters, the system

can reach different stable states (i.e. attractor) depending on the initial conditions and on random fluctuations. The two most important emergent properties are robustness, which is the ability of a system to maintain its functions under diverse conditions and plasticity, which is the ability of a system to readily adapt to new, different, or changing requirements. Robustness results from redundancy, which gives rise to multiple interactions of multiple entities. Moreover, failures of a few members are rapidly compensated by the remaining members. Plasticity refers to organisational adaptations that occur without any change of the behavioural rule at an individual level. This collective behaviour is modulated by perturbation which comprises extrinsic (environmental) stimulations and intrinsic (programmed) stimulations. These stimulations are responsible for biasing rather than altering the responses.

One distinguishing feature between conventional swarming and biological cell swarming is that conventional swarms consist of homogeneous or heterogeneous types of entities, whereas a biological cell will consist solely of heterogeneous types of entities representing different biomolecular species. Each biomolecular species is restricted to certain kinds of behaviour, with constraints on interaction and collaboration with other biomolecular species. Since constraints can reduce uncertainty, this naturally leads to the formation of order out of chaos, amongst the activities of native biomolecular species. Native biomolecules are defined as complex biopolymers (functional products) created by the cellular machinery using its own genetic information. In contrast, metabolites are perceived to be cellular resources in various forms. Native biomolecules collaborate via direct and indirect interactions. Direct interactions occur, when native biomolecules come into physical contact, such as in complex formation, signalling and regulatory activities. Indirect interactions occur when native biomolecules share resources, such as metabolites. These indirect interactions, known as stigmergy, are a key concept in the field of Swarm Intelligence (Parunak 2003), and result from the self-organising mechanism of spontaneous indirect coordination between agents due to the shared environment which means they can sense and modulate. This produces complex intelligent behaviour in the absence of planning, control and direct communication between agents, and supports efficient collaboration between very simple agents, which lack any memory, intelligence or knowledge of each other (Wikipedia Contributors 2010b).

In existing SI techniques, the behaviour/characteristic of computational individual is reproduced by directly interacting with neighbours in space and these interactions are non selective. Whereas in collective behaviour of biomolecules, the behaviour/characteristic of computational individual (species) is reproduced by causally influencing other computational individuals (species) that form the autocatalytic set. Moreover these interactions are highly selective. The social phenomena which contributed to the existing SI techniques pursue goals/objectives which are spatial in nature and strive to explore and exploit solutions in space within the bounds of their temporal constraints. By contrast, biological cells pursue goals/objectives which are temporal in nature and strive to explore and exploit solutions in time within the bounds of their spatial constraints. The noteworthy distinctions with respect to existing techniques are shown in Table 4.1.

Table 4.1: Distinguishing features between existing swarm intelligence techniques and collective behaviour of biomolecules

SI Techniques	Biomolecular Inspiration
In ACO pheromones produced by computational individuals degrade with time	Computational individuals do not produce pheromones but the individuals degrade with time
In CA and SDS computational individuals dynamically adapt by imitating their neighbours behaviour/characteristics	Computational individuals cannot imitate their neighbours (exceptions: Prions are infections conformational states of proteins that can convert other native state versions of the same protein to an infectious conformational state)

4.2.2 Analysing and Designing Swarm Systems

Every living system has to deal with spatial and temporal constraints. Cellular organisation exists in a confined space and cells are compelled to manage their biochemical activities within this confined space. Unlike conventional swarm systems, that tend to address problems which are spatial in nature, such as locating resources or paths in space, cellular systems tend to be more concerned with problems

that are temporal in nature, such as transformations. For example in a cellular environment resources are in different forms and these resources have to be in specific forms to be accessible by a particular enzyme. In a cellular environment, the emphasis is more on accessibility of resources, rather than locating resources in space. While identifying paths in space concerns representation of an appropriate sequence of locations in space, identifying paths in time concerns representation of an appropriate sequence of events, which separates causes and effects. Biochemical activities facilitate identifying appropriate paths in time, amongst diverse stimulations and responses that occur in a cell. Although cellular organisation tends to have spatial issues, they represent constraints for the temporal objectives (timely responses) of a cell. In contrast the objectives of conventional swarming systems are spatial in nature and temporal issues represent the constraints for a system. The adaptive requirement determines whether objectives become temporal or spatial in nature (see Figure 4.1).

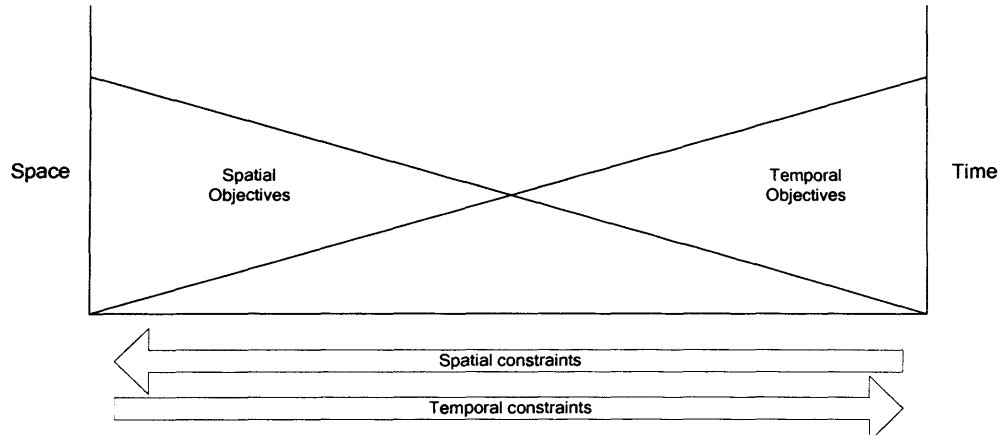


Figure 4.1: Relationship between spatial and temporal issues in situated systems.

Swarming is appropriate for system features such as discreetness, deprivation, distribution, decentralisation and dynamism. Swarm engineering (SE) is a methodology used to engineer system functionalities through emergence. SE is a process for careful design of a group of agents to have a predictable global behaviour (Kazadi 2003). SE uses a middle meeting methodology, that generates a swarm condition which is followed by generating a set of agent behaviour to satisfy the given swarm condition. More details on engineering swarming systems can be found in (Parunak 2003; Parunak and Brueckner 2004). Since swarm systems have no external control mechanisms, they can be represented virtually by any size without major modifications to the underlying architecture. Moreover the basic architecture remains the same, which gives design flexibility for applications. Swarm systems are ideal for

unknown and unpredictable situations. As a whole, these systems have the ability to adapt to rapid changes in a manner that cannot be achieved by centralised control systems. Since the constituent entities in a swarm system only communicate locally, the issue of communication delays are eradicated. Moreover individual entities do not depend on instructions, which enable them to react quickly and consistently in their environment. This leads to a higher fault tolerance, due to lack of failure by centralised control. Hence, modelling local interactions within a biological cell has two implications. It is significant, when there is communication congestion, and these interactions are sufficient to maintain a wide range coordination within the cells.

There are many issues that must be considered when programming a swarming system. The behaviour of a swarm system must be sensible to someone outside the system boundary. This means that progress of the system should be perceivable. The term organisation in self-organisation has distinct but related meanings, such as a mapping, a process or a structure. Mapping can facilitate comparison of the degree of organisation between two systems or the same systems at different times. Many definitions for self-organisation have been proposed. From a physical perspective, it is defined as a process that reduces the entropy of a system without external intervention. Various criteria for mapping organisation have been suggested, among these entropy and symmetry are common suggestions (Parunak and Brueckner 2004). However, these methods are too abstract and can only be applied to organisation of spatial structures, rather than functional organisation which is temporal in nature. Further these criteria seem to be too abstract and do not have any practical use in measuring intercellular organisational behaviour. Intracellular organisation cannot be observed and neither can it be empirically measured, because functional organisation is more of a temporal phenomena rather than spatial phenomena. These temporal phenomena, which manifest as temporal symmetry, are reproducible rhythmic behaviour of a cell. The following mathematical/quantitative oriented definition for self-organisation has the ability to measure intracellular organisational behaviour. "Self-organized behaviour in a complex system involving multiple performance measures is a sequence of system states corresponding to movement along a Pareto optimal frontier" (Fleischer 2005). In this context functional organisation is measured in terms of efficiency of diverse functional units that constitute intracellular organisational behaviour.

Dynamic systems have to deal with changing requirements, and swarming is an ideal approach for dealing with dynamically changing requirements. The constituent entities of the system do not encode system level behaviour explicitly. The consequence of this is that entities will not require modification with changing requirements. The scope of cellular organisation is characterised by the amount of change to which a cell's adaptive requirements are susceptible. This is how frequently the adaptive requirements change. If the extent of change is small, cells are more likely to anticipate and deliver acceptable performance within the bounds of their original configuration. This configuration will determine when, where and what native biomolecules should be functional. However, when the extent of change is huge, there is more value for cells to adapt to unanticipated requirements. Timely response is a crucial factor for the overall fitness of a cell. As rate of change begins to outpace the intracellular rate of information/signal migration and transformations within a cell, they will constantly find themselves providing solutions to obsolete problems. This will lead to the collapse of the cellular organisation. Information/signal migration represents a signal physically moving in space, and information transformation represents information existing in different formats, such as genetic, transcript or protein, where transformations among these formats occur via transcription and translation. Based on timescales chemical transformations appear to provide major constraints to arriving at a solution. Except for a genetic format, heritable information present in other formats will contribute to epigenetic inheritance. Epigenetic inheritance often lasts for one or two generations, because, information stability decreases from a genetic format to the transcript format and to a protein format.

4.2.3 Collective Intelligence Framework

The aim of this approach is to integrate biomolecular activities occurring within the gene, transcript, protein and metabolite spaces, so that interactions across spaces can be studied. The approach utilises a multi-objective topology based on Collective Intelligence to model adaptive dynamics of biological cells. Adaptive dynamics at a physiological timescale, occur due to biomolecular interactions rather than genetic adaptation. The performance/fitness interaction is a fundamental criterion used to modularise biomolecular interactions from the bottom-up. Further this criterion also

facilitates identification of cooperation and competition between biomolecular species, which in turn act as organisational constraints on the biological adaptations that occur from physiological to evolutionary timescales. This approach is based on a meta-formalism, that can be used for modelling complex and self-organising systems (Fleischer 2005). This formalism is based on three foundational components. The first is based on a set of first principles which include relevant laws of nature such as evolution and thermodynamics. The second is the dynamic framework based on a concept of multi-objective topology aided by Pareto optimality, which provides a novel way to characterise system interaction, behaviour and efficiency on different scales. The third is the problem framework, which is based on Cellular Intelligence.

The laws of evolution (selection and self-organisation) and thermodynamics are used as the governing principles of cellular optimisation. While the driving force at a fundamental level is the natural propensity of biochemical systems to reach thermodynamic equilibrium (ΔG tends towards zero), at an organisational level it is the natural propensity of biological systems to maintain biological equilibrium. The goal of this approach is to model a biological cell as a swarm intelligence system and to elicit self-organising mechanisms, which allow internal cellular organisation as a whole to behave intelligently in a coordinated manner as a result of direct and indirect biomolecular interactions.

Preliminary research work has been conducted to develop a Collective Intelligence based cell modelling and simulation environment. This environment was used to identify major factors that contribute to a self-organising process of functional products. Some self-organising mechanisms, that were investigated are forms of positive and negative feedbacks amongst functional products, amplification of fluctuations and multiple interactions amongst multiple functional products (Bonabeau, Dorigo et al. 1999). Various internal constraints were also considered to analyse their impact on the self-organising process. In particular, time and energy requirements for activities of functional products and cellular thermodynamic requirements were considered to model molecular level constraints associated with cellular level activities. Biomolecular activities contribute to the internal organisation of a biological cell, and these cannot be directly observed or empirically measured. Since the chemical activities of molecules, rather than their abundance, provide an accurate description of a chemical system, the performances of functional products

are analysed, based on their level of activities. However effects of these activities contribute differentially to the objectives (higher tasks) based on the nature of influence (positive or negative), it could have on those tasks with time.

Organisational behaviour within a cell corresponds to its functionalities, which manifests due to diverse activities of functional products in space and time. These activities cause interactions among functional products, which are brief encounters when considering the timescales for noticeable impact to occur at the cellular resolution. Due to the dynamic nature of the objectives, biological systems are forced to meet adaptive requirements by pursuing the objectives aided by its constituents, thus giving rise to biological processes which are seen as biological functions. In the context of a cellular system these constituents are functional products with the potential to perform their intended activities. However these activities have to be orchestrated in order to pursue objectives/tasks against perturbation and uncertainty, thus causing activities of diverse species of functional products to fluctuate. Since pursuing objectives/tasks is a temporal phenomenon, simplification of biological complexity is achieved by mechanistically constructing/deconstructing the global tasks of a cellular system into basic tasks required to pursue them. The performance of tasks is quantified, in order to quantify the functions of cellular systems.

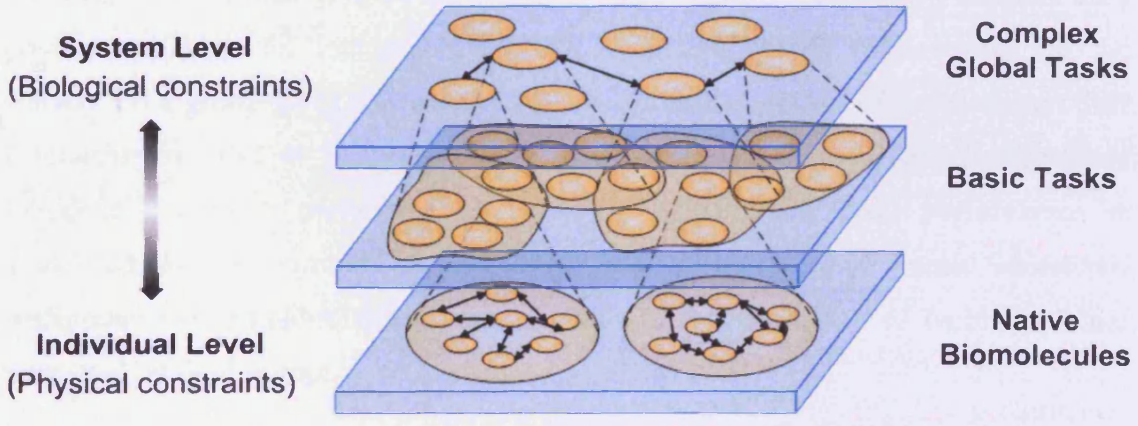


Figure 4.2: Complexity barrier between macroscopic cellular and intercellular processes that emerge from basic biomolecular activities.

Collective Intelligence logic considers the principles and properties of self-organising processes and this facilitates understanding of fundamental and organisational principles of biological systems, which define the possible organisational space of life (Kitano 2007). Identifying and modelling major self-organising mechanisms of a cell

is the most crucial task to be undertaken in mapping the problem space to the function space of a cell. The function space is deconstructed into individual molecular level tasks, which comprise contributions from individual biomolecules, group level tasks which comprise contributions from respective biomolecular species, and team level tasks comprising contributions from heterogeneous species of biomolecules. Categorising the collective behaviour of functional products, in terms of objectives/tasks can deconstruct the global objectives/tasks of a cell into the basic tasks required to pursue them. Representing objectives/tasks using multi-objective topology provides, a concurrent and hierarchical view of cellular dynamics by constructing higher cellular tasks/objectives using feasible solutions, originating from lower level tasks/objectives and the constraints associated with them. Coordination among the basic and complex global tasks is achieved via various regulatory mechanisms, which control competition using negative feedback mechanisms, whereas cooperation is favoured by positive feedback mechanisms. This framework facilitates mechanistic identification of relationships between basic tasks performed by a biomolecular species. Performance interactions among tasks are fundamental to modelling propagation of impact among the basic tasks, which are constituents of the higher tasks (see Figure 4.2). Performance is measured at each task level, ranging from basic tasks to complex global tasks of a cell. Factors that affect performance at the level of functional product are efficiency and stability. Efficiency depends on a product's affinity for interaction, and the time and energy requirements for the activity. At a group level, the net activity of a particular species of functional product is considered, and at the cooperative module level, where different species of functional products participate to complete a particular task, performance is associated to the number of completed task within a time frame. Moreover, performance of complex tasks, which constitute a combination of basic tasks, are measured by their output.

4.3 Modelling the Characteristics and Behaviour of Biomolecules

Biomolecules have evolved into different forms and are classified as native or foreign biomolecules. Native biomolecules represent chromosomes, transcripts and proteins,

while foreign biomolecules are metabolites. Chromosomes are composed of genes. Transcripts are further categorised as mRNA, tRNA, rRNA and smallRNA. Proteins are classified as enzymes, signalling, regulatory, transport and mechanical proteins (see Figure 4.3). Metabolites are broadly classified as primary and secondary metabolites. In addition to contributing to certain regulatory (metabolic regulation) and signalling activities, metabolites provide the source of matter and energy required to build and sustain the cellular organisation.

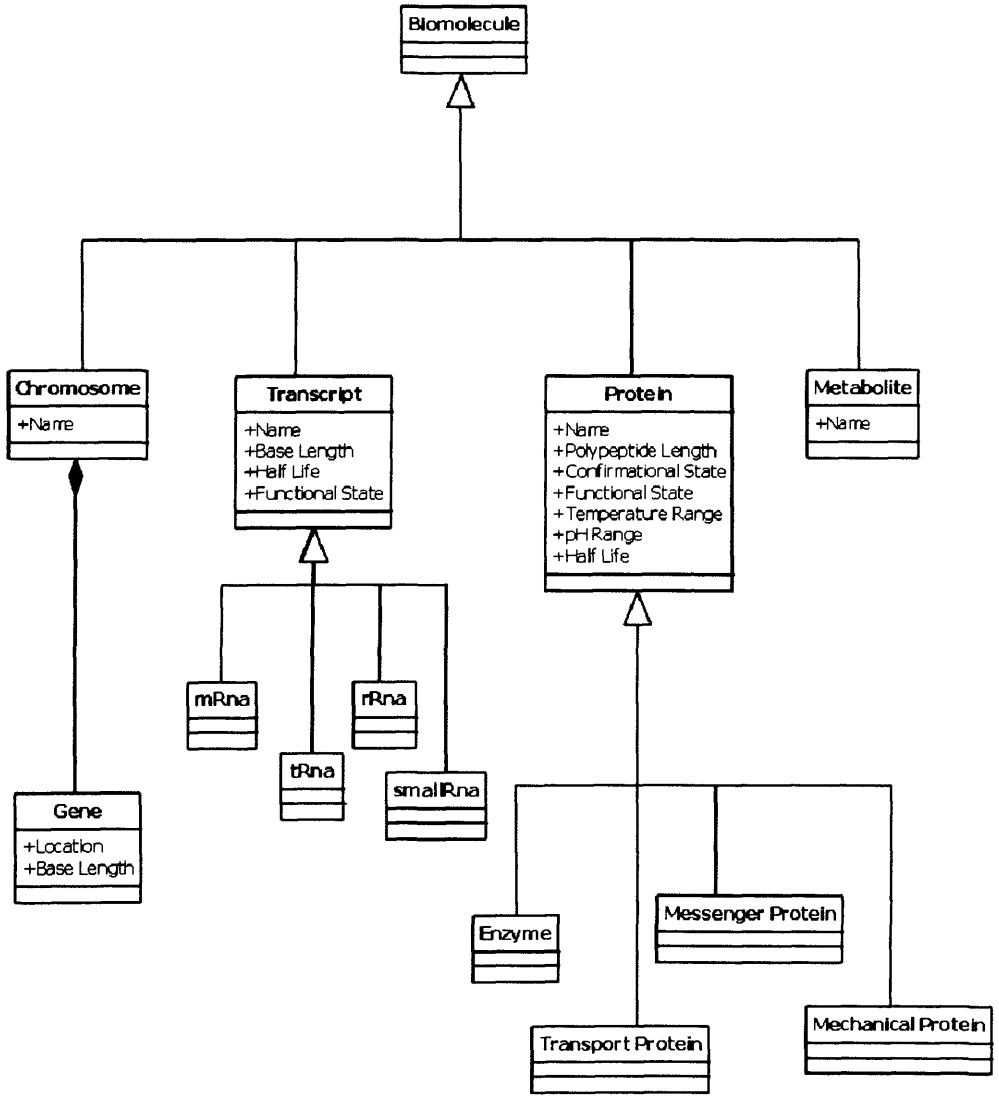


Figure 4.3: Class diagram of biomolecules

The reactive behaviour of native biomolecules is modelled using state machines, where states and transition between states, are represented in agents. Proteins can exist in different conformational states, phosphorylation states or functional states

(active or inactive). When active the protein can exist in idle or performing states (see Figure 4.4). Other noteworthy properties, which proteins can have, are stability to temperature and pH. These properties determine a protein's performance. Also transcripts and proteins are vulnerable to degradation which is represented by its half-life.

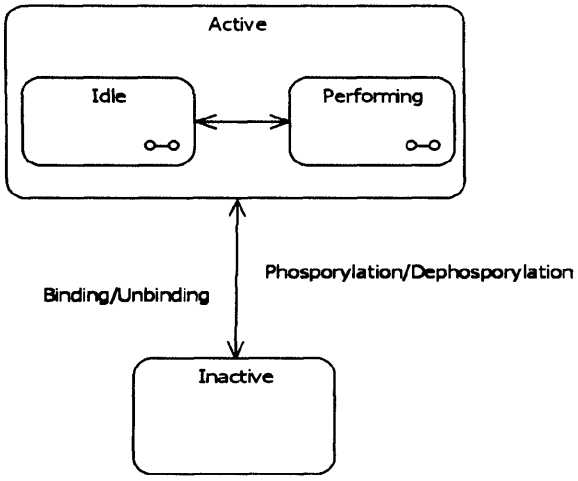


Figure 4.4: Typical functional states of a biomolecule

4.4 Modelling Interactions of Biomolecules

The significance of modelling biomolecular activities, as opposed to biomolecules themselves, is that biomolecular activities represent interactions that take place in order to sustain a biological cell. These activities cause direct and indirect interactions among biomolecules and have a distinct location in space and time, modelled as biomolecular events. Measuring the activities of proteins rather than their actual abundance reveals effective abundance. Apart from contributing to molecular crowding, biomolecules which are merely occupying the cellular environment, will not have any major effect on cellular processes. A biomolecule's contributions are judged by its activities. As explained in Section 3.5.2.3, the chemical activities of molecules provide the most accurate description of a chemical system. Nevertheless, the dynamic state of chemical systems are described in terms of concentrations, as an approximation to chemical activity based on an assumption that the difference between the concentration (actual population) and the chemical activity (effective population) is insignificant. However in biological cells, where functional products are complex molecules, and only certain states out of all possible states, have the

ability to perform an intended activity, there is a significant deviation between the actual population and the effective population. For example, proteins that can undergo various post-translational modifications can lead to various phosporylation states or conformational states. These states can affect a protein's activity and can cause significant deviations, when representing proteins with their actual abundance. Hence actual abundance does not reflect the true dynamic state of a cell.

Time and energy requirements are crucial, when modelling biomolecular events, since time, energy and efficiency are some of the biologically relevant properties. Efficiency can be represented at an individual and organisational level. At individual level, the time and energy requirements of a biomolecular activity play a significant role in biomolecular efficiency. At organisational level the concurrent orchestration of biomolecular activities to provide best possible solutions to multiple objectives, play a significant role in cellular efficacy. The duration (time requirements) of various biomolecular activities are shown in Figure 4.5. The temporal scale represents intra-molecular and inter-molecular dynamics (molecular interactions). A typical temporal scale for cellular phenomena varies from femtoseconds (10^{-15} s) to hours (10^{3} s). Atomic interactions/quantum dynamics take place in the order of picoseconds (10^{-12} s). Many conformational changes in macromolecules are microsecond (10^{-6} s) events. Metabolic activities, which are characterised by enzymatic reaction, take place in the order of milliseconds (10^{-3} s). Many non-covalent interactions (molecular binding reactions) between macromolecules as found in signal transduction activities and regulatory activities occur in the range of nanoseconds (10^{-9} s) to microsecond (10^{-6} s) (Stryer 1988). The duration of biomolecular activities is significant in the organisation of cellular activities, because the perception of time is different at every level of biological organisation. For example, our perception of a second is like eternity in the atomic or molecular world (Cox 2008). The causal episodes of activities decrease from atomic to molecular and then to cellular scale, which makes biological processes appear slower, when moving up the biological organisation strata.

The energy content of some biochemical entities are shown in Figure 4.6. However at the level of molecular resolution, energy requirements for biomolecular activities are characterised in terms of ATP. While phosporylation activities require a single molecule of ATP, the ATP requirement for transcription and translation activities depends on the length of the gene and transcript, respectively.

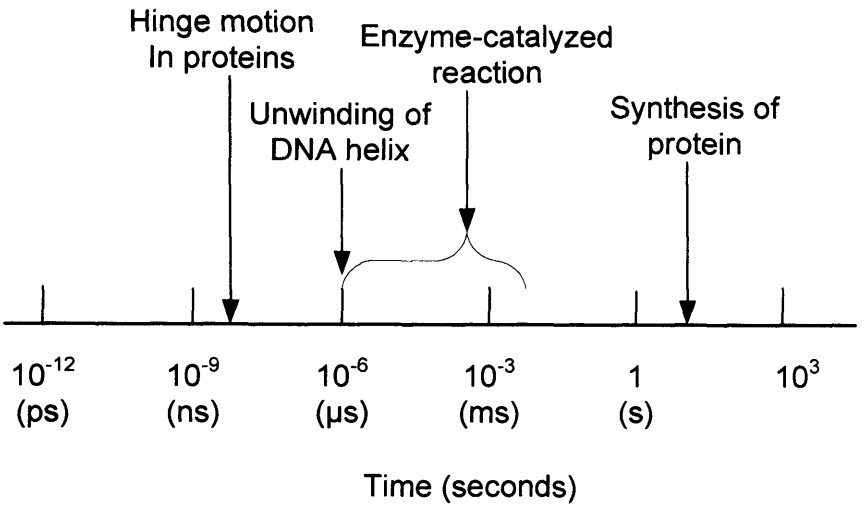


Figure 4.5: Typical event durations of some actions in biological systems (adapted from (Stryer 1988))

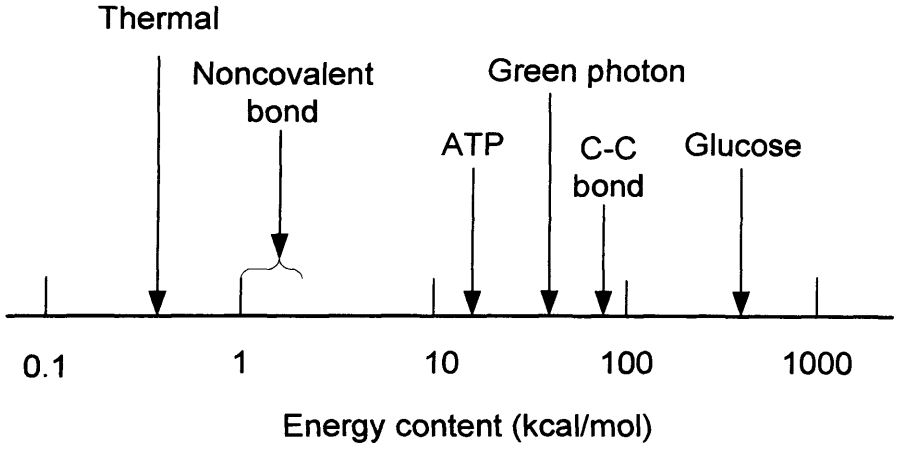


Figure 4.6: Typical energy content of some molecules and bonds in biological systems (adapted from (Stryer 1988))

Molecular abundance, molecular crowding and confinement also determine the probability of reactants meeting, since they affect mobility. Table 4.2 shows typical amounts of biomolecules found in biological cells (Takahashi, Yugi et al. 2002).

Table 4.2: Abundance of intracellular compartments and biomolecules in cells

	Prokaryote	Eukaryote
Compartments	10	10^{3-4}
Biomolecules	10^{13-14}	10^{17-18}
Biomolecular Species	10^{3-4}	10^{4-5}

A typical *E.coli* prokaryotic cell contains an estimated 50 million molecules. This includes all macromolecules, metabolites, cofactors and ions (Broderick, Ru'aini et al. 2005). A typical *Mycoplasma* bacterium only contains an estimated 1 million molecules (Broderick, Ru'aini et al. 2005). Hence, we can conclude that although cells contain a huge number of molecules, only a fraction of them are involved in the formation of a dynamic intracellular organisation. The spatial scale represents the volume of different biological entities. While volume of containers (cells and organelles) facilitates in calculating abundance of biomolecules, the volume of biomolecules facilitates in calculating their mobility in a cellular environment. Spatial scale intracellular components (Schnell, Grima et al. 2007) are shown in Table 4.3.

Table 4.3: Lengths of cellular components

Components	length
Cell	10-100 micrometers
Cell nucleus	5 micrometers
Mitochondrion	2 micrometers
Ribosome	30 nanometers
Protein	4-10 nanometers
Small molecule (e.g. H_2O)	0.5-1 nanometer

4.4.1 Direct Interactions

Direct interactions occur when biomolecules come into physical contact with each other. This could occur between native biomolecules, during complex formation, signalling or regulatory activities, or between native biomolecules and foreign biomolecules during enzyme catalyzed reactions. All biological structures and processes depend on an interplay of non-covalent and covalent interactions. Reversible biomolecular interactions, which are mediated by non-covalent forces, are at the heart of the dynamics of life. For example, recognition of substrates by enzymes and detection of signalling molecules are mediated by non-covalent bonds. Fundamental non-covalent bonds include electrostatic, hydrogen and van der Walls bonds, which differ in geometry, strength and specificity (Stryer 1988). The probability of physical interactions depends on the affinities of the biomolecules caused by non-covalent forces.

4.4.2 Indirect Interactions

Indirect interactions occur due to influence/causality. An ideal way to represent causality, which can have positive or negative effects, is via performance interactions. Positive performance interactions occur due to cooperation among native biomolecules, and negative performance interactions due to competition among native biomolecules. Based on these criteria, fundamental units of cooperation are identified among the many biomolecular interactions. Identifying how these units are configured, can facilitate the understanding of cellular organisational strategies, such as how cells have evolved to manage cooperation and competition among these units by the regulations imposed. For example, enzymes cooperate and compete indirectly by physically interacting with respective metabolites.

4.4.3 Local Interactions

Due to the low number of certain native biomolecules, binding sites in the genome, and molecular crowding, a cellular environment appears to be spatially heterogeneous and stochastic, hence the need for a stochastic approach to represent discrete biomolecular activities occurring across space and time. Different regions of a cellular environment will have different compositions of biomolecules. To model these stochastic fluctuations, biomolecular interactions have to be modelled, based on locally available information. Local interactions are the direct and indirect interactions, which occur in the neighbourhood of biomolecules. However, there is no global interaction with a cell, since biomolecules are reactive entities, functioning without any cognitive ability. They simply interact without any global awareness.

4.5 Modelling Collective Behaviour of Biomolecules

Intelligent cellular organisation emerges out of biomolecular interactions in space and time, which contribute to the collective behaviour of biomolecules. Collective behaviour of biomolecules are organised into a nested objective hierarchy. Objectives range from being physical to chemical and biological, when traversing from molecular resolution to cellular resolution. Mutual dependency of cellular objectives on lower objectives contributes to the combined complexity. Identifying orthogonal objectives will be a key to representing cooperative modules. Theoretically

orthogonal objectives are fully independent of each other and their combined complexity is simply additive. However, in reality orthogonal objectives are only approximately orthogonal. The lowest objectives are found at the resolution of biomolecules, where they are physical and based on covalent (processing) and non-covalent (binding) interactions. Orthogonal objectives can also be identified from a physical and chemical perspective. While physical space is independent, chemical space is constrained by physical objectives. Objectives at the level of biomolecular species, which also define their niche in a cell, are the collective efforts of member species. Interactions between the lower objectives occur via feedbacks. Collective autocatalysis is a form of positive feedback, which promotes the cohesiveness of lower level objectives and positively contributes to higher level objectives of a cell. Collaboration among lower objectives is promoted. In contrast, competition among lower level objectives promotes negative feedback (influence) amongst objectives. Both representation and analysis of collective behaviour of biomolecules, can facilitate analysis studies of a cell. Biomolecular activities contribute to the internal organisation of a biological cell, their abundance is typically abstracted into a network topology to analyse population dynamics in space and time. This topology has become an idealisation of reality (Stelling, Sauer et al. 2004). However to analyse the degree of organisation, performance of various intracellular organisations, such as modules will have to be measured, as performance reflects the behaviour of modules. Clearly biomolecular activities rather than their abundance, will be significant in measuring performance of their respective modules. Any performance deviation of any of a module's member biomolecular species, will have a direct impact in that module.

4.5.1 Biological Constraints of Intracellular Organisation

Biologically relevant constraints are time, energy, matter and space. Of these time and energy play a major role in regulating biomolecular activities. Matter organised in different ways carries information required to perform activities. This information is present as a one-dimensional sequential format (a gene), which is transformed into three-dimensional structural formats (native biomolecules). These structures determine the qualitative features of a biomolecule, which in turn determines the activity it performs. Out of the space of possible activities, which various

biomolecules can perform in living systems, only a subset of these activities will be available for a certain living system. Feasible solutions can only be provided based on available activities and their regulation. While most stimulations are native as they are configured to regulate by the genome, remaining ones are foreign in nature, such as metabolic regulation and extrinsic signals. A cell requires various biomolecular activities to be performed to sustain biological equilibrium. Time becomes a limiting factor, when activities have to be performed within a stipulated time. Energy becomes a limiting factor, when activities have to be performed with efficient use of energy. Since two kinds of responses can emerge to meet this demand, cellular organisation has the option of a trade off between time and energy, which is determined by both qualitative and quantitative responses of biomolecular activities. Qualitative responses of biomolecular species are determined by the time it takes to perform a stipulated activity, reusability (the number of activities a native biomolecule can perform during its existence) and energy usage. Quantitative responses are determined by investment of time, energy and material needed to produce the required amount of biomolecules to perform the activities. It is analogous to measuring work by using time units, such as person hours to complete a job. If time is of the essence, a larger work force will be allocated, otherwise a smaller work force will be sufficient. Various levels of regulation have evolved to quantitatively modulate the amount of biomolecular activities.

4.5.2 Physical Constraints of Biomolecular Interactions

The fundamental nature of all biomolecular interactions is energy, and the science of thermodynamics provides a valuable tool to help comprehend energy (Stryer 1988). The organised nature of living systems seems astonishing, given it emerges from a chaotic world of non-living objects. Nevertheless, the organisation perceptible in a biological cell emerges from biological activities, that are subjected to the same physical laws (in particular the laws of thermodynamics), that govern all physical activities. The physical systems' hierarchy comprises quarks, which form sub-atomic particles, which in turn form atoms, and finally molecules. All physical systems in the universe follow the second law of thermodynamics and proceed in an exergonic direction, which is $\Delta G < 0$; in a direction that tends to lower the free energy of a system by expending energy in the form of work. Hence a system that is far from

equilibrium ($\Delta G = 0$) has the potential to do work. The second law of thermodynamics predicts the direction of change for any biochemical reaction by stating that entropy must increase for any spontaneous process (Grace 2004). Entropy is a measure of the degree of randomness or disorder in a system. When activities or events are equally probable in space and time, complete randomness/disorder will prevail within the system. However, due to naturally occurring constraints the probabilities of activities/events are altered in space and time, which paves the way for emergent patterns in self-organising physical systems. Although at an individual level events are discrete and appear to be chaotic, at the collective level they appear to be in a continuum and ordered based on statistical interpretation of the events. These patterns form the traces of order at the higher level in hierarchical systems, such as biological systems. Hence constraints, which reduce uncertainty by altering probabilities at a particular level, are the causes of order within a system. The order that emerges can have different consequences, such as producing normal or pathological behaviours at the systems level, where the level of selection also begins.

Thermodynamic equilibrium, which is approximated to chemical equilibrium at constant temperature and pressure, is the most probable system state amongst the physical constraints. This dynamic equilibrium is the natural tendency of biochemical systems, which constrain their biological objectives. This steady state is perceived differently in different domains, such as classical thermodynamics, statistical thermodynamics (statistical mechanics or molecular thermodynamics), or chemical kinetics. In classical thermodynamics and chemical kinetics, chemical equilibrium is viewed at the macroscopic level. At equilibrium, a net conversion of reactants to products cannot be observed and $\Delta G = 0$ by definition. Moreover, in chemical kinetics, at equilibrium the rate of forward and reverse reactions are equal. Like classical thermodynamics, chemical kinetics deals with macroscopic/aggregate variables. In particular, rates at which various chemical transformations occur, such as reaction rates, rate constants, rate of catalysis (turnover number) and half-life.

The power of classical thermodynamics is driven by its ability to model the overall behaviour of a system without knowing molecular details. The complimentary domain of statistical mechanics, applies the laws of physics to individual particles (molecules, atoms and photons) to deduce the behaviour of macroscopic systems, by considering the statistical behaviour of a large number of particles (Wolfe 2001). Statistical

thermodynamics was born with the work of Boltzmann, which set the stage for redefining chemistry, in terms of probability (Hanson and Green 2008). It has the ability to make macroscopic predictions, in terms of microscopic properties. The laws of thermodynamics apply to macroscopic systems. At the microscopic level, they are applied statistically to a large number of molecules, but not to individual molecules. The force of nature driving all chemical reactions in all systems, living and non-living, is basically that systems are going from less probable states (states having a small number of possible configurations) to more probable states (states having many possible configurations) (Wolfe 2001). Hence the most probable distribution represents the equilibrium. In a system of interacting particles the energy is shared within particles, which will reach a state where global statistics are unchanged in time.

Figure 4.7 shows a typical chemical reaction based on statistical thermodynamics, where the probabilities of events contributing to a forward or backward reaction are equal, when a chemical system is at equilibrium.

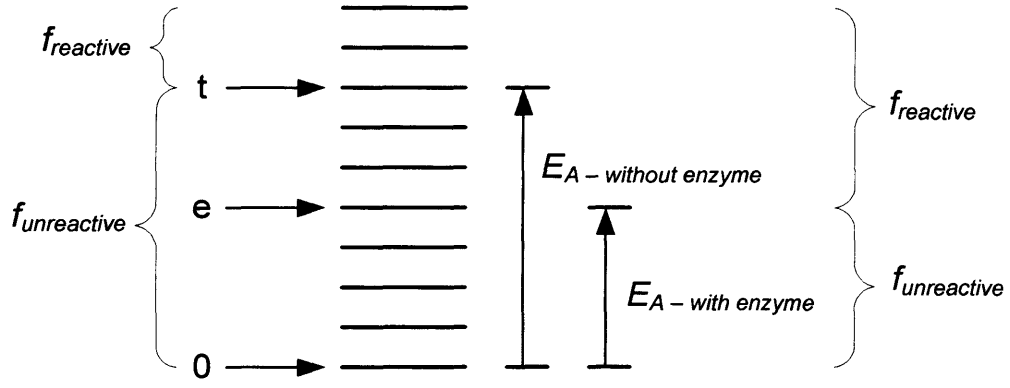


Figure 4.7: Fraction of particles that will react at any given temperature depends upon how many particles have at least the energy of Level e in the presence of an enzyme or Level t in the absence of an enzyme.

Gibbs free energy (G) is the capacity of a system to do work at constant temperature and pressure. This is a condition that is applicable for most biochemical reactions (Grace 2004). It is the most widely used thermodynamic property in biochemistry, and is also an appropriate thermodynamic property to cover molecular resolution, where the distinction between heat and other forms of energy disappears (Wolfe

2001). Table 4.4 shows the physical properties applicable at the level of cellular resolution and at the level of molecular resolution. The Gibbs free energy change predicts whether reactions (interactions) can occur spontaneously, or whether energy must be supplied for a reaction to occur. For any spontaneous reaction, the free energy of the products must be less than the free energy of the reactants, i.e. ΔG must be negative (Grace 2004). The value of ΔG depends on the standard free energy change for the reaction ($\Delta_r G^\circ$) and concentrations of reactants and products. A critical point is that the metabolic processes are governed by the activities of key enzymes rather than by the law of mass action (Stryer 1988).

Table 4.4: Macroscopic and microscopic views in cellular thermodynamics

	Macroscopic	Microscopic
Physical entity	Cell	Molecule
Domain	Thermodynamics	Newtonian or quantum mechanics
Temperature	Well defined and measurable	Not applicable
Pressure	Well defined and measurable	Not applicable
Work, heat and kinetic energy	Clearly distinguished	Distinction almost disappears

Every reaction has a propensity of reaching an equilibrium state. This is driven by Gibbs free energy, since propensity is due to a system trying to minimise its Gibbs free energy. Within a cell chemical reactions are driven towards a local thermodynamic equilibrium, where the intensive properties vary in space and time, as opposed to a global thermodynamic equilibrium, where intensive properties are homogeneous throughout a system. However, a local thermodynamic equilibrium varies so slowly, that one can assume thermodynamic equilibrium within a particular neighbourhood.

The ΔG of a reaction is influenced by the ΔG_f^0 of reactants and products and the abundance of reactants and products. For a simple reaction ΔG is given by Equation 4.1.

$$\Delta G = \Delta G^o + RTln\frac{[B]}{[A]} \qquad \text{Equation 4.1}$$

where $\Delta_r G^0 = G_f^0{}_B - G_f^0{}_A$ is the difference in standard free energy of formation between molecules A and B, R is the universal gas constant, and [A] and [B] are the concentrations of A and B. At equilibrium, when ΔG_r is zero, the difference in standard free energy between molecules A and B is exactly compensated by their concentration difference (Dogterom 2001). Although the value of $\Delta_r G$ determines, whether a reaction can occur spontaneously, it does not make any predictions about the speed of the reaction (Dogterom 2001). The reaction rate is controlled by activation energy, which determines the number of successful associations that overcome the activation barrier.

Figure 4.8 shows a typical enzyme catalysed reaction. For a single elementary reaction k_{cat} which is the rate of catalysis for a function of k_3. The turnover number which is equal to kinetic constant k_3 of an enzyme is "*the number of substrate molecules converted into products by an enzyme molecule in a unit time, when the enzyme is fully saturated with substrate*" (Stryer 1988). The catalytic step is assumed to be constant for a particular temperature and pH, since it is a property of the enzyme. The turnover cycles (processing time) for enzymes are distinct. Carbonic anhydrase has one of the shortest known turnover cycles namely, 1.7 microseconds per cycle. However, the ultimate limit on the value of enzymatic velocity is set by k_1, which is the rate of formation of ES complexes. However, this rate cannot be faster than the diffusion controlled encounter of an enzyme and its substrate (Stryer 1988). Diffusion limits the value of k_1, so that the velocity of a chemical reaction cannot be higher than between 10^8 and 10^9 M^{-1} S^{-1}. In fact the catalytic velocity of a enzymes actylcolinesterase, carbonic anhydrase and triosephosphate are between 10^8 and 10^9 M^{-1} S^{-1}, which shows that they have attained kinetic perfection (Stryer 1988). Their catalytic velocity is restricted only by the rate, at which they encounter substrate (i.e. k_1) in a cellular environment.

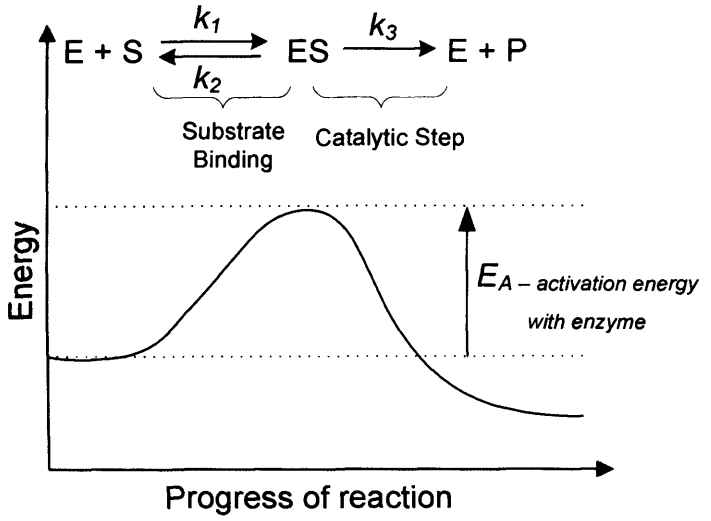


Figure 4.8: An elementary enzyme catalysed reaction.

4.6 The Cellular Environment

The physical space, which consists of native biomolecules and their environment is perceived as the Euclidean space, where spatial and temporal dimensions are distinct. The environment space represents resources, such as metabolites, ions and other macromolecules (i.e. foreign biomolecules). The activities amongst native biomolecules, and between native biomolecules and foreign biomolecules occur in the physical space. Enzymes are major contributors to cellular metabolism. In a typical enzyme catalysed reaction, the enzymes represent the native biomolecules and the metabolites are represented as part of the environment. Every region of the metabolite space holds a local population. These spaces are used by various chemical reactions to produce or consume metabolites. The reactants continuously diffuse based on diffusion gradients. Hence every metabolite space will be in constant flux, due to various biochemical reactions occurring in different regions of the cellular environment. The reactants and products at a particular locality will determine its free energy levels and the probability of a reaction occurring. This also affects the probability of substrates encountering relevant enzymes in the specified region. Enzymes assist in this process by accelerating the attainment of equilibrium but do not shift their position. The catalytic velocity is restricted by the rate at which the enzymes encounter substrates (k_1) in the cellular environment. The ATP which is known as the universal currency of free energy in biological systems is a widely used

metabolite to couple with non-spontaneous reactions in metabolism. Yet the ATP → ADP is always maintained far from equilibrium (i.e. Δ G is always negative) in the cells.

4.7 Concluding Remarks

The Collective Intelligence approach is well suited to represent the adaptability that is a feature of the collective behaviour of biomolecules. Cellular Intelligence is defined as the ability to regulate when, where and what biomolecular activities occur to maintain biological equilibrium in diverse environments. Hence modelling the collective behaviour of biomolecules will involve representing cellular adaptation utilising Swarm/Collective Intelligence. The concepts of self-organisation and emergence underlie swarming and these systems are inherently adaptive, robust, flexible, stochastic and concurrent. The first step towards modelling intracellular organisational behaviour is understanding the mechanisms that foster collective behaviour among biomolecules. The main features of a Swarm Intelligence approach involve forms of limited or minimal communication and/or interaction, large numbers of interacting entities with limited reach, and some global efficient, emergent or self-organising behaviour (Fleischer 2003). Further the four basic ingredients for the manifestation of self-organisation are (Bonabeau, Dorigo et al. 1999): Forms of positive feedback, forms of negative feedback, amplification of fluctuations, multiple interactions of multiple entities. The existing Swarm Intelligence techniques are not able to represent the intracellular adaptive dynamics since they pursue goals/objectives which are spatial in nature and strive to explore and exploit solutions in space within the bounds of their temporal constraints. Hence new techniques based on biomolecular inspired mechanisms will have to be developed in order to pursue goals/objectives which are temporal in nature and strive to explore and exploit solutions in time within the bounds of their spatial constraints. The Collective Intelligence framework is based on a meta-formalism, which can be used for complex and self-organising systems. The problem framework is based on Cellular Intelligence, that represents a biological cell's ability to organise and adapt to perturbation and uncertainty, which reflects on the characteristics of intelligence. The fundamental principles utilised are self-organisation and thermodynamics to represent biological and physical constraints, respectively. The dynamic framework utilises

multi-objective topology as a core of the model and describes the logic of Collective Intelligence, which is used to construct/deconstruct tasks for intracellular organisational behaviour of the cell in physiological timescale. The next chapter provides the model specification, which implements a SwarmCell model, and utilises an agent based formalism in the wider framework of Collective Intelligence to conduct analysis studies of a biological cell.

Chapter 5

A Collective Intelligence Approach to Modelling Intelligent Cellular Organisation

"It is possible make things of great complexity out of things that are very simple. There is no conservation of simplicity."

- Stephen Wolfram

5.1 Overview

The aim of this chapter is to provide a model specification based on the problem definition and model requirements discussed in Chapters 2 and 3 respectively. Section 5.2 describes the model specification. It provides an overview, giving an idea of the model's focus, resolution and complexity. Section 5.2.1 specifies the purpose of implementing the SwarmCell model. Section 5.2.2 specifies the representation of biomolecules and their attributes, and describes the spatial and temporal scales utilised to model interactions. Section 5.2.3 describes the execution of the model in terms of scheduling sub-models. Section 5.2.4 describes design concepts of the model in terms of representing emergence, adaptability, objectives, learning, prediction, sensing, interactions, stochasticity, collectives and observation. These design concepts facilitate integration of the agent based formalism into a wider framework of Collective Intelligence. A description of the model in detail is given in Section 5.2.5 through the model initialisation process. Section 5.2.6 covers the model inputs during the simulation and Section 5.2.7 the sub-models details. The model specification provides required functionalities to implement a SwarmCell model and conduct analysis studies from molecule to cell level using simulation experiments.

5.2 The Model

The model specification of the SwarmCell prototype is described, using the ODD protocol for agent based models (Grimm, Berger et al. 2006), developed by the open agent based modelling consortium (Open ABM Consortium 2010). The prototype utilises an agent based formalism to model the collective behaviour of native biomolecules. Concepts from Complex Adaptive Systems (CAS) are applied to build a mechanistic model of a biological cell. The physical space represents the core of the model, where interactions among native biomolecules and resources occur (see Figure 5.1). The resources are modelled as part of a native biomolecule's environment, and the rules for biomolecular mobility and interactions (i.e. agent-agent and agent-environment interactions) are also represented. The model is targeted for use in analysis and design studies. For design studies, it requires development of appropriate adaptive algorithms, based on Collective Intelligence, especially biomolecular inspired algorithms. Tradeoffs between details of reality derived from theoretical foundations, i.e. from fundamental principles, and the feasibility of modelling with relevant data has been an ongoing issue, during the mechanistic model development process. The level of model detail required is determined by the aim, hypothesis and available data for the investigation.

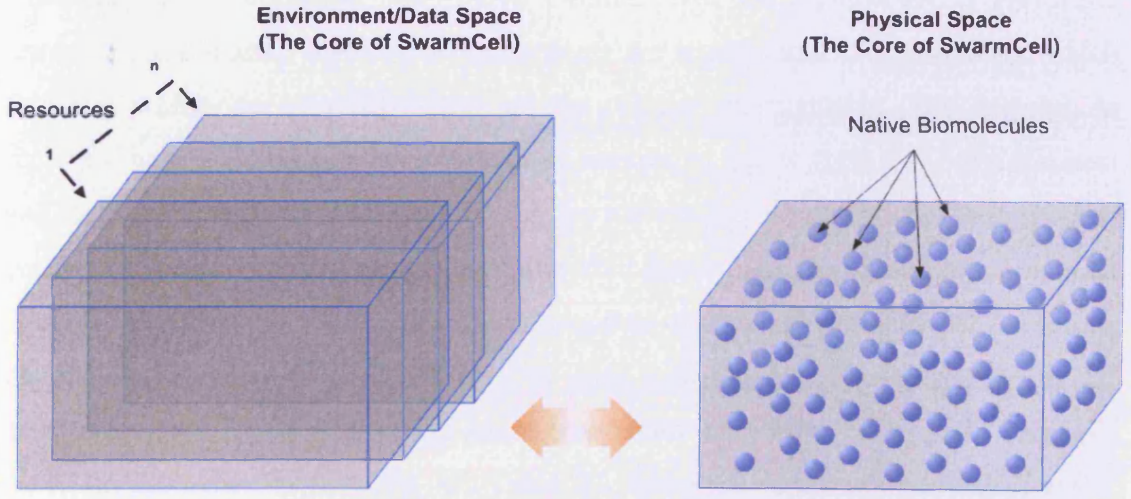


Figure 5.1: The core of the SwarmCell model representing the physical space

5.2.1 Purpose

The purpose of a SwarmCell model is to study collective behaviour of native biomolecules, constituting a biological cell by applying the principles of self-organisation and logic of Collective Intelligence to model intracellular tasks and associated constraints. The aim of the study is to define cellular functions in the context of a multi-objective topology and implement them as an *in silico* mechanistic model, to study the performance of intracellular functions by measuring activities of diverse species of functional products. The model integrates biomolecular activities occurring within the gene, transcript, protein and metabolite space. It represents various stages of regulation to correlate between perturbations and performances of a cell.

5.2.2 Entities, State Variables and Scales

The model utilises a bottom-up approach, where the lowest and highest levels of model representation are at the molecular and cellular resolution, respectively. The model comprises the following types of entity; native biomolecules - represented as individual entities, biomolecular species and tasks - represented as collectives, grid cells represented as spatial units, i.e. environmental conditions that vary over the physical space in the molecular environment, and resources such as metabolites represent the environment. The native biomolecules are represented at molecular resolution, while their metabolite counterparts are represented as populations, which fluctuate within the physical space of the cellular environment. The decision to represent native biomolecules as individual entities, is due to their low copy numbers and state specific behaviours. In contrast the metabolites do not posses state specific behaviour and are usually found in greater abundance. The environment is modelled as a Euclidean space, where spatial and temporal dimensions are discrete. This three dimensional space is divided into distinct grids with the time dimension represented as discrete time steps. The native biomolecules are characterised by state variables based on their role in the cells, see Table 5.1. Metabolites are represented as scalar fields, since they do not have state specific behaviour as found in native biomolecules. Each distinct grid in the metabolite space contains an abundance (number) of various metabolites, ions and other resources.

Table 5.1: State variables and types of biomolecules

Biomolecule	State variables	Types
Genes	Name, Location, Base length	N/A
Transcripts	Name, Base length, Half-life, locality and Functional state	mRNA, tRNA, rRNA and smallRNA
Proteins	Name, Locality, Polypeptide length, Conformational state, Functional state, Temperature range, pH range and Half-life	Enzymes, Transport proteins, Messenger proteins and Mechanical proteins

Biomolecular species are characterised by state variables, such as its biomolecular species name, number of individual biomolecules present, and their overall activity levels. Since the functional product representing a gene can eventually transform into a transcript or protein, the overall activities of these functional products will reflect on the significance of respective genes.

Scales represent the nature of the objectives/tasks forming the complex hierarchy of mutually dependent activities of a cell. Cellular functions are quantified in terms of performance of solutions, which are constructed/deconstructed in terms of objectives/tasks of a cell. The solution space is deconstructed into individual molecular level tasks - consisting of the contribution made by individual biomolecules, group level tasks - consisting of the contribution made by the same biomolecular species, and team level tasks - consisting of the contribution made by the heterogeneous species of biomolecules. Categorising the collective behaviour of functional products in terms of objectives/tasks can deconstruct the global objectives/tasks of a cell into the basic tasks required to pursue them. Hence these tasks are characterised by state variables, such as their objective and performance. The construction of global tasks from basic tasks requires associating performance from basic to global tasks of a cell. Performance can be a measure of a biomolecular species' activity level in the case of tasks containing a single biomolecular species, and the number of completed tasks in the case of tasks containing more than one biomolecular species.

The temporal scale, which begins at milliseconds, can represent most of the native biomolecular activities. Although certain enzymes perform their activities in the order of microseconds, and molecular binding activities take place in the order of nanoseconds their interactions are treated as instantaneous. Since the model is using a bottom-up approach, only a single timescale is used to represent discrete biomolecular activities. However, a time step can be represented at any timescale suited to the analysis study. Hence, usually a time step represents one millisecond and the simulations run from several seconds to minutes depending on the scenario. The spatial scale is represented on a nanometre scale, where each grid cell represents 200 nanometres and the physical interaction space comprises 2 x 2 x 2 micrometer^3(8 μm^3).

5.2.3 Process Overview and Scheduling

The model proceeds in millisecond time steps. Processes scheduled are based on model scenarios (simulation experiments). Within each time step, two processes are executed in the following order: biomolecular movement, biomolecular interactions (interactions between native biomolecules, and between native biomolecules and resource molecules). Processes such as biomolecular degradation and reproductive errors are also executed depending on the scenario. Since every type of interaction has a specific time duration (event interval), the biomolecular event intervals are scheduled using a dynamic event scheduler. The events are scheduled asynchronously, since pre-conditions for a particular event must be satisfied before the event can commence. Events are terminated when post-conditions are satisfied for a particular event. The initiation of events and the spacetime intervals among events are modelled stochastically. Between each time step, a diffusion of resources occurs and the abundance of resources are updated synchronously by the scheduler for every grid cell.

5.2.4 Design Concepts

5.2.4.1 Emergence

Analysing emergent behaviour of a cell, which occurs as a result of biomolecular activities, is one of the objectives of this study. The lowest abstraction level is

represented by biomolecules, and the highest abstraction level is represented by the organisation of a cell. The emergent dynamics occur, due to uncertain nature of the cellular environment, as explained in section 2.4. Global level analysis of emergent behaviour is conducted using a network topology and a multi-objective topology. The network topology, which provides a sequential representation of the emergent dynamics, is used to trace the flow or propagation of resources, energy and signals within a cell. Biochemical pathways will emerge as a result of these fluxes. The multi-objective topology, which provides a concurrent representation of emergent dynamics, is used to capture the performances of a native biomolecular species and their collaborations, which provide vital solutions to higher level objectives/tasks of a cell. The rules that define the goals of biomolecules, aim to produce generalisable outcomes for the heterogeneous swarm. Although deliberately designing agents with global awareness may sound interesting and satisfying, it will not lead to generation of emergent outcomes of intracellular organisational behaviour. The logic for reactive agents is much simpler than that of intelligent agents.

5.2.4.2 Adaptation

Adaptations occur at individual and organisational levels in a cell. Instability due to variations, is the primary contributor to adaptations in biological systems. Variations are caused by alterations in biomolecular activity. At the individual level, variations occur at evolutionary timescales, where the sources (genetic) of native biomolecules are altered, causing the rules of behaviour to change. These rules govern the state, transition of states and state specific behaviour. However, at the physiological timescales, the rules do not change, rather execution of the rules changes the states of native biomolecules via transitions, which trigger state specific behaviour. These are simply reactive behaviour, having predictable state specific responses to the dynamically changing environment. However, when and where these rules are executed by biomolecules depends on the stimulations they receive from their local environment. Hence by influencing a particular biomolecular response via intrinsic (programmed stimuli) and extrinsic stimulations, a biological cell is able to adapt to a dynamically changing environment at the physiological timescale. At organisational level, adaptations occur at physiological, developmental and evolutionary timescales. While at physiological and developmental timescales, the state transitions and state specific behaviour of native biomolecules contribute to organisational adaptation, at

evolutionary timescales changes in individual rules contribute to organisational adaptation. The propagation of phases of a cell cycle, are periodic adaptations, which are guided mainly by intrinsic stimulations. The conformational changes of biomolecules that occur by sensing the environment are an individual level adaptation to maintain molecular stability.

5.2.4.3 Objectives

A biological cell is organised into objectives/tasks, which range from basic to complex global level tasks of a cell. The tasks at an individual level of native biomolecules are basically to perform intended activities, which are basic and independent. These tasks are common to individuals from a particular biomolecular species, and include simple biochemical transformations and binding activities. However the collective performance of individuals from distinct biomolecular species will determine the impact on a cell. This is usually the sum of the activities of a particular biomolecular species in a stipulated time frame. The fundamental unit of objective solution is a cooperative task, which consists of individuals representing one or more native biomolecular species cooperating for a common objective/task. When more than one biomolecular species is involved in the formation of a task, mutual dependency will exist among the biomolecular species. Hence there is a gradual transition from objectives being independent at the molecular level to mutual dependency of objectives at the cellular level. The mutual dependency of higher objectives is due to the concurrent effects of various lower level tasks on higher level objectives. The multi-dimensional solution space is represented by objectives, which are organised into a hierarchy, where each level is semantically different from other levels. Fitness is a measure of the performance of these objectives. At the molecular level performances are measured by the level of activities of native biomolecular species. The effect on the multi-dimensional problem space is determined by quantifying the performance of these objectives.

5.2.4.4 Learning

Native biomolecules are unable to change their adaptive traits (rules) over time by taking account of their experience, since they are simple reactive entities with no known cognitive ability. However their adaptive traits are modified as a result of random alterations in the genome and appropriate modifications will persist. Learning

is an organisational level property of a cell and is facilitated by various regulatory activities of native biomolecules. It is a process of reconfiguring by searching for when, where and what biomolecular activities should occur, based on adaptive requirements. It is a compromise between qualitative features of native biomolecules (what activities can be done under what conditions) and quantitative features of their regulation in a cell (when the activities occur) and multi-cellular structures (where the activities occur). However, the process of learning does not occur at physiological timescales.

5.2.4.5 Prediction

Predictability refers to the proactive behaviour of a biological cell, where it is more likely to anticipate and deliver timely responses to perturbations. However this behaviour is most prominent in evolutionary timescale, where regulation of intracellular activities is gradually shifted to a point of intended activity. This is a reconfiguration process, and acceptable performances can be observed within the bounds of their original configuration. Cells have developed rhythmic behaviour (oscillations) in regulating biomolecular activities between stimulations and responses. The frequency of occurrence of these stimulation and responses will reinforce the rhythms.

5.2.4.6 Sensing

A native biomolecule is able to sense itself and its immediate environment. Hence it is assumed to know its own location and state, so that it applies its state specific behaviour. The local environment of a molecule is its neighbouring molecules, located within a grid cell.

5.2.4.7 Interactions

Redundant interactions of specialised biomolecules will give rise to extremely concurrent biomolecular activities, distributed in time and space within a cellular environment. Direct interactions occur when native biomolecules come into physical contact with each other. With enzymes these interactions will produce catalytic activities and with signalling molecules, these interactions produce binding activities. Signalling activities contribute to feedback among biomolecules. Indirect interactions occur, when enzymes use shared environment to interact with mediating metabolites.

Interactions in the form of influences between tasks/objectives occur via competition and cooperation. Cooperating biomolecular species will have beneficiary effects on each other, whereas competing biomolecular species will have inhibitory effects on each other.

5.2.4.8 Stochasticity

Stochasticity of intracellular activities manifests in many forms. These include fluctuation in abundance and distribution of biomolecules, and in activities and their effects. These factors contribute to cellular dynamics at the physiological timescale. Fluctuations in terms of where, when and what biomolecular activities occur, is regulated by various feedback mechanisms. Feedback can be specific to a species of biomolecules, (via direct physical contact as in activation and inactivation), or to a group of biomolecular species, or in general to all biomolecular species (saturation). Stochasticity caused by an uneven distribution of biomolecules in space and time, is modelled as biomolecular mobility using Brownian dynamics. Although the duration of biomolecular activities can fluctuate at the evolutionary timescale due to alterations in their rules, they remain relatively constant at the physiological timescale. Hence, while duration of activities remains constant, the intervals amongst activities are modelled stochastically. The presence of neighbouring molecules will have an important effect on free energy levels of the reactants, products and intermediate states. The total density of non-water molecules in a cell is very high and can be referred to as a crowded environment. The effects due to macromolecular crowding may alter both the reaction equilibrium and reaction rates in a non-specific way. Hence, native biomolecules of the same species can have heterogeneous behaviour due to altered states.

5.2.4.9 Collectives

Collectives can be members from the same biomolecular species or different biomolecular species collaborating for a common objective/task. The type of mutual dependency (competition or cooperation) amongst objectives depends on the performance interaction between tasks. While positive feedback facilitates cooperation, negative feedback controls competition. A cooperative module is defined as a group of one or more species of a functional product, collaborating for a common objective/task. At the cellular level mutual dependency of biomolecules facilitates

cooperation, and controls competition to form an organisational closure. Hence, positive feedbacks coupled to negative feedbacks facilitate in stabilising the cellular organisation. The mutual dependency of biomolecular activities causes concurrent effects on higher objectives. This enables capture of indirect (invisible) interactions, which occur due to causally linked activities.

5.2.4.10 Observation

The dependent variables represent data points for the simulation, which are used for analysis of the experimental results. These variables represent the abundance of biomolecules, the levels of native biomolecular activities, simulation time, free energy levels, and states of the chemical systems. To analyse intracellular activities, the aggregate variables are abstracted into a network topology and a multi-objective topology to provide a sequential and concurrent view of the intracellular dynamics, respectively. The population level variables, such as abundance of metabolites, are analysed using dynamic graphs and abstracted using a network topology for qualitative visualisation of flux directions. The level of activities is used to measure performances of tasks/objectives and is abstracted using multi-objective topology. In the multi-objective topology, biomolecular activities are abstracted based on competition and cooperation of native biomolecules. The performances of these tasks are analysed, based on the ensemble activities that vary with time and are based on comparing relative performances among tasks, see Figure 5.2.

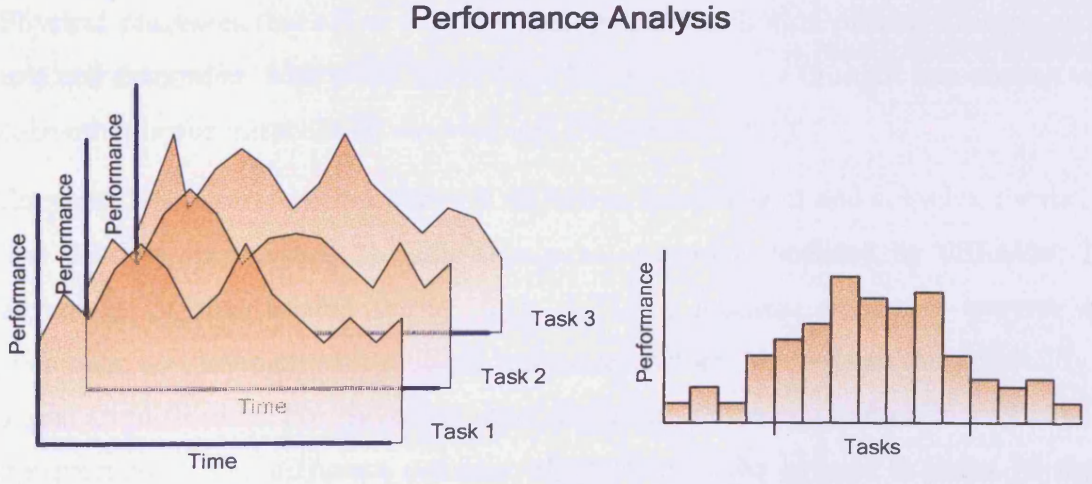


Figure 5.2: Analysis of performances

The platform accommodates multi-scale visualisation and semantic zooming to analyse intracellular activities between levels of molecular resolution and cellular resolution.

5.2.5 Initialisation

Although various species of native biomolecules have distinct locations in the cellular environment, due to alack of molecular distribution information, a significantly small proportion of molecular species are randomly distributed in the intracellular environment. Further to represent various cell types (humans have over 200 different cell types) a cell model must be able to initialize with an appropriate molecular population to represent differentiated cell types.

5.2.6 Input Data

The model does not use input data to represent time varying processes.

5.2.7 Sub-models

The sub-models specify some of the universal constraints, contributing to internal organisation of a cell, such as biomolecular mobility, biomolecular interaction, biomolecular degradation and error frequencies in transcription and translation.

5.2.7.1 Biomolecular mobility

Physical processes that affect mobility and spatial distribution of native biomolecules in a cell determine, where and when these biomolecules are brought into contact with each other in the intracellular environment (Dogterom 2001).

Enzymes find their substrate through diffusion, confinement and complex formation. The delivery of enzymes to their substrates is mostly mediated by diffusion. The advantage of this is that during their diffusive journey, an active enzyme can encounter and interact with multiple substrates, which give rise to the possibility of signal amplification. For travelling short distances, diffusion is an efficient means of transportation. The diffusion constant of proteins in the cytosol is about 10 times lower than in pure water. However, the diffusion constant for small ions in a cell is virtually the same as pure water (Dogterom 2001). Table 5.2 provides some diffusion constants for some metabolites, ions and macromolecules.

Table 5.2: Diffusion Coefficients in aqueous solutions (Source (Nobel 2009))

Substance	Diffusion Constants	
	$D_j(m^2s^{-1})$	$D_j(\mu m^2s^{-1})$
Small solutes in water		
Alanine	0.92×10^{-9}	920
Citrate	0.66×10^{-9}	660
Glucose	0.67×10^{-9}	670
Glycine	1.1×10^{-9}	1100
Sucrose	0.52×10^{-9}	520
Ca^{2+} (with Cl^-)	1.2×10^{-9}	1200
K^+ (with Cl^-)	1.9×10^{-9}	1900
Na^+ (with Cl^-)	1.5×10^{-9}	1500
CO_2	1.7×10^{-9}	1700
Globular proteins in water		
Molecular mass (kDa)		
15	1×10^{-10}	100
1000	1×10^{-11}	10

A random walk in Brownian motion is a sequence of steps of constant size δ in a space at regular time intervals τ, where each subsequent step is chosen to move in a new randomly chosen direction. The Einstein equation shows that the average distance Δr that a molecule performing a random walk travels from its starting point, increases with the square root of time, and depends on the dimensionality of diffusion (see Equations 5.1, 5.2 and 5.3).

For three dimensional diffusion

$$\Delta r = \sqrt{(6Dt)} \qquad \text{Equation 5.1}$$

For two dimensional diffusion

$$\Delta r = \sqrt{(4Dt)} \qquad \text{Equation 5.2}$$

For one dimensional diffusion

$$\Delta r = \sqrt{(2Dt)} \qquad \text{Equation 5.3}$$

Where $D = \delta^2\tau$ is the diffusion coefficient

Appendix F:3 and F:4 show the relationship between diffusion constant vs distance travelled for particular time steps and time required to travel particular distances. For a typical protein of the size of a few nanometres, it will take about 0.03 s to travel 1 μm in three dimensional diffusion space, assuming a diffusion constant of 5 $\mu m^2 s^{-1}$. Table 5.3 and Table 5.4 list distances travelled for particular timeframes and the time required travelling particular distances, respectively in three dimensional diffusion space.

Table 5.3: Distance travelled during various time steps based on diffusion constant

Diffusion Constants	Distance travelled		
	$second^{-1}$	$millisecond^{-1}$	$microsecond^{-1}$
Typical size protein			
5 $\mu m^2 s^{-1}$	5.5 μm 5500 nm	0.17 μm 170 nm	0.0055 μm 5.5 nm

Table 5.4: Time required travelling a particular distance based on diffusion constant

Time required	Diffusion Constants	
	5 $\mu m^2 s^{-1}$	66 $\mu m^2 s^{-1}$
0.1 μm/100 nm	0.3 ms	0.025 ms
0.2 μm/200 nm	1.3 ms	0.10 ms
1 μm	0.03s/30 ms	2.5 ms
1 mm	9h 15 minutes	42 minutes

Mobility/diffusion is normalised between the mobility of individual native biomolecules and the population based diffusion used by resources, based on Table 5.4. The native biomolecules are scheduled to move every 1.3 ms and resources are scheduled to diffuse every 0.1 ms since every grid cell is divided into 200 nm^3.

5.2.7.2 Biomolecular interactions

Native biomolecular interactions are driven by a favourable change in free energy, which occurs when molecules interact with their substrates. Biomolecular interactions in the cellular environment are facilitated by depletion forces, random walk, directed transport, and confinement of molecules (co-localisation) to domains in the membrane or other structures. Every native biomolecular species' interactions are different. Hence the pre- and post-conditions for these interactions will also differ. Biomolecular interactions occur between native biomolecules in signalling and regulatory activities, and between native biomolecules and resource molecules in metabolic activities. Every type of interaction has a specific time duration (event interval) and is triggered at different times. The dynamic event scheduler is used to model initiation and the termination of various biomolecular events. The events are initiated, when pre-conditions for particular events are satisfied and when these events are terminated the post-conditions are satisfied for respective events.

In metabolic activities pre-conditions are determined by free energy state of a reaction (see Equation 5.4).

$$\Delta G = \Delta G^{o} + RT ln \frac{[B]}{[A]} \qquad \text{Equation 5.4}$$

Where $\Delta G^0 = G^0_B - G^0_A$ is the difference in standard free energy between molecules A and B, $R = N_A k_B$, where R is the universal gas constant and, k_B is Boltzmann's constant, and [A] and [B] are the concentrations of A and B. At equilibrium, when ΔG is zero, the difference in standard free energy between molecules A and B is exactly compensated by their concentration difference (Dogterom 2001).

The standard ΔG_r (reaction) can be calculated from the standard ΔG_f (formation) of reactants and products or from empirically measured values. The enzymes are able to sense the free energy values of their respective reactions and decide to proceed with the reactions if ΔG_r is negative. Although the value of ΔG_r determines whether a reaction can occur spontaneously, it does not make any predictions about the speed of the reaction. The reaction rate is controlled by the activation energy. It determines the number of successful associations that overcome the activation barrier.

Table 5.5: Maximum turnover number and the calculated turnover cycle of some enzymes (Source (Stryer 1988))

Enzymes	Turnover number (s^{-1})	Turnover cycle
Carbonic anhydrase	600,000	1.7 μs
3-ketosteroid isomerase	280,000	3.6 μs
Acetylcholinesterase	25,000	40 μs
Penicillnase	2,000	500 μs
Lactate dehydrogenase	1,000	1 ms
Chymotrypsin	100	10 ms
DNA polymerase I	15	66.7 ms
Tryptophan synthetase	2	500 ms
Lysozyme	0.5	2 s

The turnover cycle of enzymes, shown in Table 5.5, represents processing time of enzymes. The enzymatic activities are recorded, when processing is complete. The inputs for the process, are the reactants and the output will be the products, including by-products. The processing time is represented as event intervals.

5.2.7.3 Biomolecular Degradation

Similar to radioactive decay, biomolecular degradation is a statistical process, which depends upon the instability of particular biomolecular species. The predictions of biomolecular degradation can be stated in terms of the half-life, the degradation constant or the average lifetime. The relationship between these quantities is given in Equation 5.5.

$$T_{1/2} = \frac{ln2}{\lambda} \approx \frac{0.693}{\lambda} \approx 0.693\tau \qquad \text{Equation 5.5}$$

Where $T_{1/2}$ is the half-life, λ is the degradation constant and τ is the average/mean lifetime. The degradation process and the observed half-life dependence of ubiquitination can be predicted by assuming that individual biomolecular degradations are purely random events. If there are N biomolecules at some time t, then the number ΔN, which would degrade in any given time interval Δt, would be proportional to N. The Equation 5.6 provides the relationship between these factors.

$$\Delta N = -\lambda N \Delta t \quad \text{Equation 5.6}$$

Where λ is a constant of proportionality, which is also called the degradation constant. Table 5.6 lists the half-life, average lifetime and degradation constants for proteins containing various amino terminal residues, which were calculated using Equation 5.5.

Table 5.6: Half-life, average lifetime and degradation constants for proteins containing various amino-terminal residues.

Amino-terminal residue	Half-life (Minutes)	Average Lifetime (Minutes)	Degradation constants Minute^{-1}	Degradation constants Second^{-1}
Stabilizing				
Methionine				
Glycine				
Alanine	>1200	1731.6	0.0005775	9.625×10^{-6}
Serine				
Threonine				
Valine				
Destabilizing				
Isoleucine				
Glutamate	~30	43.3	0.0231	3.85×10^{-4}
Tyrosin				
Glutamine	~10	14.43	0.0693	1.12×10^{-3}
Proline	~7	10.1	0.099	1.65×10^{-3}
Highly destabilizing				
Leucine				
Phenylalanine				
Aspartate	~3	4.33	0.231	3.85×10^{-3}
Lysin				
Arginine	~2	2.89	0.3465	5.78×10^{-3}

Biomolecular degradation can be implemented by imposing the process using probabilities derived from the degradation constant, or by using the average life time for degradation to emerge from the process.

Probability approach: the degradation constant (constant of proportionality) of the respective native biomolecules is assigned during initialisation. These constants will depend on the resolution of the time steps and they are inherently normalised. The process of degradation occurs during every time step, where random probabilities generated by every biomolecule are compared with its degradation constant. If the condition is satisfied the biomolecule is removed from the environment.

Average life time approach: the average life of a native biomolecular species is assigned and the age of every native biomolecule is randomly determined during initialisation. From this data the remaining life time is calculated and biomolecules are removed from the environment, when they reach their life time. The process of degradation can be scheduled based on the remaining life time, or by incrementing age and checking whether the biomolecule has reached its life time during every time step. The increment depends on the resolution of the time step (minutes or seconds).

5.2.7.4 Error Frequencies

The probability p of forming a protein with no errors depends on n. This is shown in Equation 5.7.

$$p = (1- E)^n \qquad \text{Equation 5.7}$$

Where p is probability of forming a protein with no errors, n is number of amino acid residue, and e is frequency of inserting a wrong amino acid, and it is known to be 10^{-4} per amino acid residue. The error frequency of RNA biosynthesis is about 10^{-4} to 10^{-5} per nucleotide residue. The error frequency of DNA biosynthesis is about 10^{-9} per nucleotide residue. Table 5.7 lists probabilities of forming a protein with no errors, which depends on the lengths of polypeptide chain.

During protein synthesis probabilities are calculated based on length of the mRNA, and the decision to produce an error free protein is made based on a randomly generated number.

Table 5.7: Probabilities of forming a protein of various lengths with no errors

Number of amino acid residue (n)	Probability
100	0.99
200	0.98
300	0.97
400	0.96
500	0.95
600	0.94
700	0.93
800	0.92
900	0.91
1000	0.90

5.3 Concluding Remarks

This chapter has contributed to the specification of the Collective Intelligence framework utilised in the cell modelling and simulation environment. The purpose of this chapter is to describe the model's focus, resolution and complexity. The model's scope is to study collective behaviour of biomolecules constituting a biological cell. The model utilises a bottom-up approach, where lowest and highest levels of model representation are at molecular and cellular resolution, respectively. The processes scheduled are based on model scenarios, which can include various combinations of sub-models. The design concepts of the model represent emergence, adaptability, objectives, learning, prediction, sensing, interactions, stochasticity, collectives and observation. These design concepts facilitate integration of an agent based formalism into a wider framework of Collective Intelligence. The sub-models described in this chapter represent some of the universal constraints, contributing to the internal organisation of a cell, such as biomolecular mobility, biomolecular interaction, biomolecular degradation and error frequencies in transcription and translation. The model specification has provided the required functionalities to implement a SwarmCell model and conduct analysis studies from molecules to cell using simulation experiments.

Chapter 6

Swarm Based Cell Modelling and Simulation Environment

"The model is not an oracle, it is an automation of your understanding."

- John Heath

6.1 Overview

This chapter has contributed to implementation of the model specification. It is used to setup and run various simulation experiments based on specific scenarios. The model is implemented using Repast Simphony (Repast S), an agent based discrete event simulation toolkit. Section 6.2 describes the general features of the Repast S toolkit, which is followed by the description of specific features that are used to implement the model specification. Section 6.3 describes the structure of the SwarmCell simulation environment, which was the outcome of the transformation from model specification to implementation, and depends on the toolkit used. The general implementation of a SwarmCell simulation environment, consists of templates for creating various instances of functional products and the physical environment which can represent the metabolites, ions and cofactors. Section 6.4 describes the master context, which contains all components of the model. Section 6.5 describes sub-contexts, which contain the physical space, the pathway layer and the cell layer. Section 6.6 describes implementation of functional products. Section 6.7 describes the shared environment in detail. Section 6.8 describes the model's user interface. Observation components were implemented to analyse activities of the functional products and determine abundance of metabolites.

6.2 Repast Simphony Simulation Package

The model is implemented using Repast Simphony version 1.2, an agent-based modelling and simulation toolkit (North 2006). Repast S is organised into Contexts, Projections and Agents (Howe, Collier et al. 2006). The Context is a form of "proto-space", which provides a container that can maintain a localised state for agents. A context's state can maintain multiple interaction spaces called Projections. Projections are designed, so that they can be used to represent a wide range of abstract spaces, from graphs to grids to realistic geographic spaces. Importantly, projections and agents or individuals are independent of one another. Agents can be agnostic to the type of projection in which they are interacting, and projections can be agnostic to the type of agents, whose relationships they maintain. Finally, the context provides a logical location to maintain agent behaviours that is dependent on localised agent interactions and the environment. The model is developed by creating the main Context and the members of the Context can be Agents, sub-Contexts and Projections. There are many types of Projections in Repast S, consisting of continuous space, grid, network, geography and Scalar fields. Figure 6.1 shows the relevant components utilised for the implementation of the SwarmCell environment.

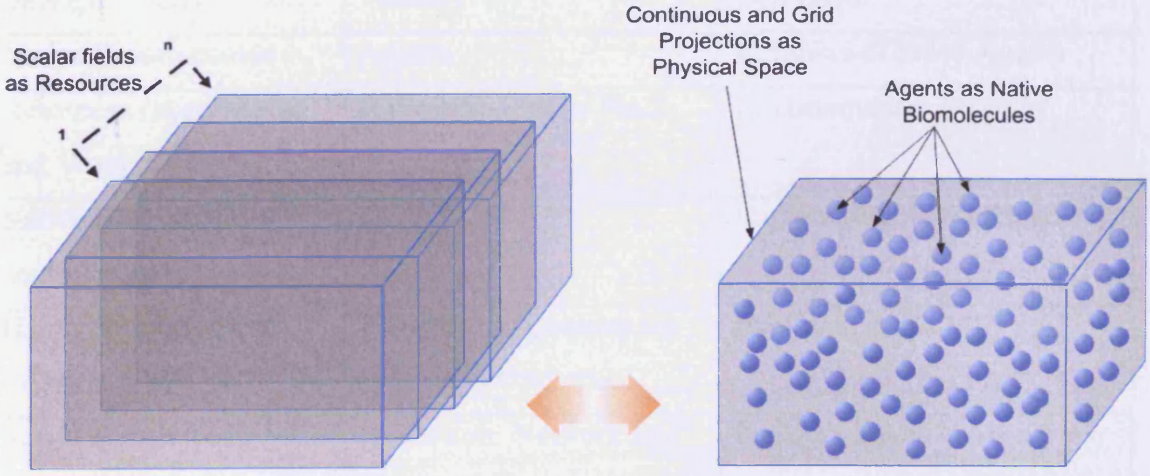


Figure 6.1: Repast S components utilised for the SwarmCell model.

6.3 Model Implementation

The model is implemented using Eclipse and libraries of Repast S version 1.2. The structure of the SwarmCell simulation environment, which was the outcome of the

transformation from model specification to implementation, is based on Repast S. Table 6.1 shows the relationship between the components described in the model specification, the components of Repast S utilised, and representation of those components in the implemented cell model. The general implementation of a SwarmCell simulation environment consists of templates for creating various instances of biomolecules and their environmental resources. In addition the observation components are implemented to analyse activities of native biomolecules and abundance of native and foreign biomolecules. Dynamic graphing is utilised for quantitative analysis of intracellular organisational behaviour. The activities of biomolecules are abstracted into reaction dynamics to observe changes in biomolecular abundance, and performance dynamics to observe changes in biomolecular performances. Qualitative visualisation is performed by representing reaction dynamics in a network topology, and the performance dynamics in a multi-objective topology.

Table 6.1: Relationship between components described in the model specification, Repast S and model implementation

Conceptual model	Repast S Objects	Implementation
Cell Model	Context	SwarmCell core
Biological cell	Context	Cell Layer
Native Biomolecules	Agents	Instances of Proto Agents
Resources (Metabolites and ions)	Projection: Scalar Field	Environment
Native Biomolecules' confinement	Context	Physical/interaction Space
Native Biomolecules' Locality	Projections: Continuous space and Grid space	Euclidian space
Native Biomolecular Interactions	Projection: Network and Graphs	Pathway Layer
Observation	Graphs	Visualisation of output

6.4 The Master Context

The master Context initialises the cell model by creating sub-contexts. It builds and returns a context based on the information provided in the context builder ('CellModelContext'). Building a context consists of filling it with agents, adding projections and creating attributes to pass parameters during initialisation. When the master context is executed, the system will provide a created context based on information given in the sub-contexts, which is specified in Section 6.5. When called for sub-contexts, each sub-context that was added when the master context was built will be executed to create the sub-context. The "model.score" file (see Figure 6.3) is used to create the user interface for the cell model. The structure of the model is shown in Figure 6.2.

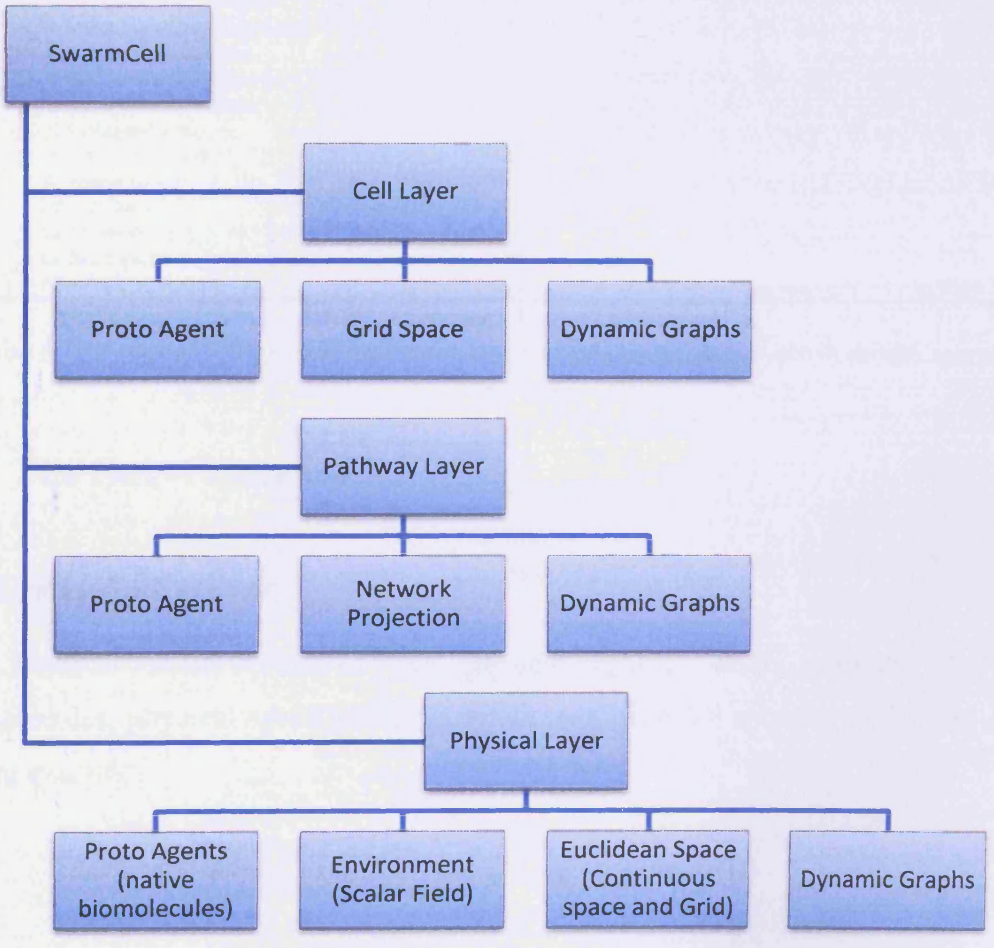


Figure 6.2: Structure of the SwarmCell environment.

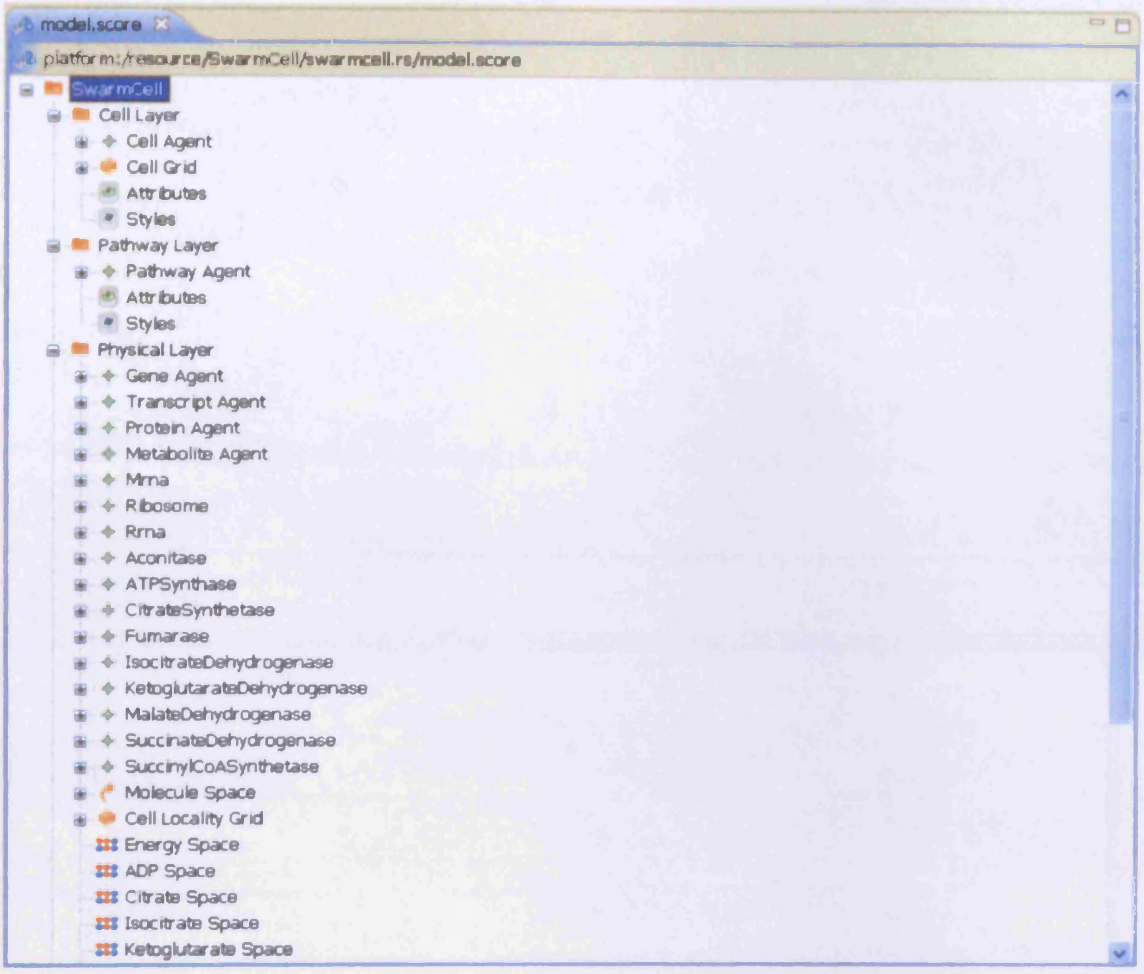


Figure 6.3: Templates for native biomolecules and resources as present in model.score file.

6.5 The Sub-Contexts

6.5.1 Physical Layer

The Physical layer initialises the physical space, which consists of native biomolecules, physical space and environments. Figure 6.4 shows the physical space during execution.

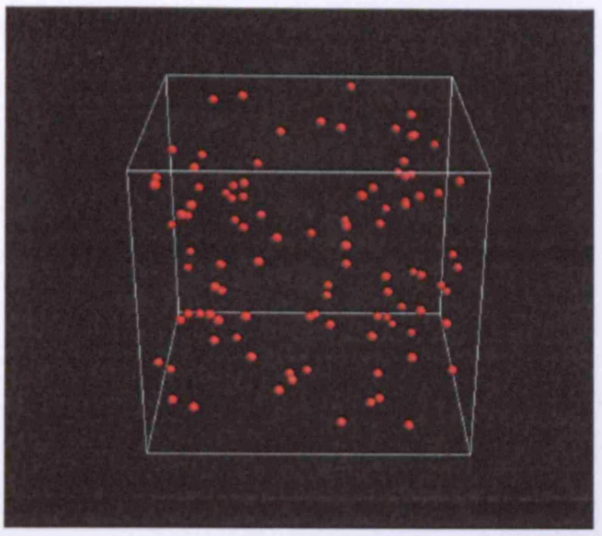


Figure 6.4: Physical Space during execution.

6.5.2 Pathway Layer

The Pathway layer initialises the network projection, which captures the population dynamics of interacting biomolecular species and provides a network representation of the flux directions. Figure 6.5 shows the Pathway layer during execution.

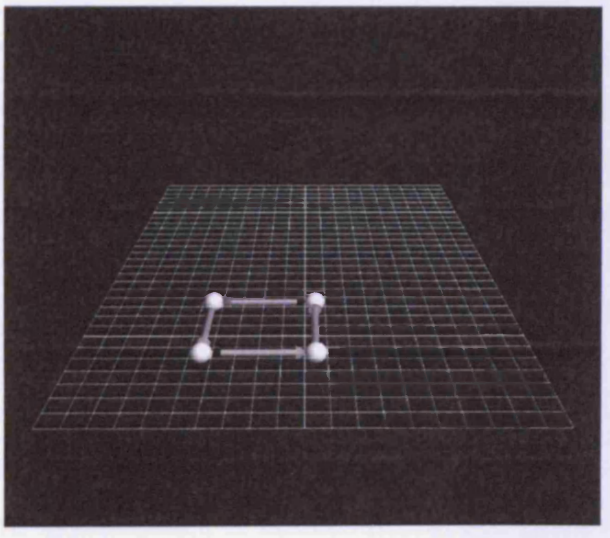


Figure 6.5: Pathway Layer during execution.

6.5.3 Cell Layer

The Cell layer initialises the observation component, which captures the global chemical phases of biomolecular interactions, such as the chemical equilibrium and thermodynamic phase which is utilised to generate the dynamic graphs during

simulation. It also controls the diffusion of scalar fields representing metabolites. Figure 6.6 shows the Cell layer during execution.

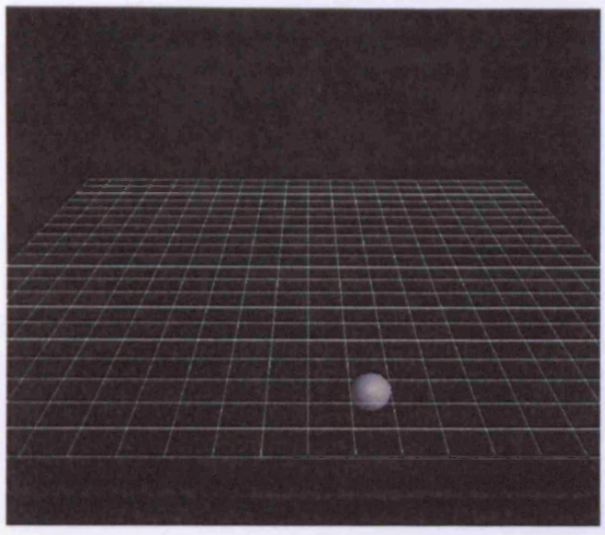


Figure 6.6: Cell Layer during execution.

6.6 Native Biomolecules

Since every native biomolecular species has a unique behaviour, templates are created for each participating native biomolecular species. The variables are represented as attributes and the behaviours as methods. The common attributes and methods are abstracted into higher classes. The highest class for native biomolecules is "MoleculeAgent" which has a direction, speed for mobility and movement.

6.7 Environments

Each foreign biomolecular species (metabolites) is represented in a three dimensional Scalar Field Projection. These projections are subdivided into grid cells, which hold a scalar value representing abundance of metabolites in that grid cell. Figure 6.7 shows scalar field space during execution.

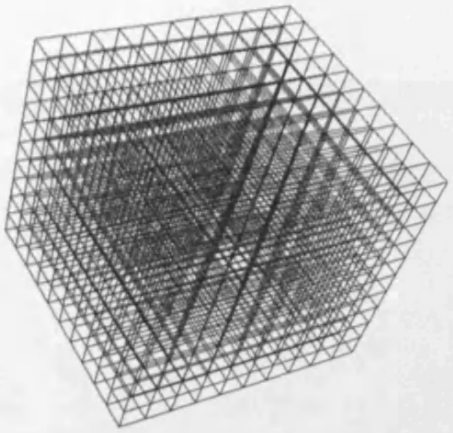


Figure 6.7: Scalar field space during execution.

6.7.1 Observations

To observe emergent phenomena various graphs and visualisation modules were implemented. The pathway layer and the cell layer are used to visualise biomolecular activities based on network topology. The cell layer represents the orthogonal states of a cell by observing the performance of objectives. The pathway layer observes bio-molecular activities, flux directions, magnitude and bio-molecular population changes in the core of SwarmCell, which is represented by the physical layer.

The observation agent captures values of all required dependent variables of the simulation. First the dependent variables are defined and instructions to capture the values are given. These values will be used to produce the required dynamic graphs of the simulation.

6.8 Model Interface

The SwarmCell interface is designed to conduct simulation experiments. It consists of two main sections, namely the experimental input and output for setting up simulation experiments. The experimental input section sets up experimental scenarios, by providing values for independent and other variables, setting up a user panel and setting up experimental observations for dependent variables. Observations include defining the types of dynamic charts and visualising components for experimental outputs. The experimental output section displays the defined visualising components

for the simulation output. Figures Figure 6.8 and Figure 6.9 show the model interface during execution.

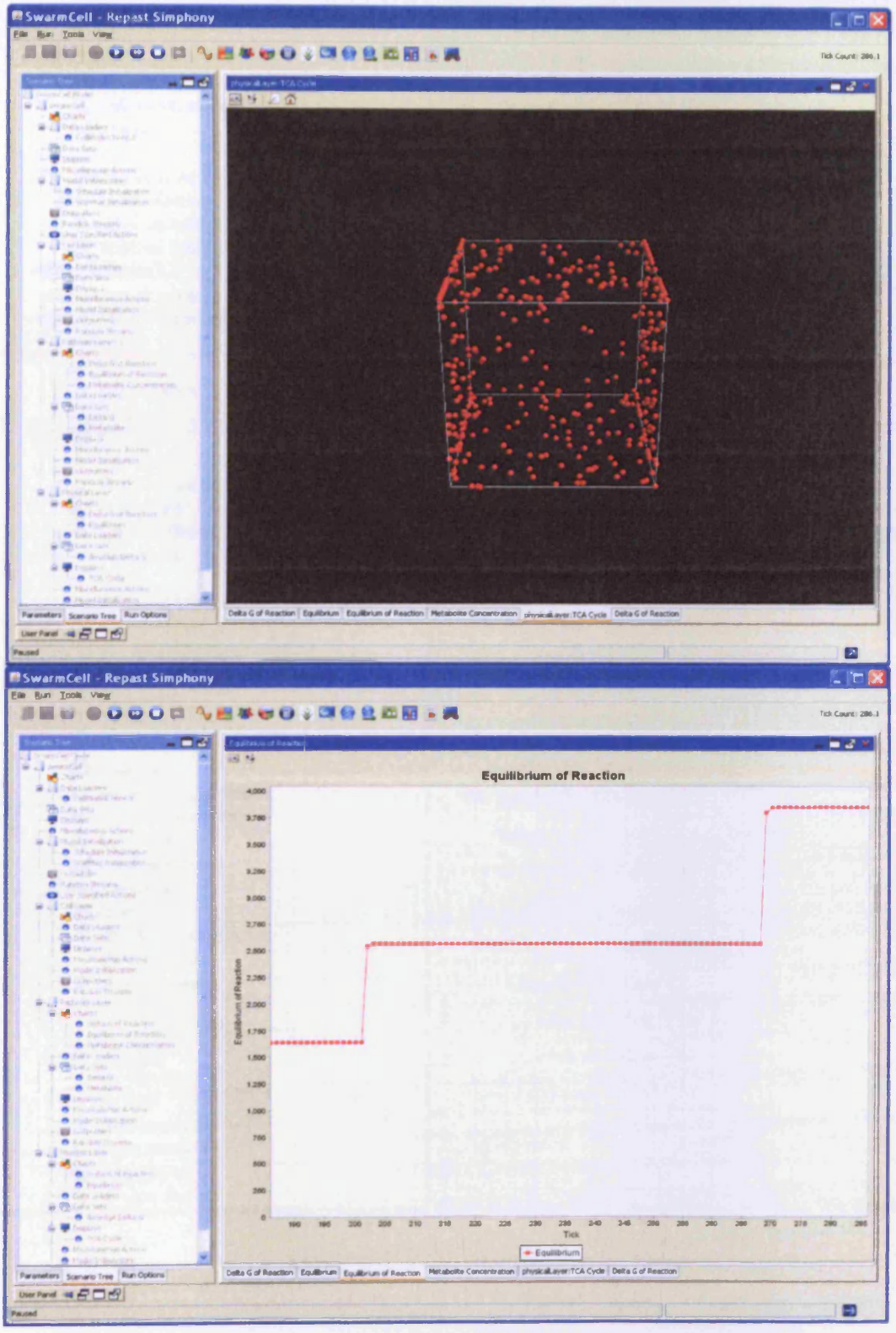


Figure 6.8: SwarmCell Interface.

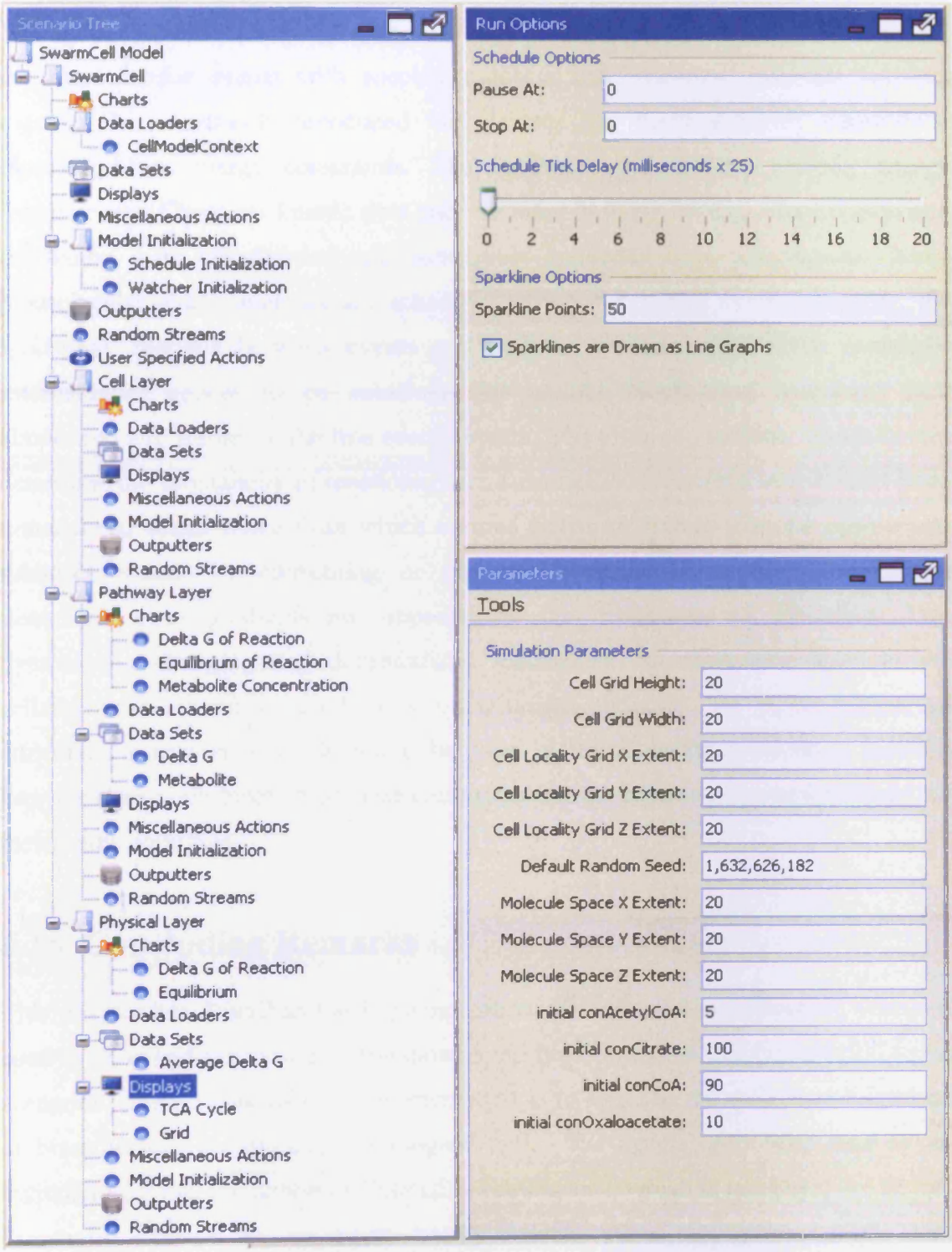


Figure 6.9: Experimental setup.

6.9 Model Execution

The agent based discrete event simulator library, Repast Simphony, is used to model the biomolecular events with specific duration and spacetime intervals between events. The constraints associated with events are thermodynamic constraints, especially free energy constraints. The molecular events have varying energy requirements. Chemical kinetic data such as rate constants for k_{cat}, transcription and translation were transformed into compatible molecular level information. While biomolecular event intervals are scheduled, using a dynamic event scheduler, the spacetime intervals between events are modelled stochastically. Since metabolic intermediates appear to be substrates for various biochemical reactions, their abundance will influence the free energy levels (ΔG) of those reactions, which in turn determine the spontaneity of reactions. Each substrate is represented as a distinct three dimensional scalar space from which various enzymes interact with the appropriate substrate spaces by consuming or producing relevant substrates. This causes disturbances, as gradients are imposed by the dissipation of substrates. This dynamically changing and decentralized mechanism of information flow in the cellular environment is used as a quantitative stigmergy for determination of enzymatic activities and interaction between different enzymes within a module. Enzymes use a combination of these common substrate environments as feedbacks for their future activities.

6.10 Concluding Remarks

This chapter has described the implementation of the model specification, which is used to setup and run various simulation experiments based on Collective Intelligence scenarios. The purpose of the implementation is to simulate the collective behaviour of biomolecules constituting a biological cell. The agents have been used as an ingredient for the simulation of Collective Intelligence, which is facilitated by Repast Simphony (Repast S), an agent based discrete event simulation toolkit. The architecture of the SwarmCell simulation environment facilitates multi-scale modelling from the level of molecular resolution to cellular resolution, which is fundamental for the study of Collective Intelligence phenomena. The ability of the architecture to represent the shared environment is also beneficial as it can model

indirect interactions amongst the biomolecules, which is a feature of Collective Intelligence. Moreover, multi-scale visualization and semantic zooming are two of the important functionalities represented by this architecture, which facilitates the analysis of Collective Intelligence phenomena in biological cells.

Chapter 7

Model Evaluation by Simulating Biological Phenomena

"Nothing in biology makes sense, except in the light of evolution."

- T. G. Dobzhansky

7.1 Overview

The aim of this chapter is to evaluate the Collective Intelligence framework by conducting a series of simulation experiments. Experiments are based on modelling the physical and biological constraints involved in intracellular organisational behaviour, which affect adaptive traits. The biomolecular organisational behaviour is analysed by quantifying cellular functions in terms of measuring performances of objectives/tasks as in Section 7.5. Various scenarios that occur within the mitochondrial environment are simulated. In Section 7.4, an enzyme catalysed reaction of the Tricarboxylic Acid (TCA) cycle is modelled to demonstrate how degeneracy in a biochemical system provides the flexibility to adapt, but recognises the contributions and compensatory adjustments made by different factors in arriving at a solution more complex. Further the thermodynamic requirements and the complex dependencies among metabolites act as constraints on attaining the chemical equilibrium and steady state. In Section 7.5 the multi-objective topology is represented by modelling two competing tasks in metabolic activities to analyse the effects of different factors, such as abundance and efficiency of enzymes on the performances of competing tasks. Based on the results of experiments, the validity of the simulation experiment and the validity of the framework are demonstrated.

7.2 The Verification of Simulation Experiments

The simulation experiments were conducted in a personal computer system with a configuration of,

- Operating system: Windows XP Professional SP3
- Processor: Intel Pentium® D 3.00 GHz
- RAM: 1 GB
- Graphics Processor: Intel® 82945G Express chipset onboard graphics

The General features of the SwarmCell simulation environment were verified during its development phase. The sub-models which consist of biomolecular degradation, biomolecular mobility, biomolecular interaction and biomolecular reproductive errors were implemented and verified before the conduct of simulation experiments. Both static and dynamic testing procedures were used during the implementation of sub-models. For static testing, code reviews, inspection and walkthroughs were utilised. For dynamic testing, test cases were used to verify the sub-models, which consisted of testing with hypothetical input parameters and comparing the output with expected output.

7.3 Simulating Biomolecular Degradation

This simulation tests one of the constraints involved in the self-organisation process specified in Section 5.2.7.3, which describes the degradation process based on average life time and probability. The two approaches are implemented and analysed to compare their accuracy.

7.3.1 Experiment Hypothesis

Biomolecular degradation can be represented at the molecular resolution level by using a probability based approach or average age approach to emulate the elimination of obsolete information in Collective intelligence systems such as a biological cell.

Self-awareness is one of the important properties of Collective intelligence. This is an observable feature of biological cells. Cells have adopted a unique strategy to continuously realise their objectives/tasks or adaptive requirements by eliminating obsolete information and generating new information in their internal organisation. The tendency for biomolecular degradation by means of random or regulated process and collective autocatalysis, provides an ideal reinforcement adaptive mechanism for a cell. The Half-life of functional products is an indication of the duration of their contribution in an intracellular environment. This experiment demonstrates two possible approaches that can be used to emulate the degradation of information, which is one of the contributing factors to self-organisation, occurring in Collective Intelligence. The expected result of this experiment would be to the comply with empirically observed half-lives of functional products. This can be utilised to study other Collective Intelligence phenomena that are a consequence of biomolecular degradation.

7.3.2 Experiment Design

This section describes the independent variables and dependent variables that were used in the simulation. Since the two approaches are to be compared for accuracy, a simple comparative design (Montgomery 2008) was chosen for the experiment. The affect of the two approaches on half-life was tested. For each approach, several simulation runs were conducted. The two approaches are considered, as the design points to model biomolecular degradation and the output of the experiment, the population of biomolecules that varies with time. Table 7.1 shows the design specification of the experiment with two design points representing the two approaches.

Table 7.1: Design specification of the experiment showing the design matrix for simulating biomolecular degradation

Design Points	Approach
Series 1	Average life based
Series 2	Probability based

The experiment consisted of two series and each series consisted of five simulation

runs due to the stochastic nature of the model. This includes the random distribution of age and the probabilistic nature of biomolecular degradation. The termination condition for every simulation run was the degradation of the biomolecular population by almost 100 percent.

7.3.3 Experiment Setup

The sub-model that was consequential for this scenario is biomolecular degradation. This part of the model was verified using the two molecular level approaches for biomolecular degradation. The parameters, shown in Table 7.2 remained constant throughout the experiment.

Table 7.2: The constants and their values used for simulating biomolecular degradation

Other Variables/Parameters	Values
Number of grid cells	1000
Size of the simulation space	2 x 2 x 2 μm^3
Initial population size	1000
Degradation constant	0.0231 $minute^{-1}$
Half-life	30 minutes
Average life time	43.3 minutes
Time Steps	minutes

7.3.4 Experiment Results

The results of the experiment are tabulated in Appendix E (a). For each series, five independent runs were performed and the number of iterations required for the population to halve was recorded from the dynamic line graphs (see Figure 7.1).

Figure 7.1: The results of the half-life experiment

7.3.5 Result Analysis

The purpose of the experiment was to compare the observed simulation results obtained from the average life based approach and the probability based approach, with the experimentally obtained results. The expected outcome was that both the approaches would comply with the expected results, in that the population should halve every 30 minutes. However, while the probability based approach complied with this expectation, the average life based approach significantly deviated from the expected results (see Figure 7.1). As discussed in Section 5.2.7.3, biomolecular degradation is a statistical process, which depends on the instability of the particular biomolecular species. The empirically observed half-life does not imply that every member of the particular biomolecular species will have the same life span, rather half-life represents the average life spans of the respective biomolecular species. In reality the absolute life spans of specific biomolecular species will differ, but will produce an average life span, which will represent the observed half-life. The problem with the average life approach is that the molecules are initialised with random age that cannot exceed the average life of the molecules. This implies that all the

molecules will degrade, when the simulation time reaches the average age of the biomolecules, which is what the results indicate. Hence, the molecules will have to be added at different times during the simulation to have the expected results. However, this will complicate the analysis, because it will be hard to trace the halving of the biomolecular population.

7.3.6 Experiment Validation

Biomolecular degradation, which acts as a negative feedback, is a ubiquitous process in the intracellular environment. Although proteins and transcripts are vulnerable to this process, the transcripts appear to be more stable than the proteins. In proteins, this process depends on ubiquitination. Although, there are no explanations for the distribution of life span, which produces the observed half-life, different species of proteins appear to have distinguishable half-lives based on their amino-terminal residue. Of theses, enzymes tend to have the shortest half-lives, while structural proteins have longer half-lives. The probability based approach has complied with the empirically observed results. However, the rate at which the biomolecules are produced, will have to be integrated with the biomolecular degradation process to build a realistic model. These two processes will produce an oscillatory behaviour, which stabilises the intracellular organisation. This framework can be used to predict the *in vitro* effects of swapping the alpha amino side chains of proteins by analysing the system level behaviour of the cell.

7.3.7 Experiment Conclusion

Biomolecular degradation is modelled at an individual level. This experiment has demonstrated by representing empirically observed half-life at the individual level by using rules, rather than at the population level which utilises rate equations. The result of this experiment complies with empirically observed half-lives of functional products, which can be utilised to study other Collective Intelligence phenomena that are a feature of biomolecular degradation.

7.4 Simulating Degeneracy in a Biochemical System

The aim of this experiment is to demonstrate how degeneracy in a biochemical system provides flexibility to adapt, but adds to the complexity of recognising the contributions and compensatory adjustments made by different factors in arriving at a solution. Further the thermodynamic requirements and complex dependencies amongst metabolites act as constraints on the attainment of a steady state. This is equivalent to chemical equilibrium for single reaction systems. The reaction chosen for this experiment is the first enzyme catalysed reaction in the TCA cycle. The enzyme involved in this reaction is Citrate Synthase. The reaction is:

$$AcetylCoA + oxaloacetate + H_2O \xrightarrow{Citrate\ Synthase} citrate + CoA + H^+$$

Every metabolite has a standard Gibbs free energy of formation, listed in Table 7.3. These values are used to calculate the standard Gibbs free energy of the reaction.

Table 7.3: Gibbs free energy of formation of metabolites involved in the reaction 1 of the TCA cycle

Metabolites	$\Delta_f G^{\circ}$ of metabolites
Oxaloacetate	-713.38 kJ mol^{-1}
Acetyl CoA	-60.49 kJ mol^{-1}
Citrate	-963.46 kJ mol^{-1}
CoA	-7.98 kJ mol^{-1}
H_2O	-157.28 kJ mol^{-1}
H^+	0.0 kJ mol^{-1}

The $\Delta_r G^{\circ}$ for this reaction is -7.5 kcal mol^{-1}/-31.4 kJ mol^{-1} where the equation used to calculate the local environment's $\Delta_r G$ is:

$$\Delta G = \Delta G^{\circ} - 2.303\ RT \log_{10} \frac{[Citrate]\ [CoA]\ [H^+]}{[Acetyl\ CoA]\ [Oxaloacetate]}$$

The enzyme is constrained by the thermodynamic property, ΔG, in the local environment. This behaviour is modelled in swarm agents, where the agents, sense the ΔG value in the local environment for a particular reaction and behave accordingly. The cellular environment is split into smaller local environments to emulate and

represent localised populations of metabolite, as opposed to global populations. The metabolites are modelled as part of the environment, where the numbers fluctuate in the local environment due to reactions and diffusion. These fluctuations will cause localised ΔG values to emerge, which will bias the behaviour of enzymes modelled as agents. Measuring the performance of functional products, such as enzymes is crucial to modelling the intracellular organisational behaviour. The performance of enzymes is a collective property, which depends on two factors, namely the efficiency and abundance of the activity. These are compromised by interacting. The efficiency of the enzyme is amplified via the redundant counterparts, and this affects the overall performance of the intended task of the particular enzyme species.

7.4.1 Experiment Hypothesis

The performance of a particular enzyme species can be represented by the enzyme's efficiency and abundance of enzyme activity

The goal of reaching equilibrium is studied against factors, such as enzyme efficiency and abundance, which are perturbed in the experiment. The expected outcome of the study is a demonstration of the impact of perturbation (change in factor levels) on the goal (time required to attain equilibrium), the sensitivity of particular perturbations on attaining the equilibrium and the identification of interaction effects between the factors. It is expected that there will be no interactions between factors, since there are no influences between the represented factors.

7.4.2 Experiment Design

The independent, dependent and other variables used in the simulation are described here. Since there are many factors, which affect the attainment of chemical equilibrium, a 2^k Factorial design was chosen (Montgomery 2008), where k corresponds to the number of factors (independent variables). The main effect of a factor is defined to be the change in response produced by a change in the level of the factor. To understand how each of the factors affects a response, two levels per factor were chosen. For each of the 2^k factor level combinations, several simulation runs were conducted. The iterations or time steps for a simulation run depend on the time taken to attain chemical equilibrium. Table 7.4 lists the factors represented in the

experiments. The abundance of enzymes represents the effective abundance and the value ranges were estimated based on the typical amounts found in a cell, which are of the order of nM. The values for the turnover cycle/processing time of wild type and mutant enzymes were obtained from the enzyme database BRENDA (Chang A., Scheer M. et al. 2009). Higher values for metabolites were chosen relative to the enzymes, due to their higher abundance. These are of the orders of nM - μM (Nazaret, Heiske et al. 2009). The relationship between biomolecular concentration and the molecules present in specific volumes are tabulated in Appendix F.

Table 7.4: Variables and their value ranges used for simulating the biochemical reaction

Variables	Values	
	Low	High
Independent Variables		
The abundance of enzyme Citrate Synthase(EA)	200	400
The turnover cycle of enzyme(ET)	66.7 ms	120.5 ms
The initial phase/state of the chemical system(EP)	-0.756 kcal mol^{-1}	-0.490 kcal mol^{-1}
Abundance of Acetyl CoA	6,000	5,000
Abundance of Citrate	180,000	200,000
Abundance of Oxaloacetate	6,000	5,000
Abundance of CoA	160,000	180,000
Dependent Variables		
Phase/State of reaction equilibrium		
ΔG of reaction		
Abundance of metabolites		
Number of iterations to reach equilibrium		

The factors that affect the time required to attain chemical equilibrium, are the abundance of the enzyme, the enzyme's processing time/turnover cycle, and the free energy phase of the chemical system. The free energy phase of the chemical system is its equilibrium phase/state. It is a phase, describing how far it is from chemical equilibrium. Since $\Delta G = 0$ at chemical equilibrium, the more negative or positive ΔG is, the further away it is from attaining chemical equilibrium. The factors, that affect

the initial ΔG phase of the chemical system, are abundance of participating reactants and products. Table 7.5 shows the design specification for the experiment with eight design points representing the three independent factors.

Table 7.5: Design specification of the experiment showing the design matrix for simulating biochemical reaction

Design Points		Factors		
		Enzyme abundance (EA)	Equilibrium phase (EP)	Turnover cycle (ET)
Series 1	(+,+,+)	High	High	High
Series 2	(+,+,-)	High	High	Low
Series 3	(+,-,+)	High	Low	High
Series 4	(+,-,-)	High	Low	Low
Series 5	(-,+,+)	Low	High	High
Series 6	(-,+,-)	Low	High	Low
Series 7	(-,-,+)	Low	Low	High
Series 8	(-,-,-)	Low	Low	Low

The experiment consisted of eight series to investigate all possible combinations of the factor levels. Each series consisted of five simulation runs due to the stochastic nature of the model, such as the random movements of the biomolecules and the asynchronous nature of biomolecular activities. The termination condition for every simulation run was the phase chemical system reaching equilibrium.

7.4.3 Experiment Setup

The sub-model that was consequential for this scenario is the biomolecular mobility and biomolecular interaction. The scenario of the experiment was verified using test cases, which consisted of hypothetical input parameters for enzyme abundance, turnover cycle and the initial phase of the chemical system, with expected output for the time required to reach chemical equilibrium. Table 7.6 shows the constant parameter values used throughout this experiment.

Table 7.6: Variables and their value ranges used for simulating biochemical reaction

Other Variables/Parameters	Values
Number of grid cells	1000
Size of the simulation space	2 x 2 x 2 μm^3
Diffusion constant of the enzyme	5 $\mu m^2 s^{-1}$
Diffusion constant of the metabolites	66 $\mu m^2 s^{-1}$
Time step	millisecond

7.4.4 Experiment Results

The results of the experiment are given in Appendix E (b). The number of iterations required to attain equilibrium was recorded from the dynamic line graphs. Each iteration represents a millisecond. The distribution of the responses is shown in Figure 7.2 and the global equilibrium constants attained by the different combinations of factor levels are shown in Figure 7.3.

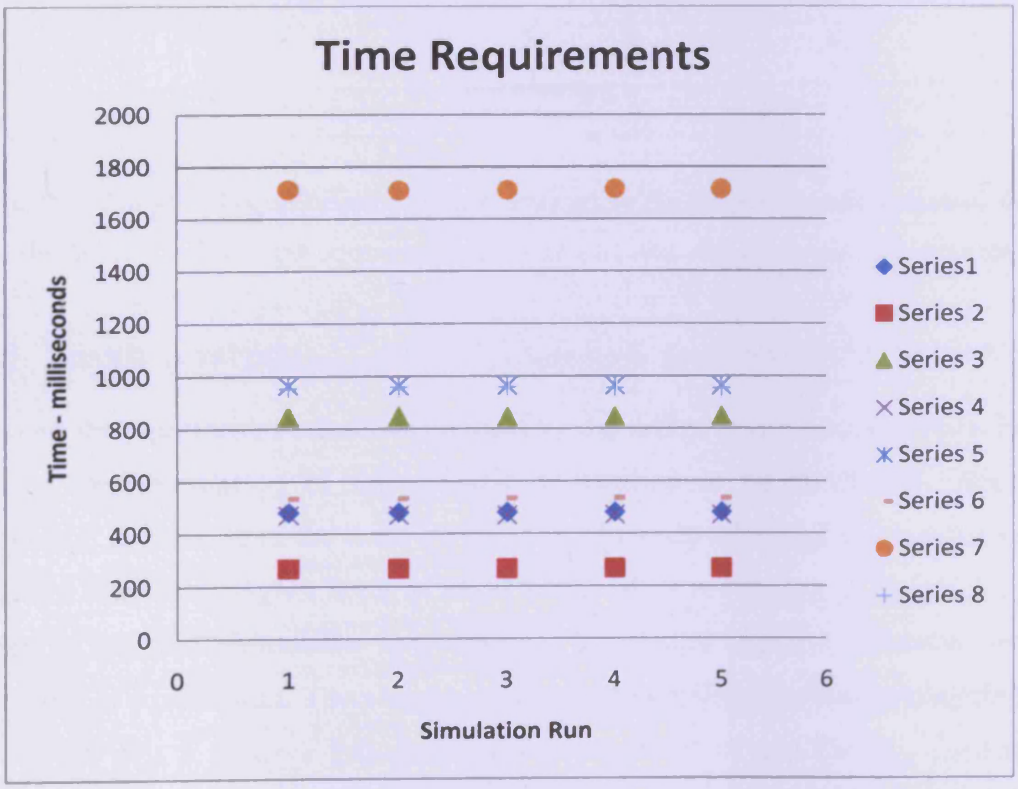


Figure 7.2: Time require to attain equilibrium with respect to enzyme abundance, processing time and the initial equilibrium phase of the chemical system. Table 7.5 shows the corresponding series numbers with the levels of factors used.

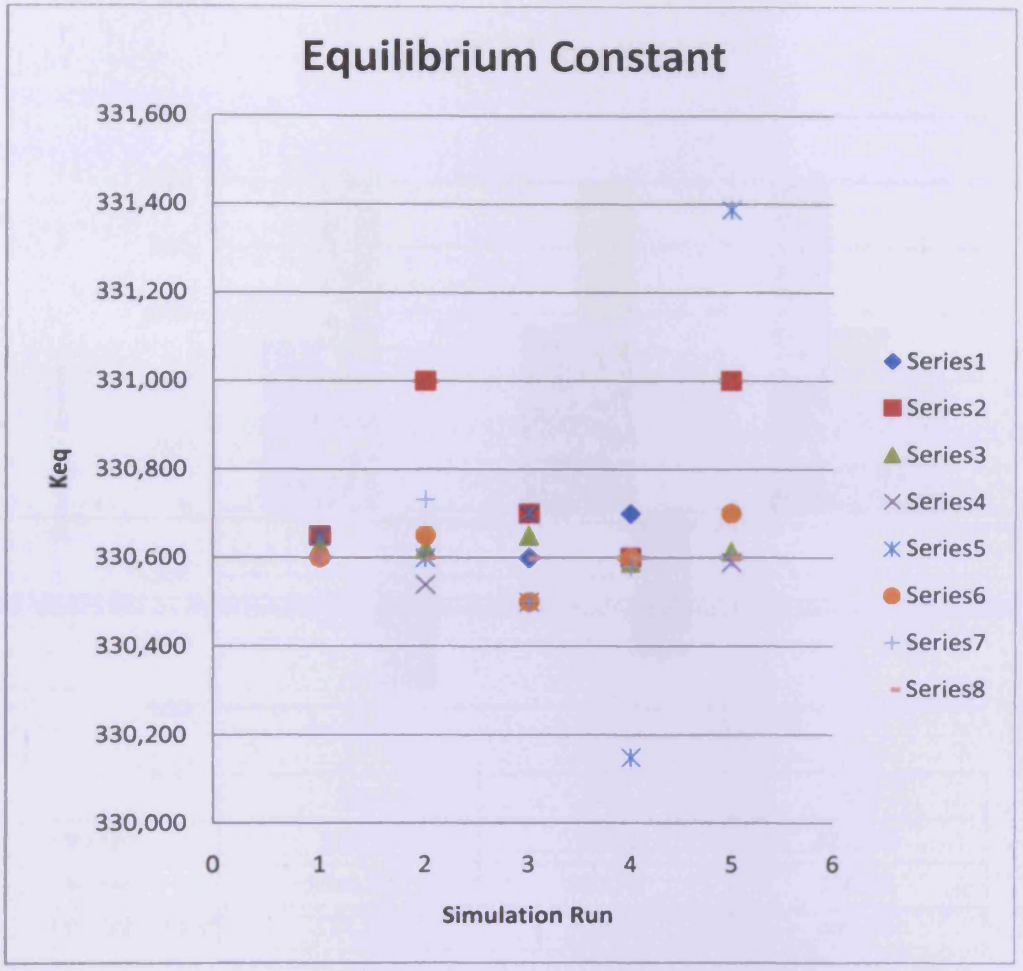
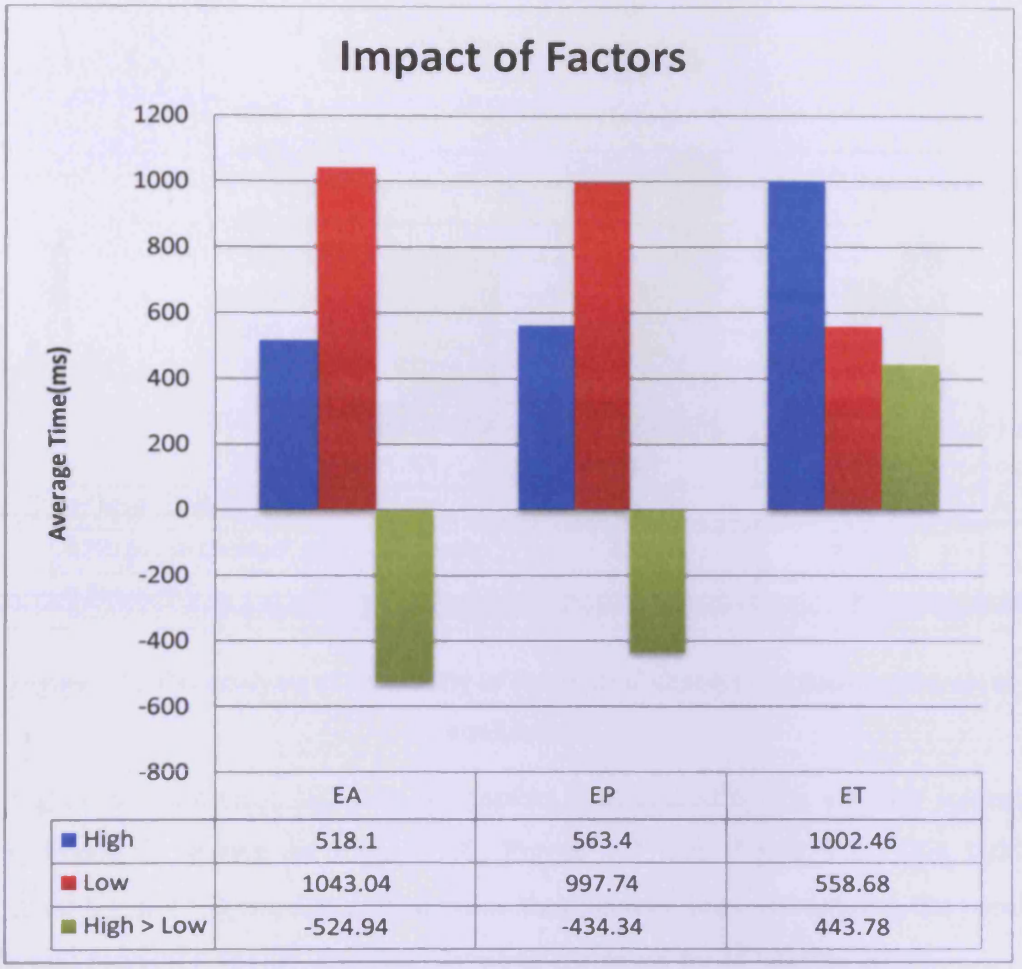


Figure 7.3: The global equilibrium constants attained by the different combinations of factor levels. Table 7.5 shows the corresponding series numbers with the levels of factors used.

7.4.5 Result Analysis

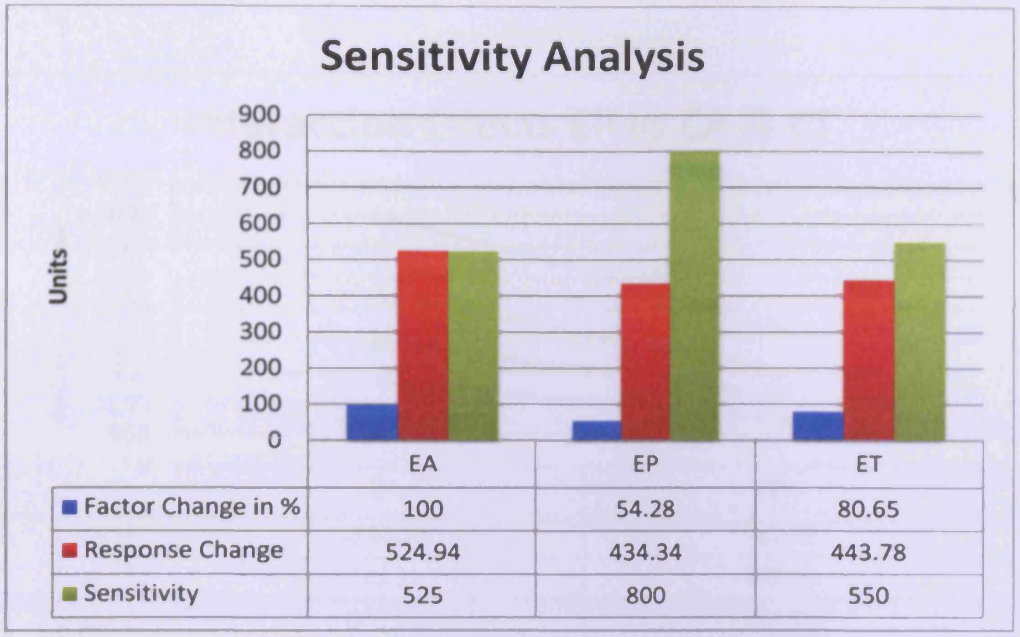
Based on the equilibrium constants attained by the different series as shown in Figure 7.3, the implementation of the scenario is verified to be functional, since the equilibrium constant is in the order of 330,000. The main effects of moving the values of factors from their higher level to lower level values are shown in Figure 7.4. The change in enzyme abundance from high to low has a negative influence on the attainment of equilibrium. The change in enzyme processing time/turnover cycle from high to low has a positive influence on the attainment of equilibrium. Further the change in the initial equilibrium phase of the chemical system from high to low has a negative influence on the attainment of equilibrium.



	EA	EP	ET
High	518.1	563.4	1002.46
Low	1043.04	997.74	558.68
High > Low	-524.94	-434.34	443.78

Figure 7.4: The effects of change in factor levels with respect to response time

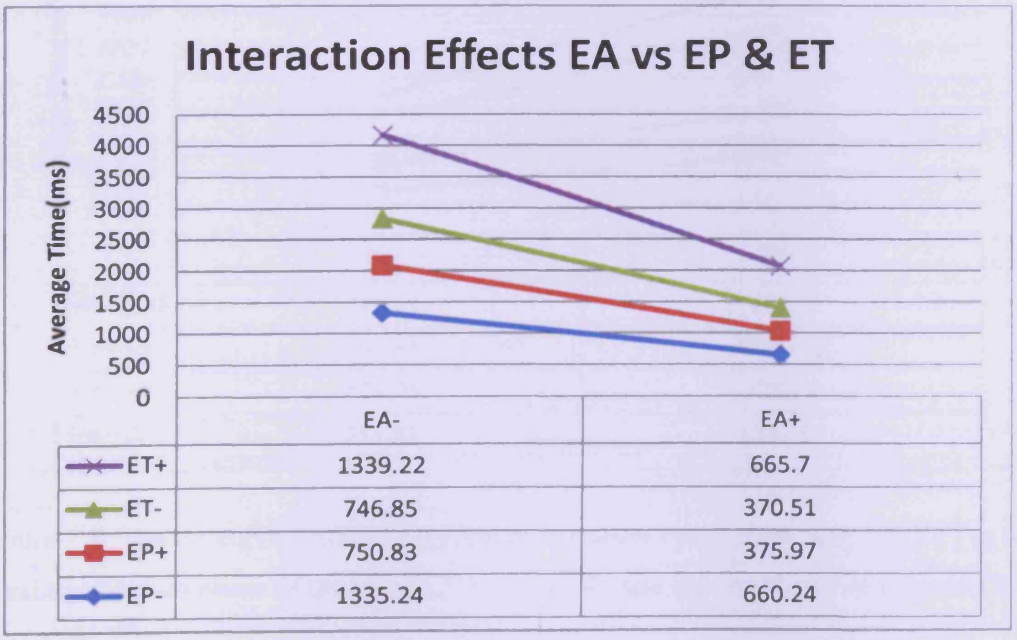
Sensitivity analysis is performed to analyse the sensitivity of changes in factor levels to specific responses, which can indicate the robustness of responses to the changes in levels of specific factors. This can help to formulate the limits between perturbation and performance, which define the boundaries of robustness. To analyse the sensitivity of response variables with respect to the changes in the levels of factors or independent variables, the magnitude between the levels of factors were compared to the magnitude of the change in responses. Figure 7.5 shows the sensitiveness of factor level changes, in reaching chemical equilibrium.



	EA	EP	ET
Factor Change in %	100	54.28	80.65
Response Change	524.94	434.34	443.78
Sensitivity	525	800	550

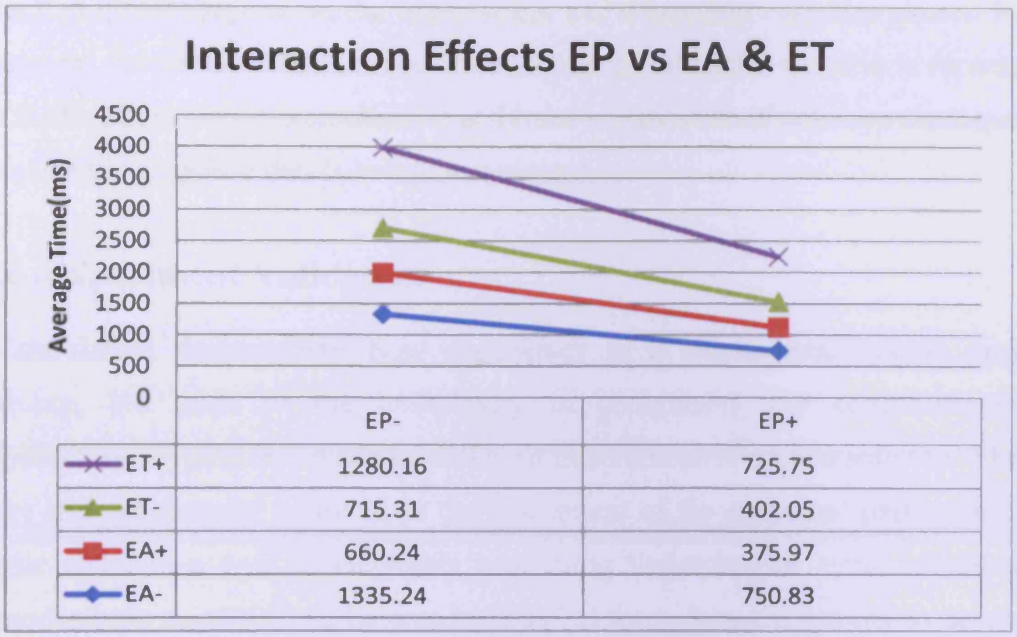
Figure 7.5: The analysis of sensitivity of factor level changes, in reaching chemical equilibrium

The degree of interaction between the factors is measured by the k-factor interaction effect, which is shown in Figure 7.6, Figure 7.7 and Figure 7.8. This value is calculated by multiplying the design point sign vectors, then multiplying the resulting vector and response vector, and then dividing the result by 2^{k-1}(Schut 2007).



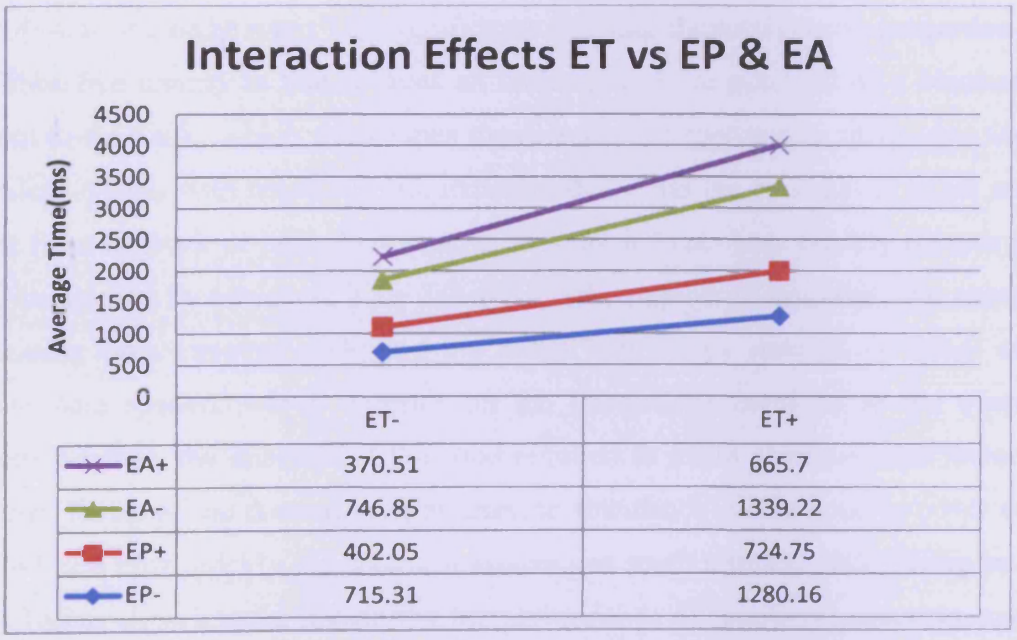
	EA-	EA+
ET+	1339.22	665.7
ET-	746.85	370.51
EP+	750.83	375.97
EP-	1335.24	660.24

Figure 7.6: The interaction effects between the abundance of enzyme (EA) vs turnover cycle of the enzyme (ET) and the initial equilibrium phase of the chemical system (EP).



	EP-	EP+
ET+	1280.16	725.75
ET-	715.31	402.05
EA+	660.24	375.97
EA-	1335.24	750.83

Figure 7.7: The interaction effects between the initial equilibrium phase of the chemical system. (EP) vs the abundance of enzyme (EA) and turnover cycle of the enzyme (ET).



	ET-	ET+
EA+	370.51	665.7
EA-	746.85	1339.22
EP+	402.05	724.75
EP-	715.31	1280.16

Figure 7.8: The interaction effects between the turnover cycle of the enzyme (ET) vs the initial equilibrium phase of the chemical system (EP) and the abundance of enzyme (EA).

Based on the analysis of the interaction between factors, there are no noticeable interactions between the factors since the difference in responses between the levels

of one factor is the same at all levels (i.e. almost parallel) to the other factors. The interaction effects depend on the independent and dependent variables chosen for the experiment. Interaction between factors will emerge when the scenario is represented with feedbacks amongst biomolecular activities and these feedbacks are consequential to the specific response that is being investigated.

7.4.6 Experiment Validation

The simulation demonstrates how degeneracy in a biochemical system provides flexibility, but adds to the complexity in recognising the contributions and compensatory adjustments made by different factors in arriving at a solution. The two factors that are directly involved in the adjustment of the enzymes' performance are enzyme abundance and an enzyme's processing time/turnover cycle. Further the thermodynamic requirements have to be satisfied for a chemical activity to occur, and the complex dependencies among metabolites will determine the reactants and the free energy phase of the chemical system. The free energy phases of the chemical system will determine how far the chemical system is from attaining chemical equilibrium or steady state. The significance of using thermodynamic properties such as Gibbs free energy is that it gives an indication of the potential of a biochemical system to do work, which determines the direction of spontaneity of the reaction. A chemical system with negative, ΔG, indicates that it has the potential to expel energy in the form of work or heat. However, it does not indicate how quickly (kinetics) the equilibrium can be achieved. This depends on the number of enzymes, the enzyme's processing time/turnover cycle and the initial free energy state (Δ G value) of the biochemical system, which depends on the metabolites involved in the chemical system. Further, the analysis of the time required to attain chemical equilibrium by different levels of the factors, such as enzyme abundance and enzyme turnover cycle, can indicate how quickly the chemical system can reach a steady state during *in vitro* perturbation experiments. Measuring the performance of functional products, such as enzymes is crucial in modelling the intracellular organisations behaviour. The complex dependencies on metabolites of various reactions of the TCA cycle are shown in Appendix B. This indicates the complexity involved in associating the performances of the enzymes, based on measuring the abundance of the metabolites. The metabolites are produced and consumed by various enzymatic reactions in the

cell, which makes it impossible to trace the performances of enzymes. The abundance of metabolites reflects on the net abundance (not gross abundance), which emerges from concurrent activities of different enzymes producing or consuming the particular metabolite. The ideal strategy is to measure the activity of enzymes, which is related to the performance of the enzyme. Further the complex dependencies also indicate the thermodynamic constraints involved in the reactions, which is ignored in chemical kinetics experiments. This correlation effect between the independent variables and the dependent variable obtained for this scenario can be compared with similar *in vitro* experiment, to ascertain the validity of the model's mechanistic structure.

To define the capacity of robustness with respect to perturbations and performances of biochemical tasks, the analysis of sensitivity of change in the levels of specific perturbation to the specific intracellular performance will be critical. It can be used to define the limits of specific perturbation, which can maintain a desired performance. The relationship between perturbation and performance is dependent on the intercellular organisational behaviour, which is regulated by numerous feedbacks amidst uncertainty.

7.4.7 Experiment Conclusion

The Collective Intelligence framework experiment has demonstrated that both qualitative and quantitative factors can compensate each other to meet a performance. The ability to measure the activities of functional products and relate them to the performance, and consequently to the intracellular functionalities is the distinctive feature of this framework. Also it has shown that the ability to represent spatial, temporal and thermodynamic constraints within the framework contributes to being able to make a more realistic representation of intracellular dynamics.

7.5 Simulating Competition in Metabolic Activities

The aim of this experiment is to use the multi-objective topology to model two competing tasks in metabolic activities and analyse the effects of different factors, such as abundance and efficiency of enzymes on the performances of the competing tasks. Measuring the performance of functional products, such as enzymes, is crucial in modelling the contributions they make as biomolecular species to the

organisational behaviour of the cell and the genes, which they represent. The performance of enzymes is a collective property, which depends on factors, such as the efficiency and abundance of activity, which are compensated by each other. The efficiency of the enzyme is amplified by their redundant counterparts, which affect the overall performance of the intended task of respective enzyme species. The pathway modelled is the biosynthesis of aromatic amino acids in *E.coli*, see Figure 7.9. This pathway diverges at chorismate into a prephenate and anthranilate branch. The problem space is represented by metabolites, which comprise chorismate, prephenate, anthranilate, L-glutamine, pyruvate and L-glutamate. The function space is represented by two groups of enzymes forming the two competing tasks, producing the metabolites prephenate and anthranilate. The critical point is that the performance of metabolic tasks are governed by the activities of key enzymes rather than by the law of mass action (Stryer 1988). The enzyme involved in the conversion of chorismate to prephenate is Chorismate Mutase. The wild type of this enzyme has a turnover number about $39s^{-1}$ and the $\Delta_rG°$ for this reaction is $-56kJmol^{-1}$(Kast, Tewari et al. 1997). The enzyme involved in the conversion of chorismate to anthranilate is Anthrinilate Synthase. The wild type of this enzyme has a turnover number around $383s^{-1}$ and the $\Delta_rG°$ for this reaction is $-183kJmol^{-1}$ (Byrnes, Goldberg et al. 2000).

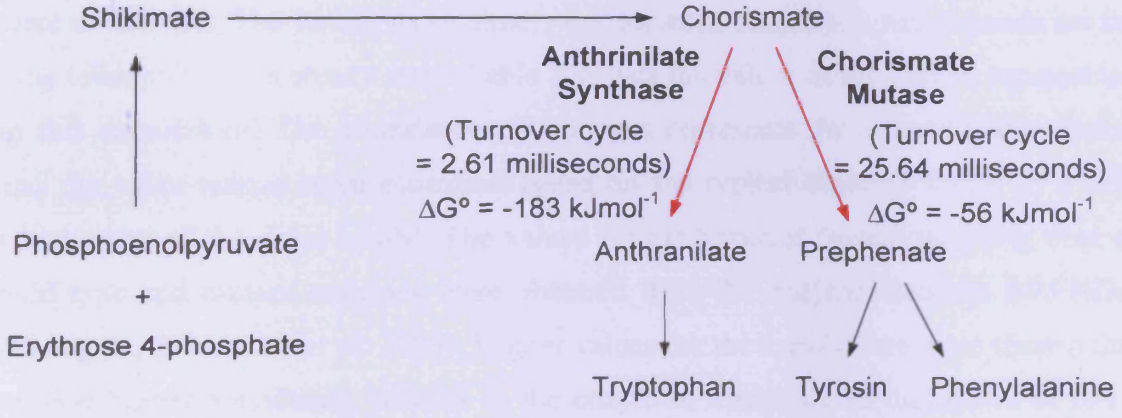


Figure 7.9: Pathway for the biosynthesis of aromatic amino acids in *E. coli*

7.5.1 Experiment Hypothesis

Competing enzyme species can have an inhibitory effect on their performances.

The overall performance of a particular enzyme species is analysed against the performance of a competing enzyme species. The competition for a common

metabolite is analysed using factors such as enzyme efficiency and abundance, which are perturbed in the experiment.

The expected outcome of the study is to demonstrate the impact of perturbation (change in factor levels) on the overall performance of a particular enzyme species. The sensitivity of a particular perturbation on the performance of a particular enzyme species and the identification of interaction effects between factors. It is expected that there will be no interactions between factors, since the scenario does not include any feedbacks between the competing enzyme species.

7.5.2 Experiment Design

The independent, dependent and other variables used in this simulation experiment are described here. Since there are many factors, which affect the performances of the group's tasks a 2^{k-p} Fractional factorial design (Montgomery 2008) was chosen, where k corresponds to the number of factors (i.e. independent variables) and p corresponds to the excluded factors, such as the regulation of the enzymes. The main effect of a factor is the change in response produced by a change in the level of the factor (Schut 2007). To understand how each factor affects the responses, two levels per factor were chosen. For each of the 2^{k-p} factor level combinations, several simulation runs were conducted. The iterations or time steps for each simulation run depends on the time taken to attain a steady state. Table 7.7 lists the values of the factors represented in this experiment. The abundance of enzymes represents the effective abundance, and the value ranges were estimated based on the typical amounts found in a cell, which were of the order of nM. The values for the turnover cycle/processing time of wild type and mutant enzymes were obtained from the enzyme database BRENDA (Chang A., Scheer M. et al. 2009). Higher values for the metabolites were chosen due to their higher abundance, relative to the enzymes. These are of the orders of nM - μM (Nazaret, Heiske et al. 2009). The relationship between biomolecular concentration and the molecules present in specific volumes are tabulated in Appendix F.

Table 7.7: The variables and their value ranges used for simulation competition and cooperation in metabolic pathway

Variables	Values	
	Low	High
Independent Variables		
The abundance of enzyme Chorismate Mutase(CMA)	100	200
The abundance of enzyme Anthrinilate Synthase(ASA)	100	200
The turnover cycle of Chorismate Mutase(CMT)	19.7ms	25.6 ms
The turnover cycle of Anthrinilate Synthase(AST)	2.6 ms	164.5 ms
Dependent Variables		
State of reaction equilibrium		
ΔG of reactions 1 and 2		
Abundance of metabolites		
Number of iterations to reach the end point		
Performances of the groups		

Although the relative values are more than sufficient to model the principles governing this experiment, absolute values obtained from experimentation were chosen as a realistic representation of the experiment. The factors that affect an enzyme's performance in performing the competing tasks in metabolic activities, are abundance of enzyme activities and the enzyme's turnover cycle. Although the turnover cycles are unique to every enzyme species, the abundance of their activities are controlled by various regulatory mechanisms, involved in the production, activation, deactivation and degradation of enzymes. The initial free energy state of chemical system remains constant throughout the experiment. The factors that affect the initial free energy state (ΔG) of the chemical system are abundance of participating reactants and products. Table 7.8 shows the design specification of the experiment with sixteen design points, representing four independent factors.

Table 7.8: Design specification of the experiment showing the design matrix for simulating competition and cooperation in metabolic pathway

Design Points		Factors			
		Enzyme abundance (CMA)	Enzyme abundance (ASA)	Turnover cycle (CMT)	Turnover cycle (AST)
Series 1	(+,+,+,+)	High	High	High	High
Series 2	(+,+,+,-)	High	High	High	Low
Series 3	(+,+,-,+)	High	High	Low	High
Series 4	(+,+,-,-)	High	High	Low	Low
Series 5	(+,-,+,+)	High	Low	High	High
Series 6	(+,-,+,-)	High	Low	High	Low
Series 7	(+,-,-,+)	High	Low	Low	High
Series 8	(+,-,-,-)	High	Low	Low	Low
Series 9	(-,+,+,+)	Low	High	High	High
Series 10	(-,+,+,-)	Low	High	High	Low
Series 11	(-,+,-,+)	Low	High	Low	High
Series 12	(-,+,-,-)	Low	High	Low	Low
Series 13	(-,-,+,+)	Low	Low	High	High
Series 14	(-,-,+,-)	Low	Low	High	Low
Series 15	(-,-,-,+)	Low	Low	Low	High
Series 16	(-,-,-,-)	Low	Low	Low	Low

The experiment consisted of sixteen series, to allow investigation of all possible combinations of the factor levels. Each series consisted of three simulation runs due to the stochastic nature of the model, such as the random movements of the biomolecules and the asynchronous nature of biomolecular activities. The termination condition for every simulation run was the complete consumption of Chorismate.

7.5.3 Experiment Setup

The sub-model that was consequential for this scenario is the biomolecular mobility and biomolecular interaction. The scenario of the experiment was verified using test

cases, which consisted of hypothetical input parameters for enzyme abundance and enzyme turnover cycle, with expected output for the performances of enzymes. Table 7.9 shows the constant parameter values used throughout this experiment.

Table 7.9: Variables and their value ranges used for simulation competition and cooperation in metabolic pathway

Other Variables/Parameters	Values
Number of grid cells	1000
Size of the simulation space	2 x 2 x 2 μm^3
Diffusion constant of the enzymes	5 $\mu m^2 s^{-1}$
Diffusion constant of the metabolites	66 $\mu m^2 s^{-1}$
$\Delta G°$ for reaction 1	-56KJ mol^{-1}
$\Delta G°$ for reaction 2	-183KJ mol^{-1}
Time step	milliseconds
The initial phase of the chemical systems	
The abundance of Anthranilate	65,000
The abundance of Pyruvate	50,000
The abundance of Lglutamine	70,000
The abundance of Lglutamate	55,000
The abundance of Prephenate	60,000
The abundance of Chorismate	30,000

7.5.4 Experiment Results

The results of the experiment are given in Appendix E (c). For each series, three independent runs were performed and the performances of the enzymes and the final levels of the metabolites were recorded from the dynamic line graphs. The distribution of the responses is shown in Figure 7.10. Figure 7.11 shows the consumption of Chorismate by the two competing tasks, involving Chorismate Mutase and Anthrinilate Synthase.

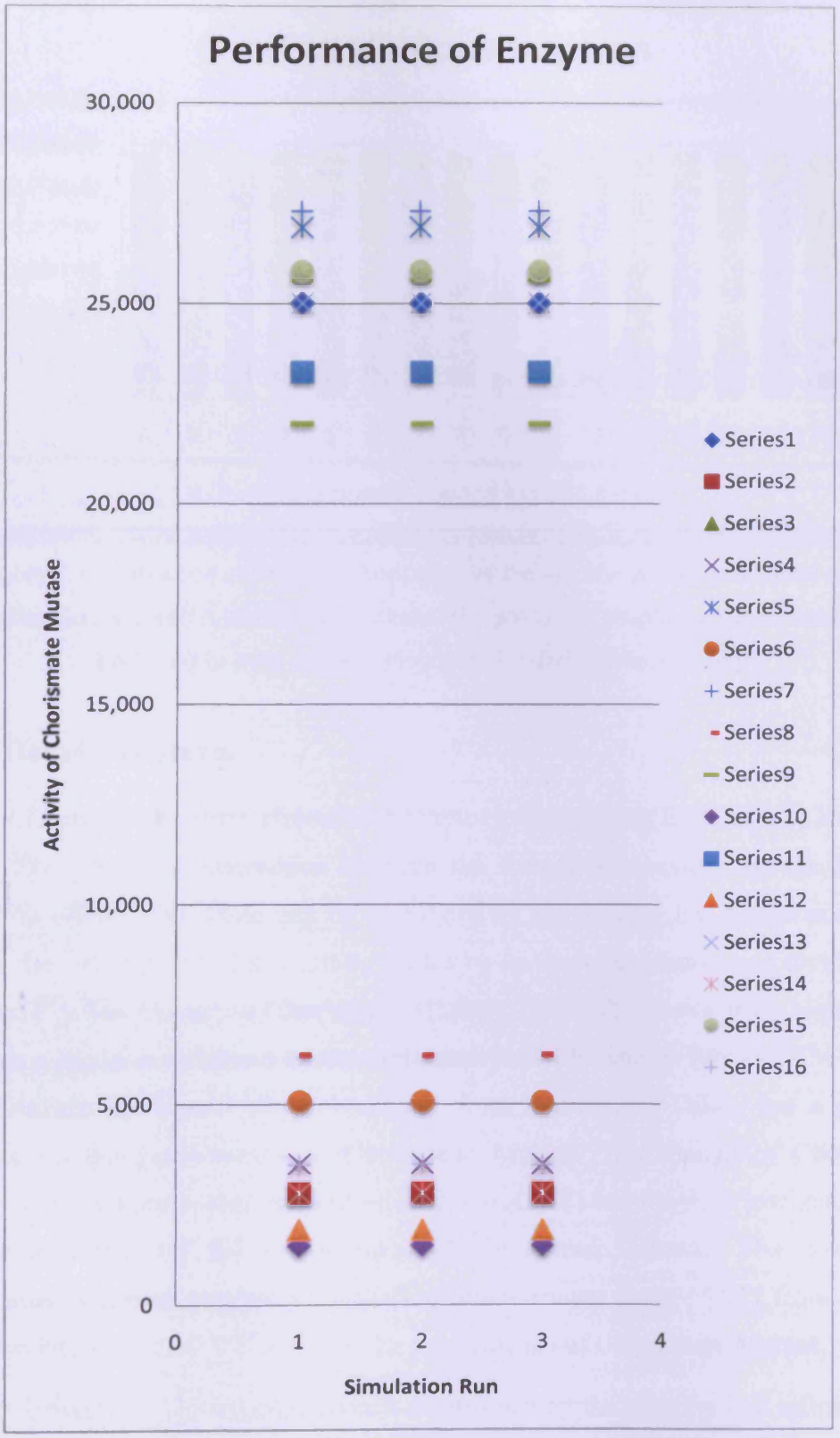


Figure 7.10: The performance of Chorismate Mutase with respect to abundance and processing time of enzyme Chorismate Mutase and Anthrinilate Synthase. Table 7.8 shows the corresponding series numbers with the levels of factors used.

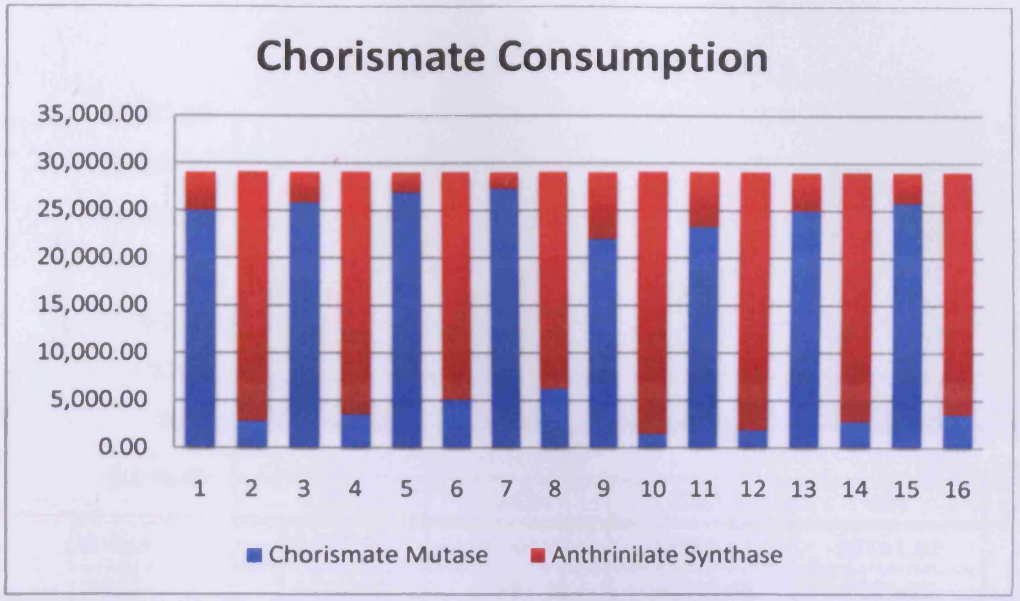
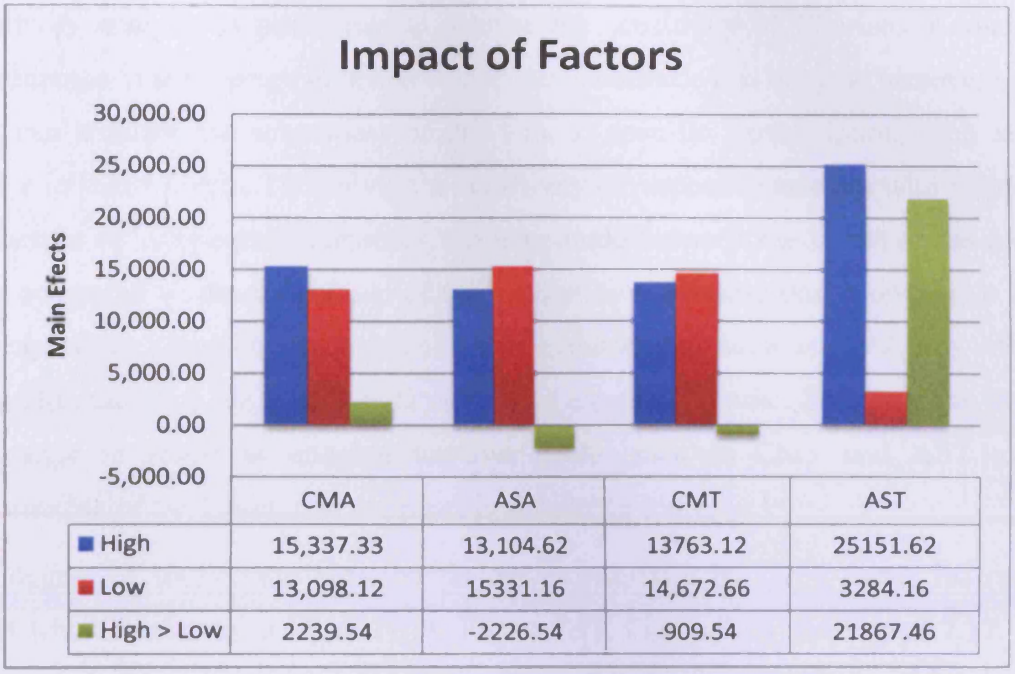


Figure 7.11: The consumption of Chorismate by the two competing tasks involving Chorismate Mutase and Anthrinilate Synthase. The amount of Prephenate and Anthranilate produced in each series is shown in blue and red respectively

7.5.5 Result Analysis

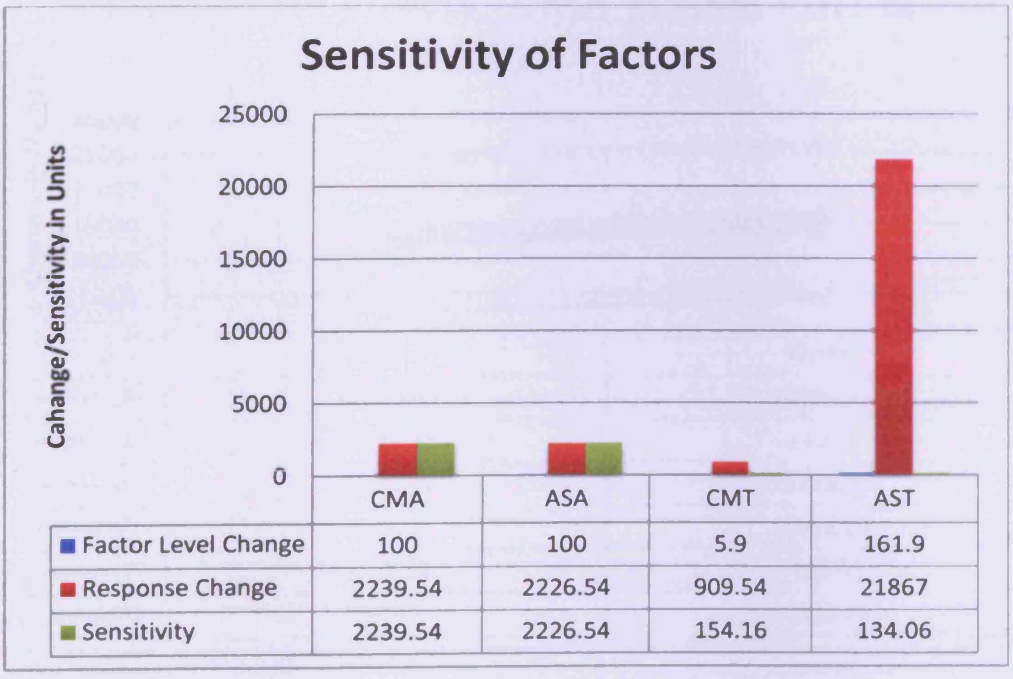
Figure 7.12 shows the main effects of the factor value moving from high to low level values. The degree of interaction between the factors is measured by the k-factor interaction effect. This value can be calculated by multiplying the design point sign vectors, then multiplying the resulting vector by its response vector, and dividing the result by 2^{k-1}. The change in Chorismate Mutase (CMA) abundance from high to low value has a negative influence on the performance of Chorismate Mutase. The change in Anthrinilate Synthase (ASA) abundance from high to low value, has a positive influence on the performance of Chorismate Mutase. The change in Chorismate Mutase enzyme's processing time/turnover cycle (CMT) from high to low value has a positive influence on the performance of Chorismate Mutase. The change in Anthrinilate Synthase enzyme's processing time/turnover cycle (AST) from high to low value has a negative influence on the performance of Chorismate Mutase.

The flux between these two directions is determined by the quantity and efficiency of the enzyme. While enzyme efficiency is a direct reflection of the quality of the coding sequence, its level of activity is reflected in its regulation. Although the cells have no control over the enzyme's efficiency, it has evolved the ability to gain control over its activity levels by regulating it quantitatively.



	CMA	ASA	CMT	AST
High	15,337.33	13,104.62	13763.12	25151.62
Low	13,098.12	15331.16	14,672.66	3284.16
High > Low	2239.54	-2226.54	-909.54	21867.46

Figure 7.12: The effects of change in factor levels with respect to the performance of Chorismate Mutase

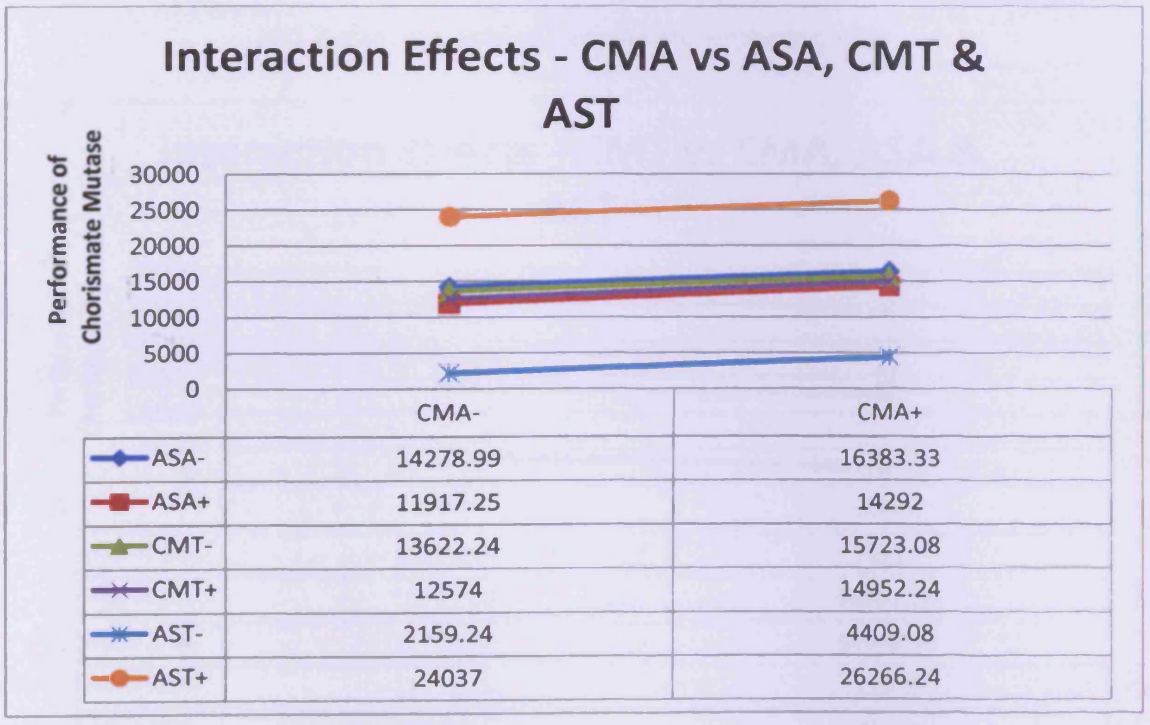


	CMA	ASA	CMT	AST
Factor Level Change	100	100	5.9	161.9
Response Change	2239.54	2226.54	909.54	21867
Sensitivity	2239.54	2226.54	154.16	134.06

Figure 7.13: The analysis of sensitivity of the performance of Chorismate mutase to the above factors

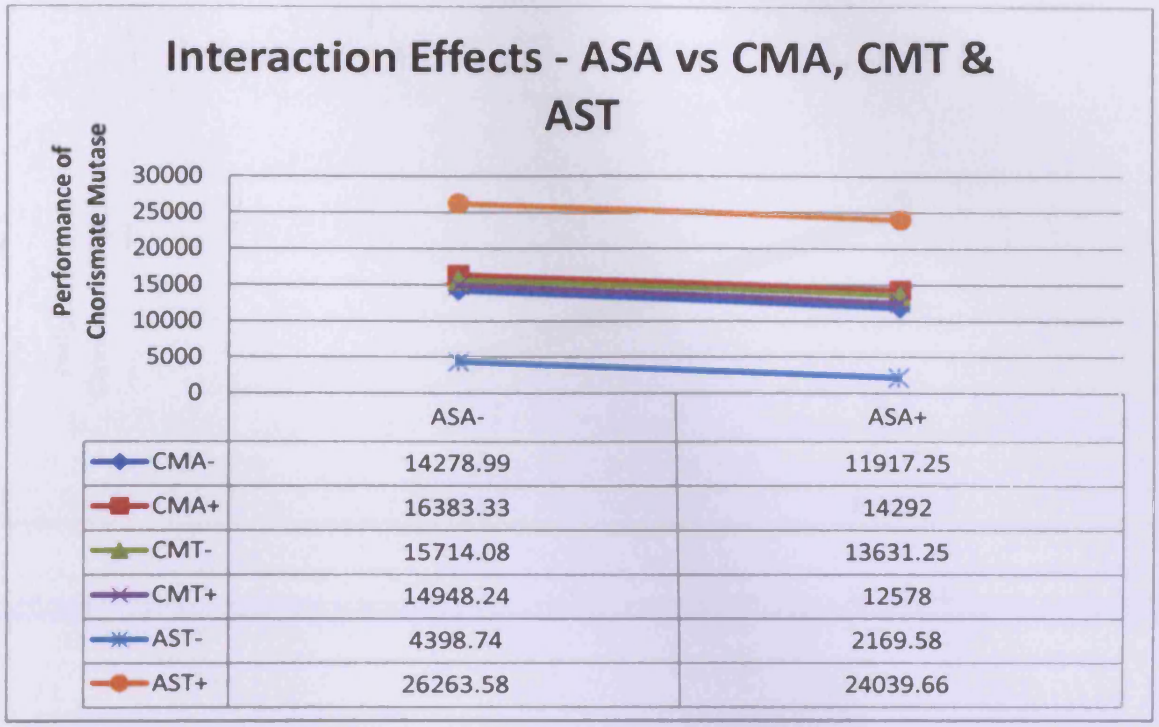
Sensitivity analysis is performed to analyse the sensitivity of Chorismate mutase's performance to the change in levels of enzyme abundance and enzyme turnover cycle. This can indicate the robustness of this task to specific perturbations, such as the change in factor levels. To analyse the sensitivity of response variables with respect to the factors or independent variables, the magnitude between the levels of the factors were compared to the magnitude of the change in responses. Based on Figure 7.13, the impact of change in the levels of enzyme abundance, such as CMA and ASA to the performance of the Chorismate mutase, is almost the same. Similarly the impact of change in levels of enzyme turnover cycle, such as CMT and AST to the performance of the Chorismate mutase, is almost the same.

The degree of interaction between factors is measured by the k-factor interaction effect which is shown in Figure 7.14, Figure 7.15, Figure 7.16 and Figure 7.17. This value is calculated by multiplying the design point sign vectors, then multiplying the resulting vector and response vector, and then dividing the result by 2^{k-1}(Schut 2007).



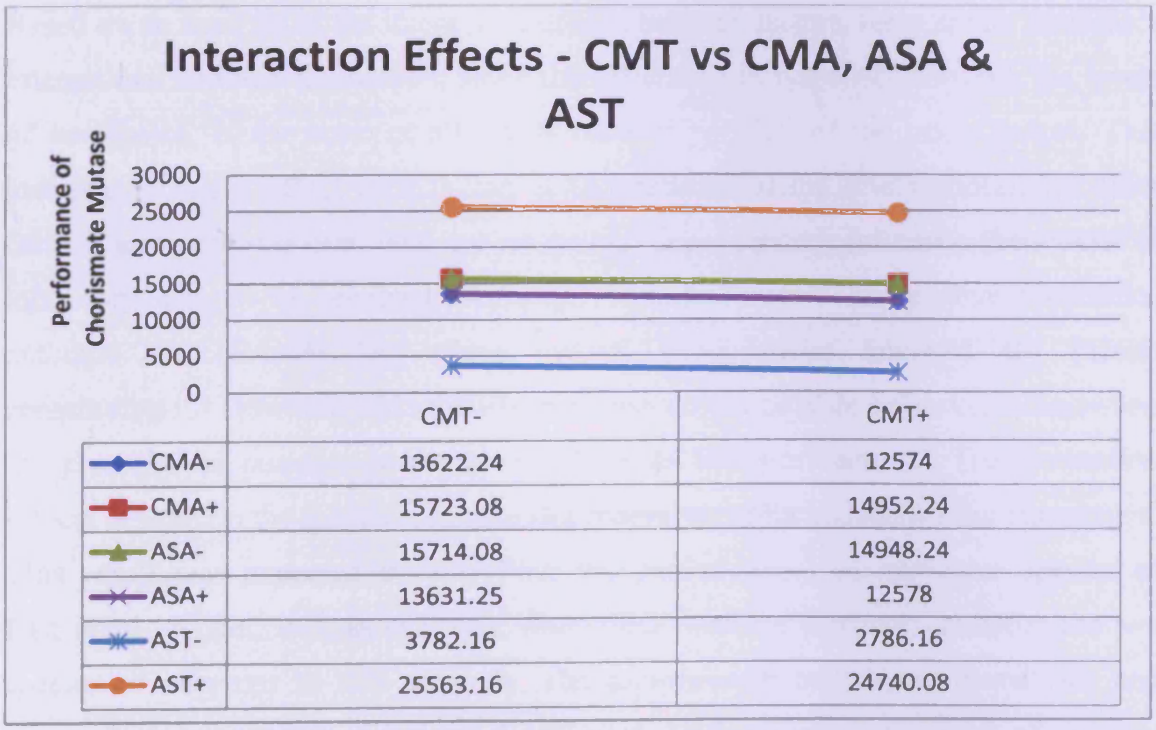
	CMA-	CMA+
ASA-	14278.99	16383.33
ASA+	11917.25	14292
CMT-	13622.24	15723.08
CMT+	12574	14952.24
AST-	2159.24	4409.08
AST+	24037	26266.24

Figure 7.14: The interaction effects between the abundance of Chorismate Mutase (CMA) vs abundance of Anthrinilate Synthase (ASA), Turnover Cycle of Chorismate Mutase (CMT) and Turnover Cycle of Anthrinilate Synthase (AST)



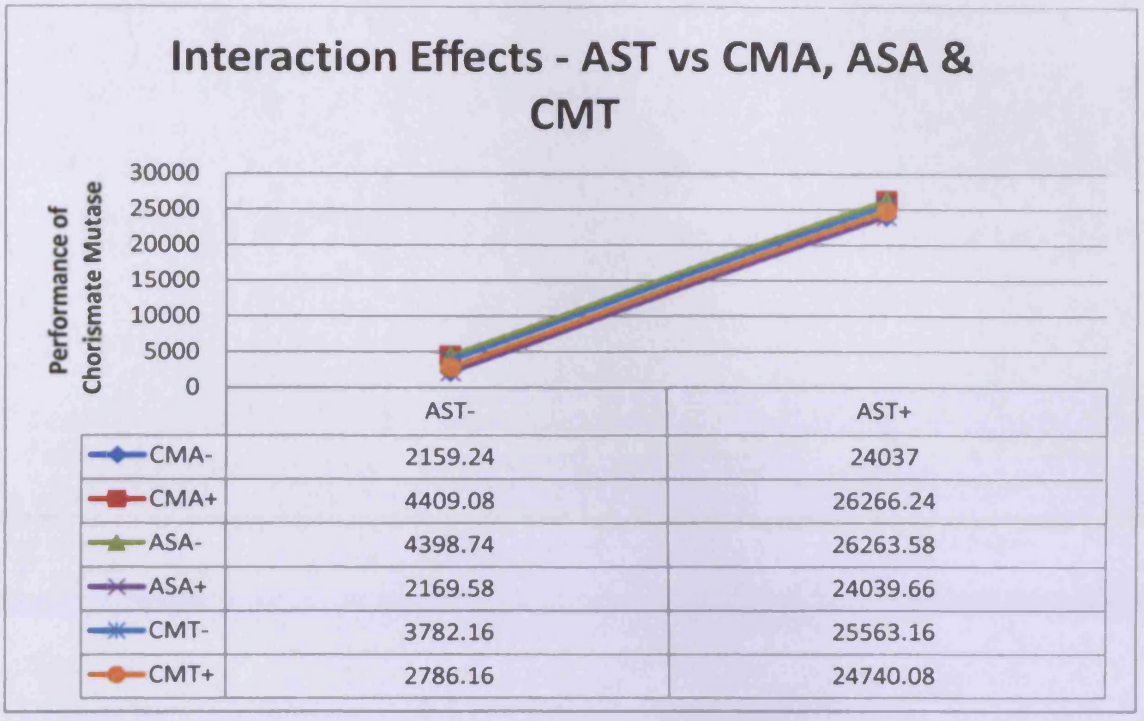
	ASA-	ASA+
CMA-	14278.99	11917.25
CMA+	16383.33	14292
CMT-	15714.08	13631.25
CMT+	14948.24	12578
AST-	4398.74	2169.58
AST+	26263.58	24039.66

Figure 7.15: The interaction effects between the abundance of Anthrinilate Synthase (ASA) vs abundance of Chorismate Mutase (CMA), Turnover Cycle of Chorismate Mutase (CMT) and Turnover Cycle of Anthrinilate Synthase (AST)



	CMT-	CMT+
CMA-	13622.24	12574
CMA+	15723.08	14952.24
ASA-	15714.08	14948.24
ASA+	13631.25	12578
AST-	3782.16	2786.16
AST+	25563.16	24740.08

Figure 7.16: The interaction effects between the Turnover Cycle of Chorismate Mutase (CMT) vs abundance of Chorismate Mutase (CMA), abundance of Anthrinilate Synthase (ASA) and Turnover Cycle of Anthrinilate Synthase (AST)



	AST-	AST+
CMA-	2159.24	24037
CMA+	4409.08	26266.24
ASA-	4398.74	26263.58
ASA+	2169.58	24039.66
CMT-	3782.16	25563.16
CMT+	2786.16	24740.08

Figure 7.17: The interaction effects between the Turnover Cycle of Anthrinilate Synthase (AST) vs abundance of Chorismate Mutase (CMA), abundance of Anthrinilate Synthase (ASA) and Turnover Cycle of Chorismate Mutase (CMT)

Based on an analysis of the interaction effects between factors, there are no noticeable interactions between the factors, since the difference in responses between the levels of one factor, is the same at all levels (almost parallel) of the other factors. This indicates, that the effect of a factor, is independent of the levels chosen for other factors, and confirms that there are no mutual dependencies, influence (beneficial or inhibitory effects) or feedbacks between factors. Positive or negative feedbacks, amongst the biomolecules, cause mutual dependencies amongst the factors constituting it. However, the interaction effects do not provide information regarding the presence of positive or negative influences between factors. The interaction effects depend on the independent and dependent variables chosen for the experiment. This result was expected in modelling the performances of particular species of functional product, such as enzymes, since there were no feedbacks between the two species of enzymes in this scenario. The factors such as enzyme abundance and enzyme's turnover cycle were associated solely to the respective species of enzyme. However, when modelling and measuring the performances of complex tasks, involving more than a single species of functional product, dependences will exist

between the factors causing interactions to occur between factors. These interactions are further complicated, when modelling complex tasks involving feedbacks between species of functional products.

7.5.6 Experiment Validation

The multi-objective topology is used to model two competing tasks in metabolic activities and analyse the effects of different factors, such as abundance and efficiency of enzymes on the performances of competing tasks. These factors act as organisational constraints, when providing solutions to cellular systems. Measuring the performance of functional products, such as enzymes, is crucial when modelling the contributions they make to the organisational behaviour of the cell. Performance of enzymes is a collective property, which depends on factors such as efficiency and abundance of activity, which compensate each other. The efficiency of the enzyme is amplified, via their redundant counterparts, which affects the overall performance of the intended task of the respective enzyme species.

The turnover number is reflected in the efficiency of enzymes, where the most efficient enzymes have very high turnover numbers. Mutations to enzymes can alter the turnover numbers, which in turn will affect the performance of the respective metabolic pathways. While some enzymes show large resistance to evolution, due to numerous inverse fitness interactions, others have fewer inverse fitness interactions and have more flexibility to alter their efficiency via mutations. A wild type enzyme is replaced by a mutant enzyme having different turnover number to model the effects at the population level. Thus an improved efficiency at the molecular level will improve the performance and hence the fitness of the respective tasks, but may have an adverse effect at a global level of cellular organisation. Hence enzymes tend to sacrifice their efficiency to improve the efficiency at the cellular level. The results show that with an increase in efficiency of a particular enzyme, the net gain of the pathway, in terms of end product production, has increased. These effects are profound in irreversible reactions, when compared to reversible reactions. Various processes can control the performance by positively or negatively influencing the task. Positive influences include the activation of enzymes and increasing the population of the enzymes. Negative influences include negative feedbacks, such as

enzyme deactivation and inhibition, and lowering the enzyme population via degradation.

7.5.7 Experiment Conclusion

This experiment has demonstrated how qualitative and quantitative features of competing enzyme species can have an inhibitory effect on the overall performances. Further, the simulation has demonstrated, that there are no feedbacks by analysing the interaction effects between factors. The simulation has made it possible to analyse the impact of a particular biomolecular species on the performance of higher tasks by altering their characteristics or silencing them. Hence, the main effects of these factors can be used to analyse the impact of biomolecular species in the presence and absence of specific biomolecular species. It has also made it possible to model the features which contribute to the intracellular adaptive dynamics, such as coordination, cooperation and competition between diverse biomolecular species.

7.6 Evaluation of the Collective Intelligence Framework

7.6.1 Functionalities Achieved

Measuring the performance of functional products such as enzymes is crucial in modelling the contributions they make to the intracellular organisational behaviour. The framework has the ability to represent the intracellular dynamics at a molecular and cellular resolution by representing the characteristics (attributes and behaviour) of functional products and by observing the system level behaviour. Further the performances of functional products can be associated with the fitness (the adaptive/evolutionary success) of the respective species of functional product and the genes involved in propagating them. To achieve this, the activities (the effective abundance) as opposed to the actual abundance of the respective functional products have to be analysed. However, conventionally the actual abundance, on which the kinetic models depend, is empirically measured rather than the effective abundance, which represents the activities contributing to cellular functions and adaptive/evolutionary success. This framework fills the gap by providing the environment to analyse the intracellular organisational behaviour, which cannot be

directly observed or empirically measured, by representing the functional products, capturing the results of their activities and providing the means to evaluate these results. Further the framework has the ability to integrate thermodynamic and energy constraints, which also have an impact on the behaviour of cellular systems. These constraints are not integrated in conventional kinetic models, which solely rely on rate laws.

The CI framework has the ability to model between the molecular and cellular resolution, by characterising the functional products in terms of their attributes and behaviour at the molecular resolution and observing their species/population level behaviour, which contributes to intracellular organisational behaviour. This requires representing multiple scales, from molecular to cellular resolution, simultaneously to analyse:

- the performances within the objective/task hierarchy,
- the timescales of molecular activities and the timescales at which their contributions can be realised,
- the energy requirements for the molecular activities and the energy production and consumption at the cellular resolution,
- the efficiency of functional product's activities, and the efficacy of the diverse objectives/tasks to which they contribute,
- the stability of the functional products, and their robustness at the cellular resolution, and
- the reactivity of functional products at physiological timescales, and the adaptability at the cellular resolution to which they contribute.

The CI framework has the ability to model adaptability at the molecular and organisational level by the characterisation of functional products which reflect on their coding sequence. These sequences act as replicators of the functional product's characteristics, and propagate to future generations of the biological systems involved. The adaptability at the organisational level is achieved via biasing the activities of the functional products in the form of internal and external stimulations to sustain the organisation without any alterations to the inherent characteristics of the functional products.

The CI framework has the ability to represent concurrency by modelling redundant counterparts of diverse species of specialised functional products at the molecular resolution. This facilitates the diverse activities of functional products to occur concurrently in space and time, and avoid the combinatorial explosion inherent in sequential representations.

Moreover, the above features facilitate the CI framework to mechanistically model and analyse the emergent properties of the cell from the molecular resolution. The analysis of the organisational behaviour within the cell is achieved by measuring the performances of the objectives/tasks. The performances of the tasks are quantified in order to quantify the functions of the cellular systems.

7.6.2 Resources Utilised

One of the aims of systems biology is to integrate heterogeneous data for developing models. The Collective Intelligence framework integrates more detailed biochemical information than the population based kinetic models. Modelling at molecular resolution required molecular level information, such as their diffusion constants, time and energy requirements for their activities, their localisations and abundance in the cell. Further biochemical thermodynamic information such as the Gibbs free energies of formations and reactions were integrated to model the physical constraints of biomolecular activity. The enzymes were represented by their processing time which is a characteristic of the enzymes and was derived from their K_{cat} values.

An agent based discrete event simulator was used to model the biomolecular activities in space and time. However modelling a complete biological cell can test the limits of the available software and hardware for simulating Collective Intelligence. These limitations included the capacity to handle billions of agents and the features of agent based simulation technologies.

7.6.3 Limitations of the Framework

Although this approach is ideal for modelling the biological cell, its main limitation is the computational requirement to model multi-cellular structures. Modelling the whole multi-cellular organism at the molecular resolution is not feasible based on the current capabilities of computational technology.

The Systems Biology Markup Language (SBML Contributors 2010) is the widely used tool for model integration and interoperability. However SBML is incompatible with the developed Collective Intelligence framework, since it is designed and used for kinetic models, where the kinetic parameters are specified in terms of the rate of change of molecular abundance. A new form of representation will be required to characterise every species of functional product in terms of their attributes and behaviour as illustrated in Section 4.3. This will require extending the Unified Modelling Language (UML) to Systems Biology (Magali and Debora Schuch da 2006).

7.7 Concluding Remarks

This chapter evaluated a Collective Intelligence framework by conducting a series of simulation experiments. Experiments were based on modelling physical and biological constraints involved in intracellular organisational behaviour, which affect the collective behaviour of biomolecules. In Section 7.3, Collective Intelligence is modelled at an individual level. This experiment has demonstrated by emulating empirically observed half-life at an individual level, rather than at the population level. This can be used to study intracellular self-organisation and Collective Intelligence phenomena which are a feature of biomolecular degradation. In Section 7.4, a Collective Intelligence experiment has demonstrated that both qualitative and quantitative factors can compensate each other to meet a performance. The ability to measure activities of functional products and relate them to performance, and consequently to intracellular functionalities is the distinctive characteristic of this framework. Also it has shown the ability to represent spatial, temporal and thermodynamic constraints within the framework contribute to being able to make a more realistic representation of intracellular dynamics. In Section 7.5, a Collective Intelligence experiment has demonstrated how qualitative and quantitative features of competing enzyme species can have an inhibitory effect on overall performance. Further the simulation has made it possible to analyse the impact of particular biomolecular species on performances of higher tasks by altering their characteristics or silencing them. Hence, the main effects of these factors can be used to analyse the impact of biomolecular species in the presence and absence of specific biomolecular species. It has also made it possible to model features which contribute to

intracellular adaptive dynamics, such as coordination, cooperation and competition between diverse biomolecular species. Based on the results of these experiments, the validity of the simulation experiment and the validity of the framework were demonstrated.

Chapter 8

Conclusion and Further Work

"Although the road ahead is long and winding, it leads to a future where biology and medicine are transformed into precision engineering."

- Hiroaki Kitano

8.1 Overview

This chapter provides answers to the investigated research questions based on the objectives and draws conclusions based on the findings of the research. Section 8.2 reviews the objectives set for the research and provides answers to the research question. This is followed by a review of the thesis in Section 8.3. Section 8.4 describes the contribution to knowledge by specifying the insights gained from the research. The chapter concludes by discussing further work and recommendations for future research that emerged from this research project in Section 8.5.

8.2 Review of Objectives

8.2.1 Characterising the major biomolecular self-organising mechanisms

The answers and conclusions for this objective and the associated questions were provided in Chapters 2 and 4. The main questions that this objective addressed are:

How self-maintenance is utilised in biological cells?

Since the cellular objectives are not maintained centrally, the cells have adopted a unique strategy to continuously realise their objectives or adaptive requirements by

removing obsolete information from the intracellular organisation. The propensity of native biomolecular degradation by means of random or regulated process and collective autocatalysis provides an ideal reinforcement adaptive mechanism for a cell. This process of native biomolecular degradation can eliminate obsolete biomolecular activities to keep cellular activities up to date, and recycle resources to maintain almost a steady biomolecular population in a resource constrained and dynamic environment.

What drives self-organisation that underlies Collective Intelligence in cells?

The process of self-organisation in biological cells is driven by organisational and physical constraints. Constraints in general, reduce uncertainty and facilitate order in biological cells. Feedbacks which range from highly specific to more general signals play a dominant role in intracellular organisational behaviour. While signals with high specificity provide precision control of biomolecular activities, signals with less specificity will have broad and vague control of activities. These varying degrees of specificity have constrained and guided self-organisation in biochemical systems. Section 4.2.1 has listed major positive and negative feedback mechanisms observed in biological cells. The cellular organisation has adopted the propensity of functional product degradation as a contributor of self-organisation of a cell. The cooperation and competition between biomolecular species contributes to the self-organisation process by acting as organisational and regulatory constraints for cellular adaptations. Enzymes, which are one of the key players in self-organisation, play a crucial role in metabolism, because they drive biologically desirable but thermodynamically unfavourable reactions by coupling them to favourable ones. The self-organisation processes in cells are non-spontaneous, because energy is required to produce various functional products to maintain order in cells. Various steady states of biological systems, which have emerged to maintain biological equilibrium far from thermodynamic equilibrium, attract non-spontaneous processes to increase order, whereas thermodynamic equilibrium attracts spontaneous processes to decrease order. The trajectory between these two biochemical system phases is controlled by metabolism, where anabolism is dominated by non-spontaneous processes, and catabolism is dominated by spontaneous processes, are coupled mostly using ATP as a shared medium.

How the limitations of every biomolecular activity will affect the intracellular organisational behaviour?

A feasible solution is only producible within the constraints of fundamental and organisational principles. The proactive nature of cellular behaviour is a result of biomolecules and their interactions in space and time. Although the rules, which remain constant at the physiological timescale, define what a particular species of biomolecule can perform, the uncertainty involved in when and where these rules are executed by redundant members of the species cause emergent behaviours in a cell. Further, each biomolecule is simply reacting in a determinate way to stimuli and in-turn responding by stimulating other biomolecules to regulate activities amongst them. Various activities are required to provide system wide responses to perturbations. However these activities have their limitations, and have to be regulated in terms of when, where and what activities should occur to provide timely responses to perturbations in a constrained environment. Based on physical constraints, every native bimolecular activity has limitations in terms of time and energy requirements. Various stages of regulation have evolved in anticipation of perturbations, which facilitated the transformation of the reactive activities of native biomolecules to a collectively proactive organisation. These regulatory mechanisms, range from transcriptional regulation (genetic level), post-transcriptional regulation, translational regulation (transcript level) and post-translational regulation (protein level). While transcriptional regulation provides slow and globalised cellular responses, post-translational regulation provides rapid and localised cellular responses. Transcriptional response is the most time and energy consuming process, since the genetic information has to be transcribed and mostly translated to produce a functional product. In contrast a post-translational response is the least time and energy consuming process, since the functional product is simply switched between an active and inactive state. The presence of higher stages of regulation such as translational and post-translational regulation, facilitate the anticipation of recurring perturbations, which also improves the performance of a cell.

8.2.2 Using a bottom-up integrative approach to model the intracellular organisational behaviour

The answers and conclusions for this objective and the associated questions were provided in Chapters 2, 3, 4 and 5. The main questions that this objective addressed are:

How to represent the collective behaviour of biomolecules in silico, to model cellular level phenomena?

The biomolecules are represented at an individual level and integrated into the Collective Intelligence framework, to study intracellular organisational behaviour of a cell. Intelligence is described as "the capability of a system to adapt its behaviour to meet its goals in a range of environment" (Fogel 2006). Intelligence is often associated with learning, which is an adaptive process. The ability to learn or adapt is one of the hallmarks of intelligent systems. This can also be witnessed in biological cells, where Cellular Intelligence emerges as an organisational level property from collective behaviour of biomolecules. Cellular Intelligence is the ability of biological cells to organise and adapt to perturbation and uncertainty, which reflects on the characteristics of intelligence. However, the process of adaptation is fundamentally different at the cellular level, since the intelligence resides not in individual native biomolecules, but in the diverse interactions/activities amongst them.

This approach facilitates analysis of the global effects of changes in behavioural rules imposed on diverse biomolecular species, where the effects of the rules are amplified due to redundant members of the biomolecular species. The representation at the level of molecular resolution also addresses the heterogeneous nature of the cellular environment and the existence of very low numbers of some functional products. Since the organisational behaviour within a cell cannot be directly observed or empirically measured, it requires a simulation framework that can represent native biomolecules, capture results of their activities and provide a way to evaluate these results.

What modelling approach can represent the intracellular organisational behaviour to study the emergence of cellular level characteristics such as adaptability, robustness and efficiency?

The cellular environment represents both biomolecules and their activities which contribute to the self * properties of a cell. The activities cause direct and indirect influences amongst various species of native biomolecules, which facilitate the self regulation of cellular processes. Agent based formalism is used in the wider framework of Collective Intelligence to model self-organisation and the emergence that occurs due to diverse biomolecular activities. Further the approach is driven by the principles of Swarm/Collective Intelligence to capture the inherent characteristics of the cell, such as adaptability, robustness and efficacy with no external supervision (Schut 2007). Some of the noteworthy properties of Collective Intelligence systems are adaptivity, emergence, global-local order, interaction, rules, redundancy, robustness and randomness (Schut 2007).

How do you functionally unite the activities of functional products from the bottom-up?

Although modularity can be observed in the biological organisation strata in terms of perceivable and physically bounded entities (molecules, organelles, cells, organs and individuals), their applicability in modularising intracellular activities of functional products into functional units, which constitute the cellular processes is doubtful. Intracellular functions, lack physical boundaries and are temporal phenomena, which emerge from the causally linked biomolecular activities. In the context of biological adaptation, function is defined as the progression along some causality, towards a goal or successful outcome. A logical approach to simplify cellular processes is by constructing/deconstructing these processes into objectives/tasks, on which selective pressure is imposed.

What are the criteria for indentifying functional units to represent intracellular tasks/objectives?

The criteria used to identify functional units by modularising the interactions among the functional products, are based on performance/fitness interactions, which emerge out of competition and cooperation amongst functional products. This is the mechanism by which evolution formed and evolved collaborative groups, containing one or more species of functional product. These functional products within a group cooperate with each other for a common objective/task. Competitive and cooperative adaptation among various biomolecular species is ubiquitous amongst their activities.

While inverse/inhibitory performance/fitness interaction exists between competing biomolecular species, positive/beneficial performance/fitness interaction will exist among cooperating biomolecular species. Beneficial and inhibitory performance interactions can reveal the organisation of the objective hierarchy in order to construct/deconstruct the tasks between molecular resolution and cellular resolution.

8.2.3 Developing a Collective Intelligence based cell modelling and simulation environment to conduct analysis studies

The answers and conclusions for this objective and the associated questions were provided in Chapters 3, 4, 5 and 7. The main questions that this objective addressed are:

What molecular level information is required to model biomolecules and their interactions?

Many macroscopic descriptions of cellular phenomena are only an approximation, idealisation and generalisation of real molecular processes. Due to insufficient description/information at this level, they often rely on probabilistic or statistical concepts. The microscopic descriptions of molecular activities are associated with detailed descriptions. However, many molecular details are insignificant, irrelevant and inconsequential to specific macroscopic phenomena (Fromm 2005). Hence, every detail at molecular resolution will not be required to represent the intracellular organisational behaviour. The level of detail required to represent phenomena will increase, when moving from population to molecular and atomic levels. Further the information used at each level is semantically different. The significant, relevant and prominent properties for activities and interactions representing intracellular organisational behaviour will have to be identified. The two types of constraints, which represent organisational and physical constraints will have to be represented, to model their effects on collective behaviour. This will require molecular level information, such as their diffusion constants, time and energy requirements for their activities, their localisations and abundance in a cell. Biomolecular activities are transformed into events, when they occur in stipulated space and time. Modelling these events will require information at molecular resolution, such as time and energy requirements to represent the respective events. Modelling event intervals will require

the diffusion constants of biomolecules, the distance amongst the biomolecules and the affinities for interaction.

Further biochemical thermodynamic information is required to model the physical constraints of the biomolecular activities. Gibbs free energy is mainly used as a thermodynamic property in biochemistry to provide quantitative answers to the probable direction of chemical reactions. Free energy is also used to represent affinities for interaction amongst biomolecules. Information such as biomolecular degradation and error frequencies for transcription, translation and replication can also be beneficial when building a comprehensive model of the cell.

What is the best approach to analyse the adaptive dynamics of biomolecular interaction?

Since the modelling of multi-scale adaptive dynamics from molecules to cell requires a mechanism based description of functional properties, which emerge as a result of molecular interactions, the study follows the bottom-up systems biology approach. This approach utilises a mechanistic model development process, where the structure of the model depends on the mechanistic principle adopted. Further the hierarchical representation of the intended study is based on a bottom-up methodology. This is because the aim of the study is to understand how biological cells dynamically adapt to multiple objectives concurrently, facilitated by constituent biomolecular activities, which require traversing from lower level molecular resolution to higher level cellular resolution. The objectives of biological systems are constantly evolving due to ever changing demands of their environment. Biological systems meet these demands by pursuing the objectives aided by their constituents, giving rise to biological processes, which manifest as biological functions. Further pathological processes have become an integral part of biological adaptation due to failure in achieving objectives caused by unanticipated constraints. The multi-objective topology provides a concurrent and hierarchical view of biological systems, whereas the network topology provides a sequential and horizontal view of biological systems.

How to measure and control organisational behaviour within the biological cell?

Numerous methods have been proposed to measure organisation in self-organising systems. However, these methods are too abstract and can only be applied to organisation of spatial structures, rather than functional organisation, which is

temporal in nature. Moreover, a mathematically oriented definition for self-organisation is proposed, which has the ability to quantify functional organisation. This is defined as "Self-organized behaviour in a complex system involving multiple performance measures is a sequence of system states corresponding to movement along a Pareto optimal frontier". In this context the functional organisation is measured in terms of efficiency of diverse functional units that constitute the intracellular organisational behaviour. Further the performances of the functional units should be measured to get the best possible optimal values as a whole.

How to measure intracellular performances?

The best approach to measure cellular performance is to identify cooperative modules and measure the performances of these modules. These performances are basically the measured activity levels of the member native biomolecular species. Measuring the activities of the native biomolecules rather than their actual abundance would reveal effective abundance. Apart from contributing to molecular crowding, biomolecules merely occupying the cellular environment will not have any major effect on the cellular processes. Their contributions are judged by their activities. As explained in Section 3.5.2.3 chemical activity of molecules provides the most accurate description of a chemical system. Nevertheless, the dynamic state of chemical systems are described in terms of concentrations as an approximation to chemical activity based on the assumption that the difference between concentration (the actual population) and chemical activity (effective population) is insignificant. However in biological cells where functional products are complex molecules and only certain states out of all the possible states, have the ability to perform the intended activity, there is a significant deviation between actual population and effective population. Hence actual abundance will not reflect the true Dynamic phase of a cell.

8.3 Thesis Review

The scope of systems biology covers top-down systems biology studies, bottom-up systems biology studies and discovering general principles of biological systems. It has been acknowledged that the success of systems biology depends not only on studies based on specific instance of life, but also on studies based on the principles governing the entire organisational space of life. Hence, modelling adaptive dynamics

is an essential requirement to understand the organisational space of biological systems. This requires the development of advanced models with molecular information that facilitate the prediction of cellular behaviour under various conditions. This is needed to reveal the cellular level characteristics and the underlying principles of cellular functions. The aim of this thesis was to investigate systems biology approaches to representing biological complexity from molecules to cells and developing computational approaches which bring abstract theories to practical use. An adaptation of a bottom-up systems biology approach and utilisation of a mechanistic model development process has created a computational model, using agent based formalism in a wider framework of Collective Intelligence to represent intracellular behavioural/functional organisation. The research contribution was determination of fundamental and organisational principles behind biological systems that define a possible design space of biological cells and applying these principles to build mechanistic models of biological phenomena.

This was followed by defining the adaptive dynamics of biological cells by utilising a multi-objective topology which differs from a conventional network topology based description of intracellular dynamics. Further, it has exemplified biological complexity from molecules to cell by deciphering a functional organisation of biological cells via a multi-objective representation of intracellular adaptive dynamics. Crucial factors involved in biological adaptation such as adaptability, robustness and efficacy in the context of multi-objective topology have been characterised. This provides a hierarchical and concurrent view of intracellular dynamics. An appropriate systems biology approach will have to be adopted to model self-organisation of biomolecular activities in order to study the emergence of intracellular behavioural organisation. Since it requires a mechanism based explanation, it has to be mechanistically modelled using a bottom-up approach and integrating molecular level information. Modelling at the level of molecular resolution will require representing molecular properties together with spatial and temporal constraints of a cellular environment. One of the challenges is, that the organisational behaviour of a cell, is not something that can be directly observed or empirically measured. Instead it needs a group of actors to represent functional products, represent a set of cellular resources utilised by these functional products, capture the results of functional products' activities and a method to evaluate these

results. The cellular activities, which correspond to a functional organisation are hierarchically organised into various basic tasks, merging to form complex and greater tasks of a cell.

The decision was to adapt a bottom-up systems biology approach and utilise a mechanistic model development process to develop a computational model, using agent based formalism in the wider framework of Collective Intelligence to represent intracellular behavioural/functional organisation. This is because the aim of the study was to understand how biological cells dynamically adapt to multiple objectives concurrently, facilitated by the constituent biomolecular activities, which require traversing from lower level molecular resolution to higher level cellular resolution. The multi-objective topology provides a concurrent and hierarchical view of biological systems, whereas a network topology provides a sequential and horizontal view of biological systems. However, mathematical models, which use a network topology, are designed to model at the population/aggregation level and are unable to model at the level of molecular resolution. The Collective Intelligence approach challenges the assumption used in classical chemistry for its applicability in cellular chemistry. Further this approach focuses on biomolecular activities rather than biomolecules because when, where and what biomolecular activities are performed are crucial for adaptive dynamics in the physiological timescale. Further it can be used to analyse the causation of biomolecular activities in space and time.

The core of the model is driven by the principles of Swarm/Collective Intelligence which capture the inherent characteristics of a cell such as adaptability, robustness and efficacy with no external supervision (Schut 2007). Modelling and simulating these characteristics is essential to truly understand the mechanism by which intracellular solutions emerge via various biomolecular activities to meet the adaptive requirements of cells. This insight is essential to understand the transformation between normal and pathological processes in cellular systems. Some of the noteworthy properties of Collective Intelligence systems are adaptivity, emergence, global-local order, interaction, rules, redundancy, robustness and randomness (Schut 2007). Out of the widely available agent based modelling and simulation toolkits, Repast Simphony was chosen mainly due to its rapid progress, versatility, support and expanding user community.

A Collective Intelligence approach is ideal to represent the adaptability that emerges out of the collective behaviour of biomolecules. Cellular Intelligence is defined as the ability to regulate when, where and what biomolecular activities occur to maintain biological equilibrium in diverse environments. Hence modelling collective behaviour of biomolecules will involve representing cellular adaptation in a Swarm/Collective Intelligence framework. The concepts of self-organisation and emergence underlie swarming and these systems are inherently adaptive, robust, flexible, stochastic and concurrent. The first step towards modelling intracellular organisational behaviour is, understanding the mechanisms that foster collective behaviour among biomolecules. The main features of Swarm Intelligence involve forms of limited or minimal communications and/or interactions, large numbers of interacting entities with limited reach, and some global efficient, emergent or self-organised behaviour (Fleischer 2003). Further the four basic ingredients for manifestation of self-organisation are (Bonabeau, Dorigo et al. 1999): Forms of positive feedback, forms of negative feedback, amplification of fluctuations, multiple interactions of multiple entities. The existing Swarm Intelligence techniques are unable to represent intracellular adaptive dynamics. Hence new techniques based on biomolecular inspired mechanisms will have to be developed. A Collective Intelligence framework is based on a meta-formalism, which can be used for complex and self-organising systems. The problem framework is based on Cellular Intelligence, that represents a biological cell's ability to organise and adapt to perturbation and uncertainty, which reflects on the characteristics of intelligence. The fundamental principles utilised are self-organisation and thermodynamics to represent biological and physical constraints, respectively. The dynamic framework utilises multi-objective topology as the core of the model and describes the logic of Collective Intelligence, which is used to construct/deconstruct tasks for the intracellular organisational behaviour of the cell in the physiological timescale.

A specification of a Collective Intelligence framework utilised for cell modelling and simulation environment was developed. The purpose was to describe the model's focus, resolution and complexity. The scope of the model is to study collective behaviour of biomolecules constituting a biological cell. The model utilises a bottom-up approach, where the lowest and highest levels of model representation are at the molecular and cellular resolution, respectively. The processes scheduled are based on

model scenarios which can include various combinations of sub-models. The design concepts of the model represent emergence, adaptability, objectives, learning, prediction, sensing, interactions, stochasticity, collectives and observation. These design concepts facilitate the integration of an agent based formalism into a wider framework of Collective Intelligence. The sub-models specify some of the universal constraints, contributing to the internal organisation of a cell, such as biomolecular mobility, biomolecular interaction, biomolecular degradation and error frequencies in transcription and translation. The model specification has provided required functionalities to implement a SwarmCell model and conduct analysis studies from molecules to cell using simulation experiments.

The implementation of the model specification followed, which was used to setup and run various simulation experiments based on Collective Intelligence scenarios. The purpose of the implementation was to simulate the collective behaviour of biomolecules constituting a biological cell. The agents have been used as an ingredient for the simulation of Collective Intelligence, which is facilitated by Repast Simphony (Repast S), an agent based discrete event simulation toolkit. The architecture of the SwarmCell simulation environment facilitates multi-scale modelling from the level of molecular resolution to cellular resolution, which is fundamental for the study of Collective Intelligence phenomena. The ability of the architecture to represent the shared environment is also beneficial as it can model indirect interactions amongst the biomolecules, which is a feature of Collective Intelligence. Moreover, multi-scale visualization and semantic zooming are two of the important functionalities represented by this architecture, which facilitates the analysis of Collective Intelligence phenomena in biological cells.

The Collective Intelligence framework was evaluated by conducting a series of simulation experiments. Experiments were based on modelling physical and biological constraints involved in intracellular organisational behaviour, which affect the collective behaviour of biomolecules. Collective Intelligence is modelled at an individual level. The first experiment demonstrated by emulating empirically observed half-life at an individual level, rather than at the population level. This can be used to study intracellular self-organisation and Collective Intelligence phenomena which are a feature of biomolecular degradation. The second Collective Intelligence experiment demonstrated that both qualitative and quantitative factors can

compensate each other to meet a performance. The ability to measure activities of functional products and relate them to performance, and consequently to intracellular functionalities is the distinctive characteristic of this framework. Also it has shown the ability to represent spatial, temporal and thermodynamic constraints within the framework contribute to being able to make a more realistic representation of intracellular dynamics. The final Collective Intelligence experiment demonstrated how qualitative and quantitative features of competing enzyme species can have an inhibitory effect on overall performance. Further the simulation has made it possible to analyse the impact of particular biomolecular species on performances of higher tasks by altering their characteristics or silencing them. Hence, the main effects of these factors can be used to analyse the impact of biomolecular species in the presence and absence of specific biomolecular species. It has also made it possible to model features which contribute to intracellular adaptive dynamics, such as coordination, cooperation and competition between diverse biomolecular species. Based on the results of these experiments, the validity of the simulation experiment and the validity of the framework were demonstrated.

8.4 Contribution to Knowledge

The research contributes to determination of fundamental and organisational principles behind biological systems that define a possible design space for biological cells and applying these principles to build mechanistic models of biological phenomena. The novelty of the thesis and its major contribution to knowledge is based on defining cellular functions in the context of a multi-objective topology and implementing this principle, as an *in silico* model, to study performances of intracellular functions by measuring activities of diverse species of functional products. This approach represents biological adaptation at the biochemical level which a network topology is unable to represent. The major contribution to computing is identifying a novel Collective Intelligence approach based on the information processing strategies of biomolecules and utilising it for modelling intracellular activities. The contributions include:

1. Use of an agent based formalism in the wider framework of Collective Intelligence, which considers principles and properties of self-organising

processes to determine fundamental and organisational principles of a biological cell (see Chapters 2 and 4).

2. Showing the significance of analysing biomolecular activities, rather than their abundance, as this provides an accurate description of a biochemical system. The biomolecular organisational behaviour is analysed by quantifying cellular functions by measuring performance of objectives/tasks formed by the activities of diverse functional products (see Sections 3.5.2.3 and 4.4).

3. Providing an environment to analyse organisational behaviour within a cell, that cannot be directly observed or empirically measured. This is achieved by using a simulation framework to represent functional products, capturing results of their activities and providing a method to evaluate these results (Periyasamy, Gray et al. 2008a; Periyasamy, Kille et al. 2008) (see Chapters 5 and 6).

4. Showing that cells have adopted a unique strategy to continuously realise their objectives/tasks or adaptive requirements (self-awareness) by eliminating obsolete information and generating new information in their internal organisation. The tendency for biomolecular degradation by means of random or regulated process and collective autocatalysis provides an ideal reinforcement adaptive mechanism for a cell (Periyasamy, Gray et al. 2008b) (see Section 2.5).

5. Implementing a novel system-theoretic approach to molecular systems biology by utilising biomolecular inspired multi-objective strategies from a Collective Intelligence perspective to capture higher level performances of a cell (Periyasamy, Gray et al. 2009) (see Section 7.5).

6. Using novel criteria to modularise interactions among functional products, which are based on performance interactions, emerging from competition and cooperation among the functional products (Periyasamy, Gray et al. 2009). Direct and inverse performance interactions can reveal the organisation of basic objectives/tasks into complex global tasks, in order to construct/deconstruct tasks between molecular resolution and cellular resolution (see Chapters 2 and 5).

8.5 Further Work

"Biology and computer science - life and computation – are related. I am confident that at their interface great discoveries await those who seek them."

- Leonard Adleman

There is an opportunity for Swarm/Collective Intelligence to unravel the hidden complexity of biological systems from molecules to cells. The most daunting task is to comprehend how biological tasks/objectives emerge as an ongoing process of optimisation to meet adaptive requirements, which is why this approach has the potential to unravel the complexity among the levels of molecular resolution and cellular resolution. While a scoring mechanism is essential to measure performance, a ranking mechanism facilitates by guiding molecular level interactions to a desired system level behaviour. Further, posing questions at the cellular resolution and seeking answers at the molecular resolution, and vice versa, is one of the challenges in multi-scale models. Scoring and ranking biomolecular activities will facilitate development of biomolecular inspired adaptive algorithms to conduct design studies. This is part of the fourth objective which has not yet been achieved. The purpose of design studies is problem solving, or seeking solutions to problems found in biological cells, namely remedies for pathological phases, or finding solutions, which engineer biological systems with new requirements. In delivering this objective we intend to address questions relating to engineering a biological cell as an *in silico* swarming system. These questions are: the construction and deconstruction of tasks from basic molecular activities to complex cellular activities of a minimal cell, the representation of the communication barriers amongst biomolecules; representation of the extremely concurrent nature of biomolecular interactions; incorporation of forms of positive and negative feedback and modelling amplification of fluctuations that give rise to solutions in a minimal cell

For this, further biomolecular optimisation strategies have to be implemented. Since optimisation strategies utilised in conventional swarm systems have adopted principles from higher levels of biological organisation, such as inter-organism adaptive processes, they do not represent adaptive strategies utilised by biomolecules. Hence, novel adaptive/optimisation strategies will have to be found and implemented

based on biomolecular self-organising mechanisms. The Collective Intelligence based biomolecular optimisation approach is based on a meta-formalism, that can be used for complex and self-organising systems (Fleischer 2005). This formalism is based on three foundational components, of which, two components have been addressed in detail. First, the problem framework is based on Cellular Intelligence, that is a biological cell's ability to organise and adapt to perturbation and uncertainty, which reflects on characteristics of intelligence. Second, the fundamental principles utilised are self-organisation and thermodynamics to represent biological and physical constraints respectively, which guides the intracellular organisation by reducing uncertainty. Third, the dynamic framework is based on the concept of Scale Invariant Pareto Optimality, which provides a novel way to characterise system interaction, behaviour and efficiency on different scales. However, as the initial stage, the multi-objective topology was utilised as the model structure to represent the logic of Collective Intelligence, which can be used to construct/deconstruct tasks for intracellular organisational behaviour of a cell in the physiological timescale. To fully comprehend Pareto Optimality, numerous intracellular objectives/tasks will have to be implemented. This will require representing a minimal cell, in itself a significant challenge.

To fully comprehend the SwarmCell framework, at the least a minimal cell will have to be implemented. This will require significant time and personnel. However this job can be simplified by attempting to model an organism with the smallest gene set. This framework can be extended for design studies such as for synthetic biology. There is growing interest in Synthetic Biology to identify the minimal genes require for a living organism. One group is focused on constructing chemical systems capable of replicating and evolving by being fed by small molecule nutrients (Forster and Church 2006). The Synthetic biology group at J. Craig Venter Institute has created the first synthetic bacterium species *Mycoplasma laboratorium* by gradually knocking out genes from *Mycoplasma genitalium* (Glass, Assad-Garcia et al. 2006), which is the natural free living organism with the smallest number of genes. However a recently discovered bacterial species *Carsonella ruddi* (an endosymbiont) is known to have the smallest genome, with an estimated 182 genes (Nakabachi, Yamashita et al. 2006).

The key features to be addressed, once a minimal cell is in place, are: the production of solutions and their persistence; emergent properties such as proactive behaviour

and robustness of a cell; modelling qualitative changes in collective behaviour of functional products via bifurcations, which produce new stable solutions due to perturbations; and modelling multi-stability, where for a given set of constraints, the systems can reach different stable states depending on initial conditions and random fluctuations (Garnier, Gautrais et al. 2007). Modelling these features will require development of biomolecular inspired adaptive algorithms to understand how novel solutions emerge based on the initial configuration and, physical and organisational constraints of a biological cell. This can also facilitate studying how cancer cells acquire unique capabilities by converging from various preliminary conditions during the process of adaptation within a multi-cellular system.

Appendix A

Publications

- Periyasamy, S., P. Kille and A. Gray. 2008. Biological Complexity in the Agent World. in *Proceedings of the IADIS International Conference Applied Computing*, Portugal, 10-13 April 2008. pp. 171-178.

Agent-based modelling and simulation (ABMS) are widely used to model complex systems in areas such as sociology, business, economics and ecology. In contrast, the use of ABMS to model biological complexity that range in scale from molecules to organisms is still in its infancy. Complex systems emerge due to local interactions between its simple entities and their environment. Modelling biological entities and their interactions provides significant challenges associated with multi-scaled spatial and temporal nature of the systems involved. We propose a novel prototype cell model implemented using the principles of Swarm/Collective Intelligence. This paper first describes the functional and non-functional requirements to implement "SwarmCell" - an Agent Based Cell Modelling and Simulation (ABCMS) environment that can be used to model and simulate local interactions between bio-molecules and their environment to predict higher level emergent structures. The paper then describes the design and the progress made in implementing the ABCMS using Repast Simphony - a general purpose ABMS environment, to represent biological complexity from molecules to cells.

- Periyasamy, S., A. Gray and P. Kille. 2008. The Epigenetic Algorithm. in *Proceedings of the IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, Hong Kong, 1-6 June 2008. pp. 3228-3236.

Evolutionary Computation (EC) paradigms are inspired by the optimization strategies utilized by biological systems. While these strategies can be found in every level of biological organization, almost all of the EC techniques which comprise techniques

from Evolutionary Algorithm (EA) to Swarm Intelligence (SI) have been inspired by organism level optimization strategies. While EA is based on trans-generational genetic adaptation of organisms (biologically inspired), SI is mainly based on intra-generational collective behavioural adaptation of organisms (socially inspired). This paper describes the optimization strategies that bio-molecules utilize both for intra-generational and trans-generational adaptation of biological cells. These adaptive strategies which are known as epigenetic mechanisms emerged long before any other biological strategy and form the basis for Epigenetic Algorithms (EGA). Further, the paper proposes an intra-generational EGA based on bio-molecular degradation and autocatalysis which are ubiquitous cellular processes and are pivotal for the adaptive dynamics and evolution of intelligent cellular organization.

- Periyasamy, S., A. Gray and P. Kille. 2008. A Collective Intelligence Approach to Modelling Intelligent Cellular Organisation. in *Proceedings of the International Conference on Systems Biology*, Gothenburg, Sweden, 23-27 August 2008. pp. 152 - 153.

Objective: The aim of this approach is to adopt principles of Collective Intelligence (CI) in representing the intelligent cellular organisation. A biological cell consists of various molecular species confined to distinct locations in the cell. These molecules have no centralised control and use a distributed problem solving strategy to sustain the cellular organisation. The main features of CI that we strive to implement are limited communication and interactions, large number of interacting entities with limited contact and some globally efficient, emergent or self-organising behaviour. We implement these features using an Agent Based Modelling and Simulation (ABMS) environment to capture the cellular level phenomena. Further we intend to progress towards developing an in silico based synthetic minimal biological cell.

Results: A prototype model using the above approach has been implemented using an ABMS toolkit. The reactive agents represent bio-molecules and the logic for these agents is much simpler than that of intelligent agents. The rules that depict the goals of the bio-molecules, aim to produce generalisable outcomes of the heterogeneous swarm. Although deliberately designing swarms to do specific cellular activities may

sound interesting and satisfying, it will be incapable of generating generalised methods to capture all cellular activities.

Conclusion: The collective intelligence approach to modelling intelligent cellular organisation is the ideal way of capturing intelligent cellular level behaviour. This behaviour is fundamental to capturing the adaptive dynamics that occurs in a cell's lifecycle. The most daunting task is to comprehend how biological structures emerge as an ongoing process of optimisation to occupy functional niches. This optimisation is powered by various levels of selection in the biological hierarchy. Further this approach could facilitate in capturing the mechanistic transition between biological and pathological processes at the cellular level and assess the impact of various molecular species on cellular level activities.

- Periyasamy, S., A. Gray and P. Kille. 2009. Multiscale Adaptive Dynamics from Molecules to Cells. in *Proceedings of the Foundations of Systems Biology in Engineering*, Denver, Colorado, USA, 9-12 August 2009: Omnipress. pp. 105 -108.

The paper proposes a novel system-theoretic approach to molecular systems biology by utilizing biomolecular inspired multi-objective optimization strategies from a collective intelligence perspective to capture the higher level performances of the cell. Based on the adaptive nature of biological systems and to achieve the aim of associating biological processes to the evolutionary mechanisms, the biological cell is represented in a multi dimensional problem, function and fitness space to analyze the multiple conflicting performances of biochemical activities. This approach could justify how cellular adaptation deals with multiple objectives simultaneously and specifies multi criteria conditions for the adaptation of intelligent cellular organisation. Further, it emphasises on optimization based analysis which deviates from the conventional mechanisms of analysing higher level cellular behaviour that uses various biochemical network based analysis techniques and methodologies.

Appendix B

The Dependencies Between Metabolites and Reactions

The complex dependencies of metabolites involved in the TCA cycle (P: Produced; C: Consumed)

Step	1	2	3	4	5	6	7	8	9
$\Delta_r G^o$ kcal mol^{-1}	-7.5	+2.0	-0.5	-2.0	-7.2	-0.8	~0.0	-0.9	+7.1
Oxaloacetate $\Delta_f G^o$ = -713.38 kJ mol^{-1}	C								P
Acetyl CoA $\Delta_f G^o$ = -60.49 kJ mol^{-1}	C								
Citrate $\Delta_f G^o$ = -963.46 kJ mol^{-1}	P	C							
CoA $\Delta_f G^o$ = -7.98 kJ mol^{-1}	P				C	P			
Isocitrate $\Delta_f G^o$ = -1156.04 kJ mol^{-1}			P	C					
Cis-Aconitate $\Delta_r G^o$ = -797.26 kJ mol^{-1}		P	C						
Malate $\Delta_f G^o$ = -682.88 kJ mol^{-1}								P	C
ά-Keto-glutarate $\Delta_f G^o$ = -633.58 kJ mol^{-1}				P	C				
Succinate $\Delta_f G^o$ = -530.72 kJ mol^{-1}						P	C		
Fumarate $\Delta_f G^o$ = -521.97 kJ mol^{-1}							P	C	
Succinyl CoA $\Delta_f G^o$ = -349.90 kJ mol^{-1}					P	C			
H_2O $\Delta_f G^o$ = -157.28 kJ mol^{-1}	C	P	C					C	
H^+ $\Delta_f G^o$ = 0.0 kJ mol^{-1}	P								P
ATP $\Delta_f G^o$ = -2292.61 kJ mol^{-1}						P			
ADP						C			

$\Delta_f G^o$ = -1428.93 kJ mol^{-1}									
NAD^+ $\Delta_f G^o$ = 1038.86 kJ mol^{-1}				C	C				C
NADH $\Delta_f G^o$ = 1101.47 kJ mol^{-1}				P	P				P
FAD $\Delta_f G^o$ = 1238.65 kJ mol^{-1}							C		
$FADH_2$ $\Delta_f G^o$ = 1279.68 kJ mol^{-1}							P		
P_i $\Delta_f G^o$ = -1058.56 kJ mol^{-1}						C			
CO_2 $\Delta_f G^o$ = -394.36 kJ mol^{-1}				P	P				

The chemical reactions in TCA cycle

Reaction	Enzyme
1 Acetyl CoA + oxaloacetate + H_2O → citrate + CoA + H^+	Citrate synthetase
2 Citrate ↔ cis-aconitate + H_2O	Aconitase
3 cis-Aconitate + H_2O ↔ isocitrate	Aconitace
4 Isocitrate + NAD^+ ↔ á-Keto-glutarate + CO_2 + NADH	Isocitrate dehydrogenase
5 á-Keto-glutarate + NAD^+ + CoA ↔ succinyl CoA + CO_2 + NADH	á-Keto-glutarate dehydrogenase
6 Succinyl CoA + P_i + ADP ↔ succinate + ATP + CoA	Succinyl CoA syntetase
7 Succinate + FAD ↔ fumarate + $FADH_2$	Succinate dehydrogenase
8 Fumarate + H_2O ↔ malate	Fumarase
9 Malate + NAD^+ ↔ oxaloacetate + NADH + H^+	Malate dehydrogenase

Appendix C

The Epigenetic Mechanisms

The trans-generational epigenetic mechanisms (Source (Allis, Jenuwein et al. 2006; Tost 2008))

Mechanism	Description
Gene silencing	Transcriptional gene silencing is a result of histone modifications. Post-transcriptional gene silencing is a result of mRNA destruction (i.e. RNAi)
Paramutation	Characteristic of a gene is remembered and observed in later generations, even if that particular version of the gene is no longer present. In is a RNA directed inheritance mechanism
Bookmarking	Transmit cellular memory of patterns of gene expression in a cell.
Genomic imprinting	Certain gene are expressed in a parent of origin specific manner (i.e. gene expression occurs from only one allele – not both allele).
Position effect	The effect on the expression of a gene when its location in a chromosome is changed.
Reprogramming	Remodeling of epigenetic markers (DNA methylation) during mammalian

	development.
Transvection	Interaction between corresponding allele of homologous chromosome which can lead to either gene activation or repression
Maternal effect	Genotype of mother is expressed in phenotype of its offspring.
X-inactivation	On of the two copies of the x-chromosome present in female mammals is inactivated.

Appendix D
The Features of Repast Simphony

Repast Simphony is a free and open source agent-based modelling toolkit that simplifies model creation and use. Repast Simphony offers users a rich variety of features including the following (Adapted from (Repast Development Team 2008)):

- Fluid model component development using any mixture of Java, Groovy, and flowcharts in each project;
- A pure Java point-and-click model execution environment that includes built-in results logging and graphing tools as well as automated connections to a variety of optional external tools including the R statistics environment, ORA and Pajek network analysis plugins, A live agent SQL query tool plugin, the VisAD scientific visualization package, the Weka data mining platform, many popular spreadsheets, the MATLAB computational mathematics environment, and the iReport visual report designer;
- An extremely flexible hierarchically nested definition of space including the ability to do point-and-click and modeling and visualization of 2D environments; 3D environments; networks including full integration with the JUNG network modeling library as well as Microsoft Excel spreadsheets and UCINET DL file importing; and geographical spaces including 2D and 3D Geographical Information Systems (GIS) support;
- A range of data storage "freeze dryers" for model check pointing and restoration including XML file storage, text file storage, and database storage;
- A fully concurrent multithreaded discrete event scheduler;
- Libraries for genetic algorithms, neural networks, regression, random number generation, and specialized mathematics;

- An automated Monte Carlo simulation framework which supports multiple modes of model results optimization;
- Built-in tools for integrating external models;
- Distributed computing with Terracotta;
- Full object-orientation;
- Optional end-to-end XML simulation
- A point-and-click model deployment system; and
- Availability on virtually all modern personal computing platforms including Windows, Mac OS, and Linux.

Appendix E

The Results of Simulation Experiments

(a) The responses of the dependent variable for simulating biomolecular degradation

Design Points		Iterations (Time – minutes)		
		1/2	1/4	1/8
Series 1 (Average life based)	Simulation run 1	23.20	33.00	38.15
	Simulation run 2	23.00	34.00	39.50
	Simulation run 3	22.50	33.50	39.15
	Simulation run 4	22.50	34.00	38.80
	Simulation run 5	23.40	33.80	39.00
Series 2 (Probability based)	Simulation run 1	28.75	59.00	87.20
	Simulation run 2	33.70	64.00	94.00
	Simulation run 3	30.00	62.30	95.00
	Simulation run 4	28.50	58.85	87.00
	Simulation run 5	30.30	58.80	87.40

(b) The responses of the dependent variable for simulating biochemical reaction

Design Points		Response (Time - ms)
Series 1 (+,+,+)	Simulation run 1	483.5
	Simulation run 2	483.5
	Simulation run 3	483.7
	Simulation run 4	483.5
	Simulation run 5	483.5
	Average	483.54
Series 2 (+,+,-)	Simulation run 1	268.5
	Simulation run 2	268.5
	Simulation run 3	268
	Simulation run 4	268.5
	Simulation run 5	268.5
	Average	268.4
Series 3 (+,-,+)	Simulation run 1	848
	Simulation run 2	848.15
	Simulation run 3	846.9
	Simulation run 4	847.5
	Simulation run 5	848.7
	Average	847.85
Series 4 (+,-,-)	Simulation run 1	473.3
	Simulation run 2	471.6
	Simulation run 3	472
	Simulation run 4	473.68
	Simulation run 5	472.5
	Average	472.62
Series 5 (-,+,+)	Simulation run 1	966.2
	Simulation run 2	965.8
	Simulation run 3	966
	Simulation run 4	965.8
	Simulation run 5	966

	Average	965.96
Series 6 (-,+,-)	Simulation run 1	535.5
	Simulation run 2	535.5
	Simulation run 3	536
	Simulation run 4	536
	Simulation run 5	535.5
	Average	535.7
Series 7 (-,-,+)	Simulation run 1	1712
	Simulation run 2	1709.03
	Simulation run 3	1711
	Simulation run 4	1715.35
	Simulation run 5	1715
	Average	1712.48
Series 8 (-,-,-)	Simulation run 1	955.5
	Simulation run 2	960
	Simulation run 3	961
	Simulation run 4	957.5
	Simulation run 5	956
	Average	958

The effects of change in factor levels with respect to iterations

Design Points	Factors			Response
	Enzyme abundance	Equilibrium phase	Turnover cycle	Time(ms)
Series 1 (+,+,+)	483.54	483.54	483.54	483.54
Series 2 (+,+,-)	268.4	268.4	-268.4	268.4
Series 3 (+,-,+)	847.85	-847.85	847.85	847.85
Series 4 (+,-,-)	472.62	-472.62	-472.62	472.62
Series 5 (-,+,+)	-965.96	965.96	965.96	965.96
Series 6 (-,+,-)	-535.70	535.70	-535.70	535.70
Series 7 (-,-,+)	-1712.48	-1712.48	1712.48	1712.48
Series 8 (-,-,-)	-958	-958	-958	958
High	518.1	563.4	1002.46	
Low	1043.04	997.74	558.68	
Effects from high to low	-524.94	-434.34	443.78	

(c) The variables and their value ranges used for simulation competition and cooperation in metabolic pathway

Design Points		Response			
		Activity of Chorismate Mutase	Activity of Anthrinilate Synthase	Prephenate Abundance	Anthranilate Abundance
Series 1 (+,+,+,+)	Simulation run 1	25,022	4000	85.022	69.00
	Simulation run 2	25,021	4000	85.021	69.00
	Simulation run 3	25,021	4000	85.021	69.00
	Average	25,021.33	4000	85,021.33	69.00
Series 2 (+,+,+,-)	Simulation run 1	2,796	26,275	62.796	91.275
	Simulation run 2	2,798	26,271	62.798	91.271
	Simulation run 3	2,801	26,264	62.801	91.264
	Average	2,798.33	26,270	62,798.33	91,270
Series 3 (+,+,-,+)	Simulation run 1	25,823	3,200	85,823	68.200
	Simulation run 2	25,822	3,200	85,822	68.200
	Simulation run 3	25,825	3,200	85.825	68.200
	Average	25,823.33	3,200	85,823.33	68,200
Series 4 (+,+,-,-)	Simulation run 1	3527	25,548	63.527	90.548
	Simulation run 2	3527	25,545	63.527	90.545
	Simulation run 3	3521	25,551	63.521	90.551
	Average	3525	25,548	63,525	90,548
Series 5 (+,-,+,+)	Simulation run 1	26,905	2,100	86.905	67.100
	Simulation run 2	26,916	2,100	86.916	67.100
	Simulation run 3	26,908	2,100	86.908	67.100
	Average	26909.66	2,100	86,909.66	67,100
Series 6 (+,-,+,-)	Simulation run 1	5,081	23,968	65.081	88.968
	Simulation run 2	5,077	23,967	65.077	88.967
	Simulation run 3	5,081	23,965	65.081	88.965

	Average	5079.66	23,966.67	65,079.66	88,967
Series 7 (+,-,-,+)	Simulation run 1	27,309	1,700	87,309	66.700
	Simulation run 2	27,310	1,700	87,310	66.700
	Simulation run 3	27,313	1,700	87,313	66.700
	Average	27,310.66	1,700	87,310.66	66,700
Series 8 (+,-,-,-)	Simulation run 1	6236	22,813	66.236	87.813
	Simulation run 2	6223	22,823	66.223	87.823
	Simulation run 3	6241	22,808	66.241	87.808
	Average	6233.33	22,814.67	66,233.33	87,815
Series 9 (-,+,+,+)	Simulation run 1	22,012	7,000	82.012	72.000
	Simulation run 2	22,009	7,000	82.009	72.000
	Simulation run 3	22,012	7,000	82.012	72.000
	Average	22011	7,000	82,012	72,000
Series 10 (-,+,+,-)	Simulation run 1	1,481	27,596	61.481	92.596
	Simulation run 2	1,480	27,593	61.480	92.593
	Simulation run 3	1,483	27,591	61.483	92.591
	Average	1481.33	27,593.33	61,481.33	92,593
Series 11 (-,+,-,+)	Simulation run 1	23,302	5,770	83.302	70.770
	Simulation run 2	23,302	5,776	83.302	70.776
	Simulation run 3	23,305	5,767	83.305	70.767
	Average	23303	5,771	83,303	70,771
Series 12 (-,+,-,-)	Simulation run 1	1,878	27,192	61.878	92.192
	Simulation run 2	1,871	27,202	61.871	92.202
	Simulation run 3	1,872	27,195	61.872	92.195
	Average	1873.66	27,196.33	61,873.66	92,196
Series 13 (-,-,+,+)	Simulation run 1	25,019	3,999	85,019	68.999
	Simulation run 2	25,017	4,000	85,017	69.000
	Simulation run 3	25,019	4,000	85.019	69.000
	Average	25018.33	3,999.67	85,018.33	69,000
Series 14 (-,-,+,-)	Simulation run 1	2,786	26,253	62.786	91.253
	Simulation run 2	2,786	26,260	62.786	91.260
	Simulation run 3	2,784	26.256	62.784	91.256

	Average	2,785.33	26,256.33	62,785.33	91,256
Series 15 (-,-,-,+)	Simulation run 1	25,814	3,200	85,814	68.200
	Simulation run 2	25,816	3,200	85,816	68.200
	Simulation run 3	25,817	3,200	85,817	68.200
	Average	25,815.66	3,200	85,815.66	68,000
Series 16 (-,-,-,-)	Simulation run 1	3,495	25,548	63.495	90.548
	Simulation run 2	3,503	25,542	63.503	90.542
	Simulation run 3	3,492	25,557	63.492	90.557
	Average	3,496.66	25,549	63,496.66	90,549

The effects of change in factor levels with respect to the performance of Chorismate Mutase

Design Points	Factors			
	Enzyme(1) abundance	Enzyme(2) abundance	Turnover cycle(1)	Turnover cycle(2)
Series 1(+,+,+,+)	+25,021.33	+25,021.33	+25,021.33	+25,021.33
Series 2(+,+,+,-)	+2,798.33	+2,798.33	+2,798.33	-2,798.33
Series 3(+,+,-,+)	+25,823.33	+25,823.33	-25,823.33	+25,823.33
Series 4(+,+,-,-)	+3525	+3525	-3525	-3525
Series 5(+,-,+,+)	+26909.66	-26909.66	+26909.66	+26909.66
Series 6(+,-,+,-)	+5079.66	-5079.66	+5079.66	-5079.66
Series 7(+,-,-,+)	+27310.66	-27310.66	-27310.66	+27310.66
Series 8(+,-,-,-)	+6233.33	-6233.33	-6233.33	-6233.33
Series 9(-,+,+,+)	-22011	+22011	+22011	+22011
Series 10(-,+,+,-)	-481.33	+481.33	+481.33	-481.33
Series 11(-,+,-,+)	-23303	+23303	-23303	+23303
Series 12(-,+,-,-)	-1873.66	+1873.66	-1873.66	-1873.66
Series 13(-,-,+,+)	-25018.33	-25018.33	+25018.33	+25018.33
Series 14(-,-,+,-)	-2785.33	-2785.33	+2785.33	-2785.33
Series 15(-,-,-,+)	-25815.66	-25815.66	-25815.66	+25815.66
Series 16(-,-,-,-)	-3496.66	-3496.66	-3496.66	-3496.66
High	122701.30/8	104836.98/8	110104.97/8	201212.97/8

	= 15,337.33	= 13,104.62	= 13763.12	= 25151.62
Low	104784.97/8 =13,098.12	122649.29/8 =15331.16	117381.3/8 = 14,672.66	26273.3/8 =3284.16
Effects from high to low Average response change	17916.33/8 = 2239.21	-17812.31/8 = -2226.54	-7276.33/8 = -909.54	174939.67/8 = 21867.46

Appendix F

The Calculation of Experimental Parameters

F:1 - The relationship between concentration and molecules in specific volumes

Concentration	Volume				
	1 μm^3	10 μm^3	100 μm^3	1000 μm^3	10000 μm^3
1 nM	6.022×10^{-1}	6.022×10^{0}	6.022×10^{1}	6.022×10^{2}	6.022×10^{3}
10 nM	6.022×10^{0}	6.022×10^{1}	6.022×10^{2}	6.022×10^{3}	6.022×10^{4}
100 nM	6.022×10^{1}	6.022×10^{2}	6.022×10^{3}	6.022×10^{4}	6.022×10^{5}
1 µM	6.022×10^{2}	6.022×10^{3}	6.022×10^{4}	6.022×10^{5}	6.022×10^{6}
10 µM	6.022×10^{3}	6.022×10^{4}	6.022×10^{5}	6.022×10^{6}	6.022×10^{7}
100 µM	6.022×10^{4}	6.022×10^{5}	6.022×10^{6}	6.022×10^{7}	6.022×10^{8}
1 mM	6.022×10^{5}	6.022×10^{6}	6.022×10^{7}	6.022×10^{8}	6.022×10^{9}
10 mM	6.022×10^{6}	6.022×10^{7}	6.022×10^{8}	6.022×10^{9}	6.022×10^{10}
100mM	6.022×10^{7}	6.022×10^{8}	6.022×10^{9}	6.022×10^{10}	6.022×10^{11}
1M	6.022×10^{8}	6.022×10^{9}	6.022×10^{10}	6.022×10^{11}	6.022×10^{12}

F:2 - The relationship between concentration and average distance amongst molecules

Concentration	Average distance
1 nM	1.469 µm
10 nM	0.682 µm
100 nM	0.316 µm
1 µM	0.1469 µm
10 µM	0.0682 µm
100 µM	0.0316 µm
1 mM	0.01469 µm
10 mM	0.00682 µm
100mM	0.00316 µm
1M	0.001469 µm

F:3 - The relationship between diffusion constants and distance travelled in specific time steps

Diffusion Constant $\mu m^2 s^{-1}$	Distance travelled (μm)			Diffusion Constant $\mu m^2 s^{-1}$	Distance travelled (μm)		
	1 s	1 ms	1μs		1 s	1 ms	1μs
1	2.44949	0.07746	0.002449	40	15.49193	0.489898	0.015492
2	3.464102	0.109545	0.003464	41	15.68439	0.495984	0.015684
3	4.242641	0.134164	0.004243	42	15.87451	0.501996	0.015875
4	4.898979	0.154919	0.004899	43	16.06238	0.507937	0.016062
5	5.477226	0.173205	0.005477	44	16.24808	0.513809	0.016248
6	6	0.189737	0.006	45	16.43168	0.519615	0.016432
7	6.480741	0.204939	0.006481	46	16.61325	0.525357	0.016613
8	6.928203	0.219089	0.006928	47	16.79286	0.531037	0.016793
9	7.348469	0.232379	0.007348	48	16.97056	0.536656	0.016971
10	7.745967	0.244949	0.007746	49	17.14643	0.542218	0.017146
11	8.124038	0.256905	0.008124	50	17.32051	0.547723	0.017321
12	8.485281	0.268328	0.008485	51	17.49286	0.553173	0.017493
13	8.831761	0.279285	0.008832	52	17.66352	0.55857	0.017664
14	9.165151	0.289828	0.009165	53	17.83255	0.563915	0.017833
15	9.486833	0.3	0.009487	54	18	0.56921	0.018
16	9.797959	0.309839	0.009798	55	18.1659	0.574456	0.018166
17	10.0995	0.319374	0.0101	56	18.3303	0.579655	0.01833
18	10.3923	0.328634	0.010392	57	18.49324	0.584808	0.018493
19	10.67708	0.337639	0.010677	58	18.65476	0.589915	0.018655
20	10.95445	0.34641	0.010954	59	18.81489	0.594979	0.018815
21	11.22497	0.354965	0.011225	60	18.97367	0.6	0.018974
22	11.48913	0.363318	0.011489	61	19.13113	0.604979	0.019131
23	11.74734	0.371484	0.011747	62	19.2873	0.609918	0.019287
24	12	0.379473	0.012	63	19.44222	0.614817	0.019442
25	12.24745	0.387298	0.012247	64	19.59592	0.619677	0.019596
26	12.49	0.394968	0.01249	65	19.74842	0.6245	0.019748
27	12.72792	0.402492	0.012728	66	19.89975	0.629285	0.0199
28	12.96148	0.409878	0.012961	67	20.04994	0.634035	0.02005
29	13.19091	0.417133	0.013191	68	20.19901	0.638749	0.020199
30	13.41641	0.424264	0.013416	69	20.34699	0.643428	0.020347
31	13.63818	0.431277	0.013638	70	20.4939	0.648074	0.020494
32	13.85641	0.438178	0.013856	71	20.63977	0.652687	0.02064
33	14.07125	0.444972	0.014071	72	20.78461	0.657267	0.020785
34	14.28286	0.451664	0.014283	73	20.92845	0.661816	0.020928
35	14.49138	0.458258	0.014491	74	21.07131	0.666333	0.021071
36	14.69694	0.464758	0.014697	75	21.2132	0.67082	0.021213
37	14.89966	0.471169	0.0149	76	21.35416	0.675278	0.021354
38	15.09967	0.477493	0.0151	77	21.49419	0.679706	0.021494
39	15.29706	0.483735	0.015297	78	21.63331	0.684105	0.021633

Diffusion Constant $\mu m^2 s^{-1}$	Distance travelled (μm)			Diffusion Constant $\mu m^2 s^{-1}$	Distance travelled (μm)		
	1 s	1 ms	1 μs		1 s	1 ms	1 μs
79	21.77154	0.688477	0.021772	90	23.2379	0.734847	0.023238
80	21.9089	0.69282	0.021909	91	23.36664	0.738918	0.023367
81	22.04541	0.697137	0.022045	92	23.49468	0.742967	0.023495
82	22.18107	0.701427	0.022181	93	23.62202	0.746994	0.023622
83	22.31591	0.705691	0.022316	94	23.74868	0.750999	0.023749
84	22.44994	0.70993	0.02245	95	23.87467	0.754983	0.023875
85	22.58318	0.714143	0.022583	96	24	0.758947	0.024
86	22.71563	0.718331	0.022716	97	24.12468	0.762889	0.024125
87	22.84732	0.722496	0.022847	98	24.24871	0.766812	0.024249
88	22.97825	0.726636	0.022978	99	24.37212	0.770714	0.024372
89	23.10844	0.730753	0.023108	100	24.4949	0.774597	0.024495

F:4 - The relationship between diffusion constants and the time required to travel a specific distance

Diffusion Constant $\mu m^2 s^{-1}$	Time required to travel a specific distance (s)						
	0.1 μm	0.2 μm	0.5 μm	1 μm	10 μm	100 μm	1 mm
1	0.001667	0.006667	0.041667	0.166667	16.66667	1666.667	166666.7
2	0.000833	0.003333	0.020833	0.083333	8.333333	833.3333	83333.33
3	0.000556	0.002222	0.013889	0.055556	5.555556	555.5556	55555.56
4	0.000417	0.001667	0.010417	0.041667	4.166667	416.6667	41666.67
5	0.000333	0.001333	0.008333	0.033333	3.333333	333.3333	33333.33
6	0.000278	0.001111	0.006944	0.027778	2.777778	277.7778	27777.78
7	0.000238	0.000952	0.005952	0.02381	2.380952	238.0952	23809.52
8	0.000208	0.000833	0.005208	0.020833	2.083333	208.3333	20833.33
9	0.000185	0.000741	0.00463	0.018519	1.851852	185.1852	18518.52
10	0.000167	0.000667	0.004167	0.016667	1.666667	166.6667	16666.67
11	0.000152	0.000606	0.003788	0.015152	1.515152	151.5152	15151.52
12	0.000139	0.000556	0.003472	0.013889	1.388889	138.8889	13888.89
13	0.000128	0.000513	0.003205	0.012821	1.282051	128.2051	12820.51
14	0.000119	0.000476	0.002976	0.011905	1.190476	119.0476	11904.76
15	0.000111	0.000444	0.002778	0.011111	1.111111	111.1111	11111.11
16	0.000104	0.000417	0.002604	0.010417	1.041667	104.1667	10416.67
17	9.8E-05	0.000392	0.002451	0.009804	0.980392	98.03922	9803.922
18	9.26E-05	0.00037	0.002315	0.009259	0.925926	92.59259	9259.259
19	8.77E-05	0.000351	0.002193	0.008772	0.877193	87.7193	8771.93
20	8.33E-05	0.000333	0.002083	0.008333	0.833333	83.33333	8333.333
21	7.94E-05	0.000317	0.001984	0.007937	0.793651	79.36508	7936.508

22	7.58E-05	0.000303	0.001894	0.007576	0.757576	75.75758	7575.758
23	7.25E-05	0.00029	0.001812	0.007246	0.724638	72.46377	7246.377
24	6.94E-05	0.000278	0.001736	0.006944	0.694444	69.44444	6944.444
25	6.67E-05	0.000267	0.001667	0.006667	0.666667	66.66667	6666.667
26	6.41E-05	0.000256	0.001603	0.00641	0.641026	64.10256	6410.256
27	6.17E-05	0.000247	0.001543	0.006173	0.617284	61.7284	6172.84
28	5.95E-05	0.000238	0.001488	0.005952	0.595238	59.52381	5952.381
29	5.75E-05	0.00023	0.001437	0.005747	0.574713	57.47126	5747.126
30	5.56E-05	0.000222	0.001389	0.005556	0.555556	55.55556	5555.556
31	5.38E-05	0.000215	0.001344	0.005376	0.537634	53.76344	5376.344
32	5.21E-05	0.000208	0.001302	0.005208	0.520833	52.08333	5208.333
33	5.05E-05	0.000202	0.001263	0.005051	0.505051	50.50505	5050.505
34	4.9E-05	0.000196	0.001225	0.004902	0.490196	49.01961	4901.961
35	4.76E-05	0.00019	0.00119	0.004762	0.47619	47.61905	4761.905
36	4.63E-05	0.000185	0.001157	0.00463	0.462963	46.2963	4629.63
37	4.5E-05	0.00018	0.001126	0.004505	0.45045	45.04505	4504.505
38	4.39E-05	0.000175	0.001096	0.004386	0.438596	43.85965	4385.965
39	4.27E-05	0.000171	0.001068	0.004274	0.42735	42.73504	4273.504
40	4.17E-05	0.000167	0.001042	0.004167	0.416667	41.66667	4166.667
41	4.07E-05	0.000163	0.001016	0.004065	0.406504	40.65041	4065.041
42	3.97E-05	0.000159	0.000992	0.003968	0.396825	39.68254	3968.254
43	3.88E-05	0.000155	0.000969	0.003876	0.387597	38.75969	3875.969
44	3.79E-05	0.000152	0.000947	0.003788	0.378788	37.87879	3787.879
45	3.7E-05	0.000148	0.000926	0.003704	0.37037	37.03704	3703.704
46	3.62E-05	0.000145	0.000906	0.003623	0.362319	36.23188	3623.188
47	3.55E-05	0.000142	0.000887	0.003546	0.35461	35.46099	3546.099
48	3.47E-05	0.000139	0.000868	0.003472	0.347222	34.72222	3472.222
49	3.4E-05	0.000136	0.00085	0.003401	0.340136	34.01361	3401.361
50	3.33E-05	0.000133	0.000833	0.003333	0.333333	33.33333	3333.333
51	3.27E-05	0.000131	0.000817	0.003268	0.326797	32.67974	3267.974
52	3.21E-05	0.000128	0.000801	0.003205	0.320513	32.05128	3205.128
53	3.14E-05	0.000126	0.000786	0.003145	0.314465	31.44654	3144.654
54	3.09E-05	0.000123	0.000772	0.003086	0.308642	30.8642	3086.42
55	3.03E-05	0.000121	0.000758	0.00303	0.30303	30.30303	3030.303
56	2.98E-05	0.000119	0.000744	0.002976	0.297619	29.7619	2976.19
57	2.92E-05	0.000117	0.000731	0.002924	0.292398	29.23977	2923.977
58	2.87E-05	0.000115	0.000718	0.002874	0.287356	28.73563	2873.563
59	2.82E-05	0.000113	0.000706	0.002825	0.282486	28.24859	2824.859
60	2.78E-05	0.000111	0.000694	0.002778	0.277778	27.77778	2777.778
61	2.73E-05	0.000109	0.000683	0.002732	0.273224	27.3224	2732.24
62	2.69E-05	0.000108	0.000672	0.002688	0.268817	26.88172	2688.172
63	2.65E-05	0.000106	0.000661	0.002646	0.26455	26.45503	2645.503
64	2.6E-05	0.000104	0.000651	0.002604	0.260417	26.04167	2604.167
65	2.56E-05	0.000103	0.000641	0.002564	0.25641	25.64103	2564.103

66	2.53E-05	0.000101	0.000631	0.002525	0.252525	25.25253	2525.253
67	2.49E-05	9.95E-05	0.000622	0.002488	0.248756	24.87562	2487.562
68	2.45E-05	9.8E-05	0.000613	0.002451	0.245098	24.5098	2450.98
69	2.42E-05	9.66E-05	0.000604	0.002415	0.241546	24.15459	2415.459
70	2.38E-05	9.52E-05	0.000595	0.002381	0.238095	23.80952	2380.952
71	2.35E-05	9.39E-05	0.000587	0.002347	0.234742	23.47418	2347.418
72	2.31E-05	9.26E-05	0.000579	0.002315	0.231481	23.14815	2314.815
73	2.28E-05	9.13E-05	0.000571	0.002283	0.228311	22.83105	2283.105
74	2.25E-05	9.01E-05	0.000563	0.002252	0.225225	22.52252	2252.252
75	2.22E-05	8.89E-05	0.000556	0.002222	0.222222	22.22222	2222.222
76	2.19E-05	8.77E-05	0.000548	0.002193	0.219298	21.92982	2192.982
77	2.16E-05	8.66E-05	0.000541	0.002165	0.21645	21.64502	2164.502
78	2.14E-05	8.55E-05	0.000534	0.002137	0.213675	21.36752	2136.752
79	2.11E-05	8.44E-05	0.000527	0.00211	0.21097	21.09705	2109.705
80	2.08E-05	8.33E-05	0.000521	0.002083	0.208333	20.83333	2083.333
81	2.06E-05	8.23E-05	0.000514	0.002058	0.205761	20.57613	2057.613
82	2.03E-05	8.13E-05	0.000508	0.002033	0.203252	20.3252	2032.52
83	2.01E-05	8.03E-05	0.000502	0.002008	0.200803	20.08032	2008.032
84	1.98E-05	7.94E-05	0.000496	0.001984	0.198413	19.84127	1984.127
85	1.96E-05	7.84E-05	0.00049	0.001961	0.196078	19.60784	1960.784
86	1.94E-05	7.75E-05	0.000484	0.001938	0.193798	19.37984	1937.984
87	1.92E-05	7.66E-05	0.000479	0.001916	0.191571	19.15709	1915.709
88	1.89E-05	7.58E-05	0.000473	0.001894	0.189394	18.93939	1893.939
89	1.87E-05	7.49E-05	0.000468	0.001873	0.187266	18.72659	1872.659
90	1.85E-05	7.41E-05	0.000463	0.001852	0.185185	18.51852	1851.852
91	1.83E-05	7.33E-05	0.000458	0.001832	0.18315	18.31502	1831.502
92	1.81E-05	7.25E-05	0.000453	0.001812	0.181159	18.11594	1811.594
93	1.79E-05	7.17E-05	0.000448	0.001792	0.179211	17.92115	1792.115
94	1.77E-05	7.09E-05	0.000443	0.001773	0.177305	17.7305	1773.05
95	1.75E-05	7.02E-05	0.000439	0.001754	0.175439	17.54386	1754.386
96	1.74E-05	6.94E-05	0.000434	0.001736	0.173611	17.36111	1736.111
97	1.72E-05	6.87E-05	0.00043	0.001718	0.171821	17.18213	1718.213
98	1.7E-05	6.8E-05	0.000425	0.001701	0.170068	17.0068	1700.68
99	1.68E-05	6.73E-05	0.000421	0.001684	0.16835	16.83502	1683.502
100	1.67E-05	6.67E-05	0.000417	0.001667	0.166667	16.66667	1666.667

Bibliography

Aebersold, R. 2005. "Molecular Systems Biology: a new journal for a new biology?" *Mol Syst Biol* 1, DOI: 10.1038/msb4100009

Alberty, R. A. 2005. "Thermodynamics in Biochemistry " *Encyclopedia of Life Sciences*, 1-10 DOI: 10.1038/npg.els.0003879

Alberty, R. A. 2009. "Resources for the Thermodynamics of Biochemical Reactions." MIT, Available at: http://web.mit.edu/chemistry/alberty/ [Accessed: October 06, 2009]

Allis, C. D., T. Jenuwein and D. Reinberg, Eds. (2006). *Epigenetics* Cold Spring Harbor Laboratory Press,U.S.

AMD. 2008. "Quad-Core AMD Opteron™ Processor with Direct Connect Architecture " Advanced Micro Devices, Inc, Available at: http://www.amd.com/us/Documents/4P_Server_Arch_Compare_44226B.pdf [Accessed: 06 May, 2010]

Ander, M., P. Beltrao, B. D. Ventura, J. Ferkinghoff-Borg, M. Foglierini, A. Kaplan, C. Lemerle, I. T. s-Oliveira and L. Serrano. 2004. "SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks " *Syst. Biol*(1), pp. 129-138.

Andersen, O. S. 2004. "A Brief History of The Journal of General Physiology." *J. Gen. Physiol.* 125(1), pp. 3-12.

Andrews, S. S. and D. Bray. 2004. "Stochastic simulation of chemical reactions with spatial resolution and single molecule detail." *Physical Biology*(1), pp. 137-151.

Avital, S., F. Jasmin, B. Dan, A. Yishai, J. S. Michael, E. J. A. Hubbard and H. David. 2008. "Toward Verified Biological Models." *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 5(2), pp. 223-234.

Bassingthwaighte, J. B., H. J. Chizeck and L. E. Atlas. 2006. "Strategies and Tactics in Multiscale Modeling of Cell-to-Organ Systems." *Proceedings of the IEEE* 94(4), pp. 819-831.

Belle, A., A. Tanay, L. Bitincka, R. Shamir and E. K. O'Shea. 2006. "Quantification of protein half-lives in the budding yeast proteome." *Proc Natl Acad Sci U S A* 103(35), pp. 13004–13009.

Bird, A. 2007. "Perceptions of epigenetics." *Nature* 447(7143), pp. 396-398.

Bonabeau, E., M. Dorigo and G. Theraulaz. 1999. *Swarm Intelligence: From Natural to Artificial Systems.*New York: Oxford University Press.

Bork, P. 2005. "Is there biological research beyond Systems Biology? A comparative analysis of terms." *Mol Syst Biol* 1, DOI: 10.1038/msb4100016

Boulianne, L., S. Al Assaad, M. Dumontier and W. Gross. 2008. "GridCell: a stochastic particle-based biological system simulator." *BMC Systems Biology* 2(1), pp. 66.

Brandon, R. N. 2001. "Philosophy of Selection: Units and Levels." *Encyclopedia of Life Sciences*, 5 DOI: 10.1038/npg.els.0003463

Brent, R. and J. Bruck. 2006. "2020 Computing: Can computers help to explain biology?" *Nature* 440(7083), pp. 416-417.

Broderick, G., M. Ru'aini, E. Chan and M. J. Ellison. 2005. "A life-like virtual cell membrane using discrete automata." *In Silico Biology* 5(2), pp 163-78

Brooks, C. L., M. Karplus and B. M. Pettitt. 1988. *Protein: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* New York: Wiley.

Bruggeman, F. J. and H. V. Westerhoff. 2007. "The nature of systems biology." *Trends in Microbiology* 15(1), pp. 45-50.

Buchanan, M. 2009. "Economics: Meltdown modelling - Could agent-based computer models prevent another financial crisis?" *Nature* 460(6), pp. 680-682

Byrnes, W. M., R. N. Goldberg, M. J. Holden, M. P. Mayhew and Y. B. Tewari. 2000. "Thermodynamics of reactions catalyzed by anthranilate synthase." *Biophysical Chemistry* 84(1), pp. 45-64.

Cannata, N., F. Corradini, E. Merelli, A. Omicini and A. Ricci. 2005. An Agent-Oriented Conceptual Framework for Systems Biology. *Transactions on Computational Systems Biology III.* Heidelberg, Springer. pp. 105 - 122.

Cardelli, L. 2005. Abstract Machines of Systems Biology. *Transactions on Computational Systems Biology III.* C. Priami, E. Merelli, P. P. Gonzalez and A. Omicini, Springer. 3737. pp. 145-168.

Castle, C. J. E. and A. T. Crooks. 2006. *Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations.* London: Centre for Advanced Spatial Analysis, University College London.

Catholijn, M. J. and T. Jan. 2007. "Agent-oriented modeling of the dynamics of biological organisms." *Applied Intelligence* 27(1), pp. 1-20.

CellML Project Community. 2001-2010. "CellML." The CellML Project, Available at: http://www.cellml.org/ [Accessed: March 08, 2010]

Chang A., Scheer M., Grote A., Schomburg I. and S. D. 2009. "BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009." *Nucleic Acids Res.* 37(Database issue), pp. D588-D592.

Church, G. M. 2005. "From systems biology to synthetic biology." *Mol Syst Biol* 1, DOI: 10.1038/msb4100007

Clayton, D. D. 1983. *Principles of Stellar Evolution and Nucleosynthesis.* The University of Chicago Press.

Cohen, I. R. and D. Harel. 2007. "Explaining a complex living system: dynamics, multi-scaling and emergence." *Journal of The Royal Society Interface* 4(13), pp. 175-182.

Corning, P. A. 1995. "Synergy and Self-Organization in the Evolution of Complex Systems." *SYSTEMS RESEARCH* 12(2), pp. 86-121.

Corning, P. A. 2002. "Thermoeconomics: Beyond the Second Law." *Journal of Bioeconomics* 4(1), pp. 57-88.

Corrado, P. 2009. "Algorithmic systems biology." *Commun. ACM* 52(5), pp. 80-88.

Coveney, P. V. and P. W. Fowler. 2005. "Modelling biological complexity: a physical scientist's perspective." *Journal of The Royal Society Interface* 2(4), pp. 267-280.

Cox, B. 2008. Do You Know What Time It Is? *Horizon*. K. Beetham. UK, BBC. 59 minutes.

Crespi, B. J. 2001. "Selection: Units and Levels." *Encyclopedia of Life Sciences*, 6 DOI: 10.1038/npg.els.0001712

Dastani, M. and J. J. Gomez-Sanz. 2005. "Programming multi-agent systems." *The Knowledge Engineering Review* 20(02), pp. 151-164.

David, L. S. and O. Virginie. 2008. "THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION?" *Evolution* 62(9), pp. 2155-2177.

Dogterom, M. 2001. "Cell Biophysics." *Encyclopedia of Life Sciences*, DOI: 10.1038/npg.els.0001271

Dubitzky, W. 2006. "Understanding the computational methodologies of systems biology." *Brief Bioinform* 7(4), pp. 315-317.

Dusenbery, D. B. 1992. *Sensory Ecology: How Organisms Acquire and Respond to Information.* W H Freeman & Co.

E-Cell Project. 2009. "E-Cell." Plone Foundation, Available at: http://www.e-cell.org/ecell [Accessed: 04 November, 2009]

Eberhart, R. C. and Y. Shi. 2007. *Computational Intelligence: Concepts to Implementations.*Burlington: Morgan Kaufmann.

Edelman, G. M. and J. A. Gally. 2001. "Degeneracy and complexity in biological systems." *PNAS* 98(24), pp. 13763-13768.

Eiben, A. E. and J. E. Smith. 2007. *Introduction to Evolutionary Computing.*Corr. 2nd printing. Springer.

Emonet, T., C. M. Macal, M. J. North, C. E. Wickersham and P. Cluzel. 2005. "AgentCell: a digital single-cell assay for bacterial chemotaxis." *Bioinformatics* 21(11), pp. 2714-2721.

Engelbrecht, A. P. 2005. *Fundamentals of Computational Swarm Intelligence.* Wiley.

Epstein, J. M. 2009. "Modelling to contain pandemics." *Nature* 460(7256), pp. 687-687.

Farmer, J. D. and D. Foley. 2009. "The economy needs agent-based modelling." *Nature* 460(7256), pp. 685-686.

Fisher, J. and T. A. Henzinger. 2007. "Executable cell biology." *Nat Biotech* 25(11), pp. 1239-1249.

Fleischer, M. 2003. Foundations of Swarm Intelligence: From Principles to Practice. *Swarming: Network enabled C4ISR.*

Fleischer, M. 2005. Scale invariant pareto optimality: a meta--formalism for characterizing and modeling cooperativity in evolutionary systems. *Proceedings of the 2005 conference on Genetic and evolutionary computation*,25 - 29, June, 2005, Washington DC, USA: ACM. pp. 233 - 240.

Fogel, D. B. 2006. *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*.3rd. Wiley-IEEE Press.

Forster, A. C. and G. M. Church. 2006. "Towards synthesis of a minimal cell." *Mol Syst Biol* 2, DOI: 10.1038/msb4100090

Fromm, J. 2005. "Ten Questions about Emergence." at: http://arxiv.org/abs/nlin.AO/0509049 [Accessed: 20 April, 2010]

Galperin, M. Y. and G. R. Cochrane. 2009. "Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009." *Nucl. Acids Res.* 37(suppl_1), pp. D1-4.

Garnier, S., J. Gautrais and G. Theraulaz. 2007. "The biological principles of swarm intelligence." *Swarm Intelligence* 1(1), pp. 3-31.

Gerstein, M. B., C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Weissman and M. Snyder. 2007. "What is a gene, post-ENCODE? History and updated definition." *Genome Research* 17(6), pp. 669-681.

Gilbert, D., H. Fuss, X. Gu, R. Orton, S. Robinson, V. Vyshemirsky, M. J. Kurth, C. S. Downes and W. Dubitzky. 2006. "Computational methodologies for modelling, analysis and simulation of signalling networks." *Brief Bioinform* 7(4), pp. 339-353.

Gilbert, N. and S. Bankes. 2002. "Platforms and methods for agent-based modeling." *Proceedings of the National Academy of Sciences of the United States of America* 99(Suppl 3), pp. 7197-7198.

Gillespie, D. T. 2007. "Stochastic Simulation of Chemical Kinetics." *Annual Review of Physical Chemistry* 58(1), pp. 35-55.

Glass, J. I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, H. O. Smith and J. C. Venter. 2006. "Essential genes of a minimal bacterium." *Proceedings of the National Academy of Sciences of the United States of America* 103(2), pp. 425-430.

Goldstein, R., P. Husbands, C. Fernando and D. Stekel. 2010. In silico Biology. *Pacific Symposium on Biocomputing*,4–8 January 2010, Kamuela, Hawaii, USA: World Scientific Publishing Co. Pte. Ltd. pp. 477-480.

Grace, S. C. 2004. Plant Bioenergetics. *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd. pp.

Grima, R. and S. Schnell. 2008. "Modelling reaction kinetics inside cells." *Essays Biochem*(45), pp. 41–56.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jorgensen, W. M. Mooij, B. Muller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Ruger, E. Strand, S. Souissi, R. A. Stillman, R. Vabo, U. Visser and D. L. DeAngelis. 2006. "A standard protocol for describing individual-based and agent-based models." *Ecological Modelling* 198(1-2), pp. 115-126.

Haider, S., B. Ballester, D. Smedley, J. Zhang, P. Rice and A. Kasprzyk. 2009. "BioMart Central Portal--unified access to biological data." *Nucl. Acids Res.* 37(suppl_2), pp. W23-27.

Hanson, R. M. and S. Green. 2008. *Introduction to Molecular Thermodynamics.* Sausalito, CA: University Science Books.

Hartman, J. L. I. V., B. Garvik and L. Hartwell. 2001. "Principles for the Buffering of Genetic Variation." *Science* 291(5506), pp. 1001-1004.

Hattne, J., D. Fange and J. Elf. 2005. "Stochastic reaction-diffusion simulation with MesoRD." *Bioinformatics* 21(12), pp. 2923-2924.

Heylighen, F. and C. Joslyn. 2001. Cybernetics and Second-Order Cybernetics. *Encyclopedia of Physical Science & Technology.* R. A. Meyers. New York, Academic Press. pp.

Hinchey, M. G., R. Sterritt and C. Rouff. 2007. "Swarms and Swarm Intelligence." *Computer* 40(4), pp. 111-113.

Hopi, E. H. and A. C. Jerry. 2007. "THE LOCUS OF EVOLUTION: EVO DEVO AND THE GENETICS OF ADAPTATION." *Evolution* 61(5), pp. 995-1016.

Howe, T. R., N. T. Collier, M. J. North, M. T. Parker and J. R. Vos. 2006. Containing Agents: Contexts, Projections, and Agents. *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects* Argonne National Laboratory, Argonne, IL USA.

Hunt, C. A., G. E. P. Ropella, S. Park and J. Engelberg. 2008. "Dichotomies between computational and mathematical models." *Nat Biotech* 26(7), pp. 737-738.

Hunter, P. 2008. "Feature: What genes remember." *Prospect*(146), 24th May 2008, pp.

Inchiosa, M. E. and M. T. Parker. 2002. "Overcoming design and development challenges in agent-based modeling using ascape." *Proceedings of the National Academy of Sciences of the United States of America* 99(Suppl 3), pp. 7304-7308.

Izaguirre, J. A., R. Chaturvedi, C. Huang, T. Cickovski, J. Coffland, G. Thomas, G. Forgacs, M. Alber, G. Hentschel, S. A. Newman and J. A. Glazier. 2004. "COMPUCELL, a multi-model framework for simulation of morphogenesis." *Bioinformatics* 20(7), pp. 1129-1137.

Jones, D. 2008. "All systems go." *Nat Rev Drug Discov* 7(4), pp. 278-279.

Joyce, A. R. and B. O. Palsson. 2006. "The model organism as a system: integrating 'omics' data sets." *Nat Rev Mol Cell Biol* 7(3), pp. 198.

Kahlem, P. and E. Birney. 2006. "Dry work in a wet world: computation in systems biology." *Mol Syst Biol* 2, DOI: 10.1038/msb4100080

Karsenti, E. 2008. "Self-organization in cell biology: a brief history." *Nat Rev Mol Cell Biol* 9(3), pp. 255-262.

Kast, P., Y. B. Tewari, O. Wiest, D. Hilvert, K. N. Houk and R. N. Goldberg. 1997. "Thermodynamics of the Conversion of Chorismate to Prephenate: Experimental Results and Theoretical Predictions." *The Journal of Physical Chemistry B* 101(50), pp. 10976-10982.

Katherine Hollywood, D. R. B. R. G. 2006. "Metabolomics: Current technologies and future trends." *PROTEOMICS* 6(17), pp. 4716-4723.

Kauffman, S. 1993. *The Origins of Order: Self-Organization and Selection in Evolution.* New York: Oxford University Press.

Kauffman, S. and P. Clayton. 2006. "On emergence, agency, and organization." *Biology and Philosophy* 21pp. 501-521.

Kazadi, S. 2003. The Genesis of Swarm Engineering. *Proceedings of the SCI2003 Conference, Special Session on Swarm Engineering,* Orlando, Florida, USA.

Kell, D. B. and J. D. Knowles. 2006. The Role of Modeling in Systems Biology. *System modeling in cellular biology: from concepts to nuts and bolts.* Z. Szallasi, J. Stelling and V. Periwal. Cambridge, MA, MIT Press. pp. 3-18.

Kennedy, J. and R. C. Eberhart. 2001. *Swarm Intelligence.* Morgan Kaufmann.

Kim, J. T. and R. Eils. 2008. "Systems Biology and Artificial Life: Towards Predictive Modeling of Biological Systems." *Artificial Life* 14(1), pp. 1-2.

Kitano, H. 2002a. "Computational systems biology." *Nature* 420, 206 DOI: 10.1038/nature01254

Kitano, H. 2002b. "Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology." *Current Genetics* 41(1), pp. 1-10.

Kitano, H. 2002c. "Systems Biology: A Brief Overview." *Science* 295(5560), pp. 1662-1664.

Kitano, H. 2004a. "BIOLOGICAL ROBUSTNESS." *Nature Reviews Genetics* 5, 826 DOI: 10.1038/nrg1471

Kitano, H. 2004b. "CANCER AS A ROBUST SYSTEM: IMPLICATIONS FOR ANTICANCER THERAPY." *Nature Reviews Cancer* 4(3), pp. 227.

Kitano, H. 2005. "International alliances for quantitative modeling in systems biology." *Molecular Systems Biology* 1, 1-2 DOI: 10.1038/msb4100011

Kitano, H. 2006. "Computational cellular dynamics: a network-physics integral." *Nature Reviews Molecular Cell Biology* 7(3), pp. 163.

Kitano, H. 2007. "Towards a theory of biological robustness." *Molecular Systems Biology* 3, 1-7 DOI: 10.1038/msb4100179

Kozlenkov, A., R. Penaloza, V. Nigam, L. Royer, G. Dawelbait and M. Schroeder. 2006. Prova: Rule-based Java Scripting for Distributed Web Applications: A Case Study in Bioinformatics. *Current Trends in Database Technology – EDBT 2006.* Heidelberg, Springer. 4254/2006. pp. 899 - 908.

Krieger, K. 2006. "COMPUTER SCIENCE: Life in Silico: A Different Kind of Intelligent Design." *Science* 312(5771), pp. 189-190.

Kwiatkowska, M. Z. and J. K. Heath. 2009. "Biological pathways as communicating computer systems." *J Cell Sci* 122(16), pp. 2793-2800.

Kyung-Joong Kim, S.-B. C. 2006. "A Comprehensive Overview of the Applications of Artificial Life " *Artif. Life* 12(1), pp. 153-182

Langton, C., Ed. (1988). *Artificial Life*. Redwood City CA, Addison-Wesley.

Langton, C. G., Ed. (1992). *Preface*. Artificial Life II. Redwood City, Addison-Wesley.

Laursen, L. 2009. "Computational biology: Biological logic." *Nature* 462, 408-410 DOI: 10.1038/462408a

Le Novere, N. and T. S. Shimizu. 2001. "STOCHSIM: modelling of stochastic biomolecular processes." *Bioinformatics* 17(6), pp. 575-576.

Lee, J. M., E. P. Gianchandani and J. A. Papin. 2006. "Flux balance analysis in the era of metabolomics." *Brief Bioinform* 7(2), pp. 140-150.

Lee, S. Y., S. Park, C. Yun and H. Yun. 2007. "WebCell." KAIST Available at: http://www.webcell.org/ [Accessed: April 11, 2010]

Lila, K. and R. Grzegorz. 2008. "The many facets of natural computing." *Commun. ACM* 51(10), pp. 72-83.

Linda, M. S., W. P. Daniel and K. Marc. 2006. An aspect-oriented approach for modeling self-organizing emergent structures. *Proceedings of the 2006 international workshop on Software engineering for large-scale multi-agent systems*. Shanghai, China, ACM. pp 59-65.

Liu, E. T. 2005. "Integrative biology and systems biology." *Mol Syst Biol* 1, DOI: 10.1038/msb4100008

Luck, M. and E. Merelli. 2005. "Agents in bioinformatics " *The Knowledge Engineering Review* 20(02), pp. 117-125.

Luke, S., C. Cioffi-Revilla, L. Panait and K. Sullivan. 2004. MASON: A New Multi-Agent Simulation Toolkit. *Proceedings of the 2004 SwarmFest Workshop*,9-11, May, 2004, Ann Arbor, Michigan USA.

Macal, C. M. and M. J. North. 2008. "Agent-Based Modeling and Simulation for Exascale Computing " *SciDAC Review*(8), pp. 34 - 41.

Magali, R.-R. and R. Debora Schuch da. 2006. Ten top reasons for systems biology to get into model-driven engineering. *Proceedings of the 2006 international workshop on Global integrated model management*,22 - 22, May, 2006, Shanghai, China: ACM. pp. 55 - 58.

Magali Roux-Rouquié and M. Soto. 2005. Virtualization in Systems Biology: Metamodels and Modeling Languages for Semantic Data Integration *Transactions on Computational Systems Biology I*. Berlin / Heidelberg, Springer. 3380. pp. 28-43.

Malone, T. W. 2006. "What is collective intelligence and what will we do about it?" MIT Center for Collective Intelligence, Available at: http://cci.mit.edu/about/MaloneLaunchRemarks.html [Accessed, 11 October 2010]

Markus, P. and B. Andreas. 2004. A Brief Top-Down and Bottom-Up Philosophy on Software Evolution. *Proceedings of the 7th International Workshop on Principles of Software Evolution*,6-7 Sept. 2004: IEEE Computer Society. pp. 131 - 136.

Materi, W. and D. S. Wishart. 2007. "Computational Systems Biology in Cancer: Modeling Methods and Applications." *Gene Regulation and Systems Biology* 2007(GRSB-1-Materi-et-al), pp. 91.

Merelli, E., G. Armano, N. Cannata, F. Corradini, M. d'Inverno, A. Doms, P. Lord, A. Martin, L. Milanesi, S. Moller, M. Schroeder and M. Luck. 2006. "Agents in bioinformatics, computational and systems biology." *Brief Bioinform* pp. bbl014.

Meuer, H., E. Strohmaier, H. Simon and J. Dongarra. 2009. "34th Top500 Supercomputer sites Report." The top 500 Supercomputer sites, Available at: http://www.top500.org/lists/2009/11 [Accessed: 10 May, 2010]

Michener, W. K., T. J. Baerwald, P. Firth, M. A. Palmer, J. L. Rosenberger, E. A. Sandlin and H. Zimmerman. 2001. "Defining and Unraveling Biocomplexity." *BioScience* 51(12), pp. 1018-1023.

Miller, P. 2007. "Swarm Theory." National Geographic, Available at: http://ngm.nationalgeographic.com/2007/07/swarms/miller-text [Accessed: 10 May, 2010]

Minar, N., R. Burkhart, C. Langton and M. Askenazi. 1996. *The Swarm Simulation System: A Toolkit for Building Multi-agent Simulations.* Santa Fe: Santa Fe Institute.

Minson, R. and G. K. Theodoropoulos. 2008. "Distributing RePast agent-based simulations with HLA." *Concurrency and Computation: Practice and Experience* 20(10), pp. 1225-1256.

Montgomery, D. C. 2008. *Design and Analysis of Experiments.*7th Edition. New York: John Wiley & Sons Ltd.

Nagpal, R. 2003. Towards a Catalog of Biologically-inspired Primitives. *Workshop on Engineering Self-organising Applications, Autonomous Agents and Multiagents Systems Conference (AAMAS).*

Nakabachi, A., A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran and M. Hattori. 2006. "The 160-Kilobase Genome of the Bacterial Endosymbiont Carsonella." *Science* 314(5797), pp. 267.

Nazaret, C., M. Heiske, K. Thurley and J.-P. Mazat. 2009. "Mitochondrial energetic metabolism: A simplified model of TCA cycle with ATP production." *Journal of Theoretical Biology* 258(3), pp. 455-464.

Nestorov, I. S., S. T. Hadjitodorov, I. Petrov and M. Rowland. 1999. "Empirical Versus Mechanistic Modelling: Comparison of an Artificial Neural Network to a Mechanistically Based Model for Quantitative Structure Pharmacokinetic Relationships of a Homologous Series of Barbiturates." *AAPS PharmSci* 1(4), pp. 5–13.

Ng, A., B. Bursteinas, Q. Gao, E. Mollison and M. Zvelebil. 2006. "Resources for integrative systems biology: from data through databases to networks and dynamic system models." *Brief Bioinform* 7(4), pp. 318-330.

Nobel, P. S. 2009. *Physicochemical and Environmental Plant Physiology.*Fourth Ed. Canada: Elsevier Inc.

Noble, D. 2008. *The Music of Life: Biology Beyond Genes.* Oxford University Press, USA.

North, M. J. and C. M. Macal. 2007. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation.* Oxford University Press, USA

North, M. J., N.T. Collier, and J.R. Vos. 2006. "Experiences Creating Three Implementations of the Repast Agent Modeling Toolkit." *ACM Transactions on Modeling and Computer Simulation* 16(1), pp. 1-25.

NRCAM. 2009. "Run the Virtual Cell." NRCAM, Available at: http://www.nrcam.uchc.edu/login/login.html [Accessed: 04 November, 2009]

Nvidia. 2009. "NVIDIA TESLA PERSONAL SUPERCOMPUTER." NVIDIA Corporation, Available at: http://www.nvidia.com/object/personal_supercomputing.html [Accessed: 06 July, 2009]

Odell, J. 2002. "Objects and Agents Compared." *Journal of Object Technology* 01(01), pp. 41-53.

OpenABM Consortium. 2010. "Open agent based modelling consortium." Arizona State University, Available at: http://www.openabm.org/ [Accessed: 22 April, 2010]

Pahle, J. 2009. "Biochemical simulations: stochastic, approximate stochastic and hybrid approaches." *Brief Bioinform* 10(1), pp. 53-64.

Palmer, D. W., M. Kirschenbaum and L. Seiter. 2005. Emergence-Oriented Programming. *2005 IEEE International Conference onSystems, Man and Cybernetics*,12-12 Oct. 2005. pp. 1441-1448.

Parunak, H. V. D. 2003. Making Swarming Happen. *Proceedings of Swarming and Network-Enabled C4ISR*,3 January 2003, Tysons Corner, VA, ASD C3I. pp. 1 - 18.

Parunak, H. V. D. and S. A. Brueckner. 2004. "Engineering Swarming Systems." *Methodologies and Software Engineering for Agent Systems* pp. 341-376.

Periyasamy, S., A. Gray and P. Kille. 2008a. A Collective Intelligence approach to modelling intelligent cellular organisation. *International Conference on Systems Biology*,23-27 August 2008, Gothenburg, Sweden. pp. 152 - 153.

Periyasamy, S., A. Gray and P. Kille. 2008b. The Epigenetic Algorithm. *IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*,1-6 June 2008, Hong Kong. pp. 3228-3236.

Periyasamy, S., A. Gray and P. Kille. 2009. Multiscale Adaptive Dynamics from Molecules to Cells. *Foundations of Systems Biology in Engineering*,9 -12 August 2009, Denver, Colorado, USA: Omnipress. pp. 105 -108.

Periyasamy, S., P. Kille and A. Gray. 2008. Biological Complexity in the Agent World. *IADIS International Conference Applied Computing*,10 -13 April 2008, Portugal. pp. 171-178.

Plimpton, S. J. and A. Slepoy. 2005. "Microbial cell modeling via reacting diffusing particles." *Journal of Physics* 16pp. 305-309

Prud'homme, B., N. Gompel and S. B. Carroll. 2007. "Emerging principles of regulatory evolution." *Proceedings of the National Academy of Sciences* 104(Suppl 1), pp. 8605-8612.

Railsback, S. F., S. L. Lytinen and S. K. Jackson. 2006. "Agent-based Simulation Platforms: Review and Development Recommendations." *SIMULATION* 82(9), pp. 609-623.

Rando, O. J. and K. J. Verstrepen. 2007. "Timescales of Genetic and Epigenetic Inheritance." *Cell* 128(4), pp. 655-668.

Regenmortel, M. H. V. V. 2004. "Reductionism and complexity in molecular biology " *EMBO Rep* 5(11), pp. 1016–1020.

Repast Development Team. 2008. "Repast Simphony." Argonne National Laboratory, Available at: http://repast.sourceforge.net/ [Accessed: 16 May, 2010]

Ridgway, D., G. Broderick and M. J. Ellison. 2006. "Accommodating space, time and randomness in network simulation." *Current Opinion in Biotechnology* 17(5), pp. 493-498.

Robin, H. 1990. "MECHANISMS FOR THE CONTROL OF GENE ACTIVITY DURING DEVELOPMENT." *Biological Reviews* 65(4), pp. 431-471.

Rodriguez, S., V. Hilaire, P. Gruer and A. Koukam. 2007. "A formal holonic framework with proved self-organizing capabilities." *International Journal of Cooperative Information Systems* 16(1), pp. 7.

Rosenfeld, J. S. 2002. "Functional redundancy in ecology and conservation." *Oikos* 98(1), pp. 156-162.

Sanford, C., M. L. K. Yip, C. White and J. Parkinson. 2006. "Cell++--simulating biochemical pathways." *Bioinformatics* 22(23), pp. 2918-2925.

Saric, T. and A. L. Goldberg. 2006. "Protein Degradation and Turnover." *Encyclopedia of Life Sciences*, DOI: 10.1038/npg.els.0005722

SBML Contributors. 2010. "The Systems Biology Markup Language." at: [Accessed: 8 May, 2010]

Schnell, S., R. Grima and P. K. Maini. 2007. "Multiscale modeling in biology." *American Scientist* 95(2), pp. 134.

Schut, M. C. 2007. *Scientific Handbook for Simulation of Collective Intelligence.* Published under Creative Commons licence.

Segré , D., D. Ben-Eli and D. Lancet. 2000. "Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies." *Proceedings of the National Academy of Sciences of the United States of America* 97(8), pp. 4112-4117.

Segré, D. and D. Lancet. 2000. "Composing life " *EMBO Rep* 1(3), pp. 217-222.

Shapiro, R. 2007. "A Simpler Origin for Life." *Scientific American Magazine* 296(6), June 2007, pp. 24-31.

Shenhav, B., A. Oz and D. Lancet. 2007. "Coevolution of compositional protocells and their environment." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1486), pp. 1813-1819.

Snoep, J. L., F. Bruggeman, B. G. Olivier and H. V. Westerhoff. 2006. "Towards building the silicon cell: A modular approach." *Biosystems* 83(2-3), pp. 207-216.

Solé, R. V., A. Munteanu, C. Rodriguez-Caso and J. Macía. 2007. "Synthetic protocell biology: from reproduction to computation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1486), pp. 1727-39.

Stelling, J., U. Sauer, F. J. D. III and J. Doyle, Eds. (2006). *Complexity and Robustness of Cellular Systems*. System Modeling in Cellular Biology: From Concepts to Nuts and Bolts. Cambridge, MA, The MIT Press.

Stelling, J., U. Sauer, Z. Szallasi, F. J. Doyle and J. Doyle. 2004. "Robustness of Cellular Functions." *Cell* 118(6), pp. 675-685.

Stiles, J. R. and T. M. Bartol. 2001. Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. *Computational neuroscience : realistic modeling for experimentalists*. E. d. Schutter. Boca Raton, Fla., CRC Press. pp. 87-127.

Stryer, L. 1988. *Biochemistry*.Third edition. New York: W.H.Freeman And Company.

Sundararaj, S., A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison and D. S. Wishart. 2004. "The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli." *Nucl. Acids Res.* 32(suppl_1), pp. D293-295.

Sutterlin, T., S. Huber, H. Dickhaus and N. Grabe. 2009. "Modeling multi-cellular behavior in epidermal tissue homeostasis via finite state machines in multi-agent systems." *Bioinformatics* 25(16), pp. 2057-2063.

SwarmWiki Contributors. 2009. "Tools for Agent-Based Modelling." SwarmWiki, Available at: http://www.swarm.org/index.php/Tools_for_Agent-Based_Modelling [Accessed: 14 July, 2009]

Szallasi, Z., V. Periwal and J. Stelling. 2006. On Modules and Modularity. *System modeling in cellular biology: from concepts to nuts and bolts*. Z. Szallasi, J. Stelling and V. Periwal. Cambridge, MA, MIT Press. pp. 41-50.

Szallasi, Z., J. Stelling and V. Periwal, Eds. (2006). *System Modeling in Cellular Biology From Concepts to Nuts and Bolts*. Cambridge, MA, MIT Press.

Takahashi, K., S. N. V. Arjunan and M. Tomita. 2005. "Space in systems biology of signaling pathways - towards intracellular molecular crowding in silico." *FEBS Letters* 579(8), pp. 1783-1788.

Takahashi, K., K. Yugi, K. Hashimoto, Y. Yamada, C. J. F. Pickett and M. Tomita. 2002. "Computational challenges in cell simulation: a software engineering approach." *Intelligent Systems* 17(5), pp. 64-71.

Terracotta, I. 2009. "Terracotta." Terracotta, Inc., Available at: http://www.terracotta.org/ [Accessed: 06 July, 2009]

Tham, M. 1998 - 2000. "Overview of mechanistic modelling techniques." University of Newcastle upon Tyne, Available at: http://lorien.ncl.ac.uk/ming/dynamics/modelling.pdf [Accessed: 10, May, 2010]

Theodoropoulos, G., Y. Zhang, D. Chen, R. Minson, S. J. Turner, W. Cai, Y. Xie and B. Logan. 2006. Large Scale Distributed Simulation on the Grid. *Sixth IEEE International Symposium on Cluster Computing and the Grid Workshops (CCGRIDW'06)*,16-19 May, Singapore,. pp. 63.

Thorne, B. C., A. M. Bailey and S. M. Peirce. 2007. "Combining experiments with multi-cell agent-based modeling to study biological tissue patterning." *Brief Bioinform* 8(4), pp. 245-257.

Tolle, D. P. and N. Le Novere. 2006. "Particle-Based Stochastic Simulation in Systems Biology." *Current Bioinformatics* 1, pp. 315-320.

Tomita, M. 2001. "Whole-cell simulation: a grand challenge of the 21st century." *Trends in Biotechnology* 19(6), pp. 205-210.

Tost, J. 2008. *Epigenetics.*Norfolk, UK: Caister Academic Press.

Turner, B. M. 2007. "Defining an epigenetic code." *Nat Cell Biol* 9(1), pp. 2-6.

van Gunsteren, W. F., D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hunenberger, M. A. Kastenholz, C. Ostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt and H. B. Yu. 2006. "Biomolecular modeling: Goals, problems, perspectives." *Angewandte Chemie International Edition* 45(25), pp. 4064-92.

Van Regenmortel, M. H. 2007. "The rational design of biological complexity: a deceptive metaphor." *Proteomics* 7(6), pp. 965-75.

Vinson, V. J. 2009. "Proteins in Motion." *Science* 324(5924), pp. 197.

Walker, D. C. and J. Southgate. 2009. "The virtual cell--a candidate co-ordinator for 'middle-out' modelling of biological systems." *Brief Bioinform*pp. bbp010.

Walker, D. C., J. Southgate, G. Hill, M. Holcombe, D. R. Hose, S. M. Wood, S. Mac Neil and R. H. Smallwood. "The epitheliome: agent-based modelling of the social behaviour of cells." *Biosystems* 76(1-3), pp. 89-100.

Wanner, B., A. Finney and M. Hucka. 2005. Modeling the E. coli cell: The need for computing, cooperation, and consortia. *Systems Biology: Definitions and Perspectives*. L. Alberghina and H. Westerhoff. Berlin Heidelberg, Springer. pp. 163-189.

Webb, K. and T. White. 2006. "Cell modeling with reusable agent-based formalisms." *Applied Intelligence* 24(2), pp. 169-181.

Westerhoff, H. and L. Alberghina. 2005. Systems Biology: Did we know it all along? *Systems Biology: Definitions and Perspectives*. L. Alberghina and H. Westerhoff. Berlin Heidelberg, Springer. pp. 3-9.

Westerhoff, H. V. and B. O. Palsson. 2004. "The evolution of molecular biology into systems biology." *Nat Biotech* 22(10), pp. 1249-1252.

Wikipedia Contributors. 2009a. "Feedback." Wikipedia, The Free Encyclopedia., Available at: http://en.wikipedia.org/w/index.php?title=Feedback&oldid=320909704 [Accessed: 11 November, 2009]

Wikipedia Contributors. 2009b. "List of software for molecular mechanics modeling." Wikipedia, The Free Encyclopedia., Available at: http://en.wikipedia.org/w/index.php?title=List_of_software_for_molecular_mechanics_modeling&oldid=287522548 [Accessed: 8 May 2009 15:21 UTC 2009]

Wikipedia Contributors. 2009c. "System dynamics." Wikipedia, The Free Encyclopedia., Available at: http://en.wikipedia.org/w/index.php?title=System_dynamics&oldid=301818810 [Accessed, 13 July 2009 06:31 UTC]

Wikipedia Contributors. 2010a. "Pareto efficiency." Wikipedia, The Free Encyclopedia., Available at: http://en.wikipedia.org/w/index.php?title=Pareto_efficiency&oldid=356274179 [Accessed: 18 April, 2010]

Wikipedia Contributors. 2010b. "Stigmergy." Wikipedia, The Free Encyclopedia., Available at: http://en.wikipedia.org/w/index.php?title=Stigmergy&oldid=350580603 [Accessed: 13 April, 2010]

Wikipedia Contributors. 2010c. "Swarm intelligence." Wikipedia, The Free Encyclopedia., Available at: http://en.wikipedia.org/w/index.php?title=Swarm_intelligence&oldid=389818877 [Accessed, 30 July 2010 13:40 UTC]

Wingreen, N. and D. Botstein. 2006. "Back to the future: education for systems-level biologists." *Nature Reviews Molecular Cell Biology* 7(11), pp. 829-832.

Wishart, D. S., R. Yang, D. Arndt, P. Tang and J. Cruz. 2005. "Dynamic cellular automata: an alternative approach to cellular simulation." *In Silico Biology* 5(2), pp. 139-61.

Wolfe, J. 2001. "Cellular Thermodynamics." *Encyclopedia of Life Sciences*, DOI: 10.1038/npg.els.0001363

Wolfe, J. 2002. "Cellular Thermodynamics." *Encyclopedia of Life Sciences*, pp. 1-11.

Wray, G. A. 2007. "The evolutionary significance of cis-regulatory mutations." *Nat Rev Genet* 8(3), pp. 206-216.