

# **INAUGURAL-DISSERTATION**

zur Erlangung der Doktorwürde der  
Naturwissenschaftlich-Mathematischen

Gesamtfakultät der

RUPRECHT-KARLS-UNIVERSITÄT

HEIDELBERG

vorgelegt von

Nurgazy Sulaimanov

aus Batken/Kirgisistan

Tag der mündlichen Prüfung: 2. Mai 2013



---

# Mathematical Modeling of Hepatitis C Virus Replication

1. Gutachter: Prof. Dr. Angela Stevens
2. Gutachter: Prof. Dr. Lars Kaderali



*Dedicated to my family and those who always support me*



# Abstract

At present, 170 million people are infected with Hepatitis C Virus, which is about 3 % of the world population. Currently available treatment is successful in only 50 % of treated patients. Further development of the treatment strategies requires a profound quantitative understanding of the viral lifecycle. This remains a major challenge, however, by combining the experimental and modeling approaches it has become possible to understand it quantitatively.

In this thesis, we develop a mathematical model to investigate the intracellular dynamics of HCV replication. In this model, two processes, the translation of viral proteins and the replication of viral genomes are considered. This model is established by using ordinary differential equations. The established model is then calibrated by using the time course data representing the viral plus-, minus-strand RNA and the polyprotein dynamics and the steady state data reflecting the ratios among the plus-, minus-strand RNA and the non-structural proteins. Subsequently, the model is validated using the independent measurements highlighting the replication deficient and synthesis inhibited HCV RNA dynamics. Furthermore, the model is used to analyze the observed difference in HCV RNA replication in the clonal Huh-7 cell lines. We demonstrate that this difference can be explained by the differential expression of the cellular host factor which involves in the replication of viral genomes. Using the model, a role of replication vesicles with respect to viral dynamics is analyzed. A sensitivity analysis is performed on the model parameters to reveal the crucial steps in the viral replication. This analysis shows that the processes in the replication vesicles are the most crucial for HCV replication. Finally, an identifiability analysis is performed to check whether the model parameters are sufficiently estimated from the measured data.



# Zusammenfassung

Heutzutage sind etwa 170 Millionen Menschen, d.h. 3% der Weltbevölkerung, mit Hepatitis C Virus infiziert. Die derzeit verfügbare Behandlung ist nur bei 50 % der behandelten Patienten erfolgreich. Die Weiterentwicklung von Behandlungsstrategien erfordert ein fundiertes quantitatives Verständnis des viralen Lebenszyklus. Dies ist eine große Herausforderung, aber durch die Kombination von experimentellen und Modellierungsansätzen ist es möglich geworden, den Lebenszyklus quantitativ zu verstehen.

In dieser Arbeit entwickeln wir ein mathematisches Modell, um die intrazelluläre Dynamik der HCV-Replikation zu untersuchen. In diesem Modell werden zwei Prozesse, die Translation der viralen Proteine und die Replikation des viralen Genoms, betrachtet. Das Modell wird durch die Verwendung gewöhnlicher Differentialgleichungen etabliert. Das Modell wird dann unter Verwendung experimenteller Zeitreihendaten der Dynamik von Plus-, Minusstrang-RNA und Polyprotein, sowie Messungen der Verhältnisse zwischen Plus- und Minus-Strang-RNA sowie der nicht-strukturellen Proteine im stationären Zustand, kalibriert. Anschließend wird das Modell mithilfe von unabhängigen Messungen der Dynamik von replikationsdefizienter bzw. Synthese-inhibierter HCV-RNA validiert. Darüber hinaus wird das Modell verwendet, um unterschiedliche Replikationsdynamik von HCV-RNA in verschiedenen Huh-7-Zelllinien zu analysieren. Wir zeigen, dass dieser Unterschied durch die differentielle Expression eines zellulären Wirtsfaktors, welcher an der Replikation des viralen Genoms beteiligt ist, erklärt werden kann. Mithilfe des Modells wird die Rolle der Replikationsvesikel in der viralen Dynamik analysiert. Um die entscheidenden Schritte in der viralen Replikation aufzudecken, eine Sensitivitätsanalyse bezüglich der Modellparameter wird durchgeführt. Diese Analyse zeigt, dass die Prozesse innerhalb der Replikationsvesikel den größten Einfluss auf die HCV-Replikation haben. Abschließend wird eine Identifizierbarkeitsanalyse durchgeführt, um zu überprüfen, ob die Modellparameter ausreichend genau auf Grundlage der gemessenen Daten geschätzt werden können.



# Acknowledgements

I would like to thank those people who helped me to realize and bring this work to the end. First of all, I am deeply grateful to my group leader Professor Dr. Lars Kaderali for offering this highly interesting research project, for his continuous support, guidance, encouragement and for the excellent work environment during my doctoral studies. I am grateful to my supervisor, Professor Dr. Angela Stevens for her continuous support, critique, helpful comments and discussions. I would like to thank our experimental collaborators Dr. Marco Binder and Professor Dr. Bartenschlager for the excellent and fruitful collaboration with support of experimental data, numerous help with biological questions, for always finding time to discuss the results. I would like to express my gratitudes to Dr. Johannes Schlöder for carefully reading the manuscript and his helpful comments on the optimization part of the thesis. I would like to thank Simon Lenz and Dr. Johannes Schlöder for their support with Parfit software and performing Multiple Shooting Algorithm and for their numerous discussions. Further, I am very thankful to my colleagues Dr. Narsis Kiani and Dr. Diana Clausnitzer for their critical reading of the thesis and useful comments. Also, I am very thankful to all my colleagues in Viroquant Research Group Modeling for their helpful discussions and support. I am very grateful for spending nice time with you while working on my project. I would also like to thank all my friends for their continuous support and encouragement. Finally, I want to thank my family for their love, support, encouragement and understanding.

Heidelberg, June 2012

Nurgazy Sulaimanov



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Contents</b>	<b>ix</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>Foreword</b>	<b>1</b>
<b>1 Introduction</b>	<b>5</b>
1.1 Hepatitis C Virus infection . . . . .	5
1.1.1 Acute infection . . . . .	5
1.1.2 Chronic infection . . . . .	6
1.2 Hepatitis C Virus lifecycle . . . . .	7
1.2.1 Model systems to study HCV lifecycle . . . . .	7
1.2.2 Genome organization . . . . .	8
1.2.3 HCV entry . . . . .	9
1.2.4 HCV RNA replication . . . . .	9
1.2.5 HCV assembly . . . . .	12
1.3 Differential equations . . . . .	13
1.3.1 Nonlinear systems . . . . .	13
1.3.2 Dynamical systems . . . . .	15
1.4 First replication model . . . . .	16
1.4.1 Model predictions . . . . .	18
1.4.2 Limitations of the model . . . . .	19

<b>2</b>	<b>Modeling HCV Replication</b>	<b>25</b>
2.1	Translation Process . . . . .	25
2.1.1	Processing of a transfected plus-strand RNA . . . . .	26
2.1.2	Translation . . . . .	26
2.2	Replication in replication vesicles (RV) . . . . .	29
2.3	Host factor assumption . . . . .	33
2.4	Complex model . . . . .	33
2.5	Experimentally determined parameter values . . . . .	35
<b>3</b>	<b>Model Calibration</b>	<b>39</b>
3.1	Inverse problem . . . . .	39
3.2	Parameter estimation . . . . .	41
3.2.1	Multiple shooting parametrization . . . . .	44
3.2.2	Generalized Gauss-Newton method . . . . .	46
3.2.3	Statistical analysis . . . . .	47
3.2.4	Software . . . . .	47
3.2.5	Genetic algorithm . . . . .	47
3.2.6	Genetic algorithm operators . . . . .	49
3.2.7	A simple genetic algorithm . . . . .	51
3.2.8	Optimization using genetic algorithm . . . . .	52
3.3	HCV replication dynamics . . . . .	55
3.3.1	Plus-strand RNA kinetics . . . . .	56
3.3.2	Minus-strand RNA kinetics . . . . .	56
3.3.3	Polyprotein kinetics . . . . .	57
3.3.4	Steady state observations . . . . .	58
<b>4</b>	<b>Model Validation and Predictions</b>	<b>61</b>
4.1	Model validation . . . . .	61
4.1.1	Replication deficient virus dynamics . . . . .	61
4.1.2	Initiation hampered HCV RNA dynamics . . . . .	64
4.2	Model predictions . . . . .	67
4.2.1	Permissiveness of HCV RNA replication in different cells . . . . .	67
4.2.2	Degradation inside the replication vesicles . . . . .	73
4.2.3	Steady state predictions . . . . .	76
4.2.4	Transfected plus-strand RNA degradation . . . . .	77
4.3	Sequential binding of ribosomes to HCV RNA . . . . .	79

---

<b>5</b>	<b>Sensitivity Analysis</b>	<b>81</b>
5.1	Local Sensitivity Analysis . . . . .	81
5.1.1	Results . . . . .	83
5.2	Global Sensitivity Analysis . . . . .	86
5.2.1	Extended Fourier Amplitude Test (eFAST) . . . . .	86
5.2.2	Results . . . . .	89
5.2.3	Local sensitivity versus global sensitivity . . . . .	94
<b>6</b>	<b>Identifiability Analysis</b>	<b>97</b>
6.1	Prior Identifiability . . . . .	98
6.2	Practical Identifiability . . . . .	99
6.3	Local Identifiability Analysis . . . . .	101
6.4	Results . . . . .	103
<b>7</b>	<b>Discussion</b>	<b>107</b>
7.1	Summary . . . . .	107
7.2	Comparison with other models . . . . .	113
7.3	Perspectives for future work . . . . .	114
7.3.1	Infection model . . . . .	114
7.3.2	Future application of the model . . . . .	115
	<b>Appendix A</b>	<b>117</b>
	<b>Appendix B</b>	<b>121</b>
	<b>Appendix C</b>	<b>129</b>
	<b>References</b>	<b>131</b>



# List of Figures

1.1	Genetic organization and polyprotein processing of HCV . . . . .	8
1.2	Hepatitis C Virus replication cycle . . . . .	12
1.3	First replication model . . . . .	22
1.4	Model fitting using the existing model with original parameter ranges	23
1.5	Model fitting using the existing model with the extended parameter ranges . . . . .	24
2.1	Mathematical model of Hepatitis C Virus replication cycle . . . . .	32
3.1	Multiple shooting approach . . . . .	45
3.2	Mutation in genetic algorithm . . . . .	50
3.3	Crossover in genetic algorithm . . . . .	51
3.4	Model calibration against experimental data . . . . .	59
4.1	Validation of the model dynamics using a replication deficient virus .	63
4.2	Model validation by the synthesis inhibited virus dynamics . . . . .	66
4.3	Host factor difference explains the cell permissiveness for HCV replication . . . . .	70
4.4	Effect of different factors on the HCV replication dynamics . . . . .	71
4.5	Correlation between virus components and cellular host factor amounts	72
4.6	Effect of protective property of the replication vesicles on the replication dynamics . . . . .	75
4.7	Steady state predictions . . . . .	77
4.8	Initial plus-strand RNA dynamics . . . . .	79
4.9	Replication dynamics in the case of sequential binding . . . . .	80
5.1	Local sensitivity analysis on the model parameters . . . . .	85
5.2	Parameter sensitivity at $t = 4$ hours . . . . .	92
5.3	Parameter sensitivity at $t = 72$ hours . . . . .	93

6.1 Correlation matrix representing the correlation between the individual parameters. . . . . 106

# List of Tables

1.1	Parameter estimates of the first replication model. . . . .	21
2.1	Definitions of the model parameters . . . . .	37
3.1	Estimated kinetic parameters and initial values using Multiple Shooting Algorithm. . . . .	53
3.2	Estimated kinetic parameters and initial values using Genetic Algorithm. . . . .	54
3.3	Values of genetic algorithm operators used in parameter estimation . . . . .	54
4.1	Parameter values in the case of inhibited HCV replication . . . . .	66
5.1	Parameter ranges for global sensitivity analysis . . . . .	95
6.1	Practical identifiability analysis on the model parameters. . . . .	105



# Foreword

Mathematical modeling plays a key role in providing a solid framework which enables us to build experiments and to generate hypotheses. The use of mathematics to model biological systems has a long history. As various models are used in an attempt to improve our understanding of a complicated phenomena, it is becoming clear that the more complex models are required to capture the rich variety of dynamics observed in natural systems. By complex models, one can imagine a large systems of differential equations which can be quite good at approximating observed behavior, unfortunately, they contain a large number of parameters which need to be estimated from the experimental data. At the same time, different models are able to produce qualitatively similar dynamics. The question is how to distinguish between them and how to determine which may be the relevant mechanism. Moreover, the model predictions depend on the estimated parameters and the model may lead to wrong predictions if the model parameters are not properly estimated. Therefore, it is becoming obvious that the close interplay between mathematical modeling and experiments is required to precisely analyze the biological processes.

Mathematical modeling to study the intracellular plus-strand RNA virus lifecycle was first applied to RNA bacteriophages,  $Q\beta$  [37]. This provided an important point for developing intracellular HCV replication models. Recently, Dahari *et al.* developed the first mathematical model to study HCV replication in Huh-7 cells [48]. This model was applied to study the mechanisms which explain the steady state dynamics based on the experimental data measured at steady state. Recent advances in HCV research and generation of new time course data for HCV lifecycle require a development of more complex and realistic models that enables us to study the host and virus interactions and to make proper predictions.

The thesis is focused on the differential equation model which based on mass action kinetics to describe the intracellular Hepatitis C Virus replication kinetics. In this work, we try to answer the question what are the underlying mechanisms which shape the HCV replication dynamics. How do the complex interactions between the

virus and the host determine the fate of viral replication?

The structure of this thesis is as follows:

**Chapter 1** introduces the biological background of Hepatitis C Virus. We give some latest facts about HCV disease and HCV lifecycle. We motivate the importance to understand HCV lifecycle in order to increase the drug efficiency against the disease. In the same chapter, we discuss the existing model for HCV replication and its advantages and disadvantages. We discuss why this model fails to explain the experimental data quantitatively and why it is important to develop a more comprehensive model.

**Chapter 2** focuses on the development of the new mathematical model for HCV replication. The model is defined in the form of ordinary differential equations and based on mass action kinetics. All individual steps of the model are discussed in details. We further discuss the model parameters which are experimentally determined.

**Chapter 3** discusses the model calibration process where the model is fitted to the experimental data. We discuss the formulation of the objective function and the optimization methods used in the model calibration. We also discuss the kinetic and steady state data which are used to verify the model efficiency. We show the importance of the cellular host factor's effect in explaining the highly dynamic replication process.

**Chapter 4** focuses on the validation of the established model and the predictions drawn from the model analysis. For the validation process, we use independent measurements which are not used in the model calibration and show that the model simulations are consistent with these data. Using the model, we examine the different HCV replication dynamics in the Huh7 cells and identify the factors which explain this observation. We further analyze the role of replication vesicles in shaping the viral replication dynamics.

In **Chapter 5**, we use the sensitivity analysis to determine the crucial steps in viral replication. A local sensitivity analysis is employed to check the sensitivity at the vicinity of the best fit values. For a more detailed analysis, a global sensitivity analysis is performed where the output effect is calculated by changing all parameters simultaneously.

**Chapter 6** discusses the identifiability of the model parameters. We discuss two types of identifiability analysis which are a practical identifiability analysis and a local version of the structural identifiability analysis. Using the practical identifiability analysis, we identify those parameters which cannot be properly estimated from the experimental data. The local identifiability analysis is performed at the best fit values which are obtained from the model calibration.

In **Chapter 7**, we conclude the thesis by giving a short overview of the results and further discuss the challenges and the future applications of the model.

In **Appendix A**, we present some mathematical formulations to improve the readability of the text. **Appendix B** provides some biological terminology which simplify the readability of the thesis. Finally, **Appendix C** provides the list of publications arising from this thesis.



# Chapter 1

## Introduction

### 1.1 Hepatitis C Virus infection

HCV is a small, single-stranded RNA virus which belongs to the *Flaviviridae* family of viruses, that cause other serious human infections, such as dengue and yellow fever [21]. Nowadays, HCV is a global health problem, causing 40% of chronic liver disease [40],[6]. Currently, approximately 170 million people are affected with HCV worldwide, comprising 3% of the global population [104], but depending on the country, the prevalence ranges from 0.1 to 12% (reviewed in [6]). According to the latest statistics, HCV infection is a leading cause of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma worldwide [97], [144], [138] and estimated to be the most common indication for a liver transplantation in most Europe countries, USA and Australia [5], [96], [80]. It is estimated that nearly 75% – 85% of HCV infected people are progressing to the chronic HCV infection [138], and at least 30% of carriers develop a progressive liver disease, including cirrhosis and hepatocellular carcinoma (HCC)<sup>1</sup> [145, 152]. HCV can be divided into 6 major genotypes (1 to 6) and 11 common subtypes (1a-c, 2a-c, 3a, 3b, 4a, 5a, and 6a) (reviewed in [6]). Geographically, HCV genotypes 1, 2, and 3 occur worldwide, whereas infection with HCV genotypes 4 and 5 occurs mainly in Africa, and HCV genotype 6 appears mainly in Asia [11].

#### 1.1.1 Acute infection

Acute infection occurs less frequently because the most of acutely infected people are asymptomatic [138]. According to one study, about 70 % to 80 % of acute infections

---

<sup>1</sup> The most common type of liver cancer.

were asymptomatic [88].

The best diagnosis for the acute infection are HCV RNA tests [6] and the viral RNA that can be detected in human blood within 1 to 3 weeks of exposure [10]. However, a beginning of the infection is frequently undetected since an icteric phase (actual phase) of typical hepatitis occurs in only 25% of acute infections [6]. Several studies suggested that the acute infection is very sensitive to a therapy than the established disease [151]. An interferon therapy performed by the German Acute Hepatitis C Therapy Group during the acute phase of infection, resulted in a remarkable SVR<sup>1</sup> of 98% [35], (reviewed in [151]). This study indicated that an early treatment of the patients with the acute hepatitis C can prevent a chronic infection. Nevertheless, it has been reported that in 50% of patients, the acute infection evolves into a chronic disease [81].

### 1.1.2 Chronic infection

The majority of HCV-infected individuals are unable to clear the virus during the acute phase and become chronically infected. Chronic hepatitis is determined as the persistence of HCV RNA in the blood for at least 6 months after the onset of infection [138]. During the first decade of infection, hepatitis C can progress very slowly without liver-specific symptoms or physiological signs [6]. Most patients experience liver related symptoms with the development of advanced liver disease 10 to 30 years after infection [152], [145], [6]. It was reported that the rate of chronic HCV infection is affected by many factors which include the age at time of infection, the gender, and the ethnicity (reviewed in [138]). The data of one study suggested that the people with HCV infection at younger age have chronic hepatitis C less frequently than those infected at older age [18]. The current results indicated that the different racial and ethnic groups have differences in the rate of disease progression and the response to treatment [138].

Cirrhosis occurs in about 10% to 15% of individuals with chronic HCV infection [104], [138] and higher levels of the alcohol consumption contributes to the development of this liver disease [138], [112], [111].

It was observed that the rate of chronic HCV infection in the patients with human immunodeficiency virus (HIV) infection, was higher than in the patients without HIV infection which shows a significance of immune response in the development of

---

<sup>1</sup>Sustained virological response is a best indicator of effective treatment and defined by the absence of detectable HCV RNA or lower limit of detection in the serum after the end of the treatment [1].

chronic hepatitis C [149], [138].

The best available treatment for HCV infection for many years was the monotherapy with interferon which produces a sustained response in  $< 20\%$  of patients [151]. The combination therapy with interferon- $\alpha$  and ribavirin made a significant improvement in the treatment [133], [151] and this treatment produced a sustained virus response (SVR) in about 40% of patients with chronic HCV infection. Later, a combination therapy using peginterferon instead standard interferon has been found to be superior [83, 42]. In spite of these developments, only about 50% of individuals with genotype 1 disease respond after the combination therapy for 48 weeks [151] which indicates that there is still a urgent need for more effective therapies.

## 1.2 Hepatitis C Virus lifecycle

Profound advances in understanding the important steps of the HCV lifecycle have been made in recent years, however, many steps remained poorly understood. Figure 1.2 depicts the HCV lifecycle which comprise a viral entry, a genome replication and a particle formation. Here, we will review the recent developments in the field of HCV lifecycle from Moradpour *et al.* [32], Bartenschlager *et al.*, [17] and Tellinghuisen *et al.* [148].

### 1.2.1 Model systems to study HCV lifecycle

HCV lifecycle has been studied using different model systems (reviewed in [32]). In 1999, Lohmann *et al.* [154] engineered a bicistronic subgenomic replicon system for HCV in Huh7 cells and later it has become the standard cell-based assay to study HCV replication. In this study, the authors reported an establishment of persistent HCV RNA replication in a human hepatoma cell line (Huh-7)<sup>1</sup> using a subgenomic genotype 1b (Con1) replicon. Later, the adaptive mutations in non-structural proteins were identified which enhance the efficiency of HCV replication [148], [16]. However, a full-length HCV genome with these adaptive mutations failed to produce an infectious virus [32]. Later, Date *et al.* [34] and Kato *et al.* [62] revealed an HCV genotype 2a strain (JFH-1)<sup>2</sup> capable of replication without any adaptation (reviewed in [148]). This genotype is capable of producing an infectious

---

<sup>1</sup>Huh-7 is a well differentiated hepatocyte derived cellular carcinoma cell line and an immortal cell line of epithelial-like tumorigenic cells that was originally taken from a liver tumor in a 57-year-old Japanese male in 1982 [2]

<sup>2</sup>A Japanese patient with rare case of acute fulminant hepatitis

virus when transfected into the Huh-7 cells [157] and this achievement led to study HCV lifecycle in vitro for the first time.

## 1.2.2 Genome organization

The HCV genome consists of a positive-sense, single-stranded RNA approximately 9600 nucleotides in length. It encodes a single ORF (open reading frame)<sup>1</sup>, that is flanked by 5'- and 3' - NCRs (non-coding regions) and the 5'-NCR contains an internal ribosome entry site (IRES) which is essential for cap-independent translation of the viral RNA and is composed of four domains numbered I to IV (Figure 1.1) [32]. It was reported that the domain I is not required for IRES activity, but the domains I and II are both essential for HCV RNA replication [32],[41]. The 3'-NCR is composed of three regions that essential for HCV replication. Translation of the HCV open reading frame yields a polyprotein precursor that is processed into the three structural (core, E1, E2) and six non-structural proteins (NS2, NS3A, NS4A, NS4B, NS5A, NS5B) (Figure 1.1) [32](For more information refer to Appendix B).

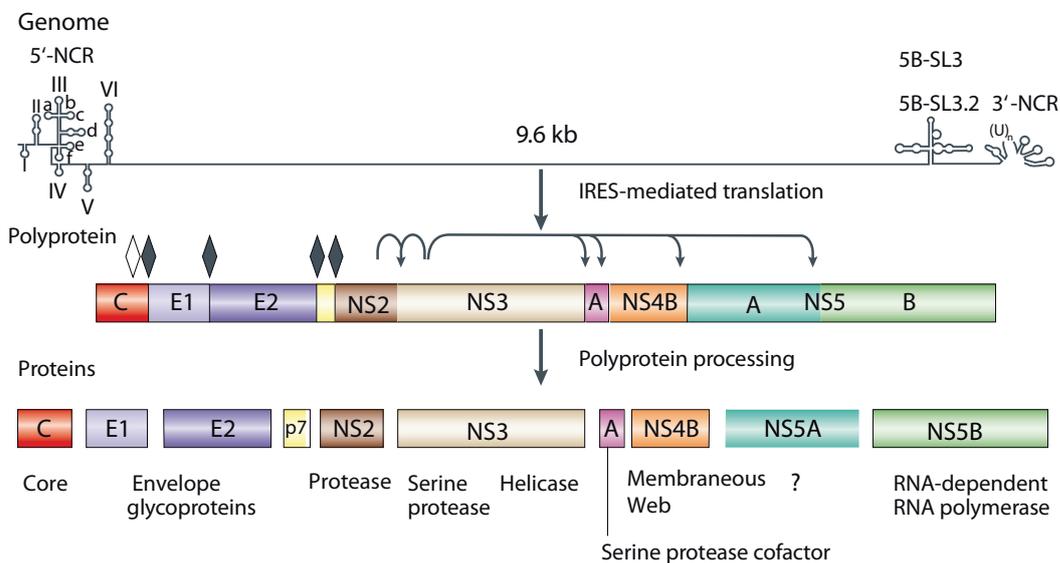


Figure 1.1: Genetic organization and polyprotein processing of hepatitis C virus (HCV) (Adapted from [98]). At the top, the HCV genome is depicted. HCV genome encodes for the polyprotein which is depicted at the middle. The polyprotein is cleaved into the core, structural and non-structural proteins which are depicted at the bottom.

<sup>1</sup>a portion of a gene which contains a sequence of bases that could potentially encode a protein

### 1.2.3 HCV entry

The first step in the virus life-cycle is the attachment of the virus particle to the host cell. This is done by a specific interaction between several receptors on the cell surface and a viral attachment protein on the surface of the particle. CD81 is a putative HCV receptor which interacts with the viral protein, *E2* [115], and it should present in a sufficient amount in the cell surface level for HCV entry to be efficient [67]. Low density lipoprotein receptor [7] is found to associate with viral particle via serving as a primary collector for further targeting to CD81 and additional receptor components (Figure 1.1) [32]. Another receptor, which binds to a viral protein is a scavenger receptor, SR-BI [134], which mediates a selective uptake of cholesteryl esters<sup>1</sup> from the high density lipoprotein (HDL) into the cellular membrane [148]. Recently, some studies [39] identified a claudin-1 (CLND1), a component of cellular tight junctions expressed at high levels in the liver cells, as an additional factor required for HCV entry, which functions at a late stage of cell entry [148]. HCV binding and internalization is initiated by the interaction between the HCV-associated lipoproteins and the lipoprotein receptors SR-BI and the virus enters the cell by endocytosis and fusion of the viral envelope which leads to the release of the viral nucleocapsid into the cytoplasm [27].

### 1.2.4 HCV RNA replication

Once the virus enters the cell, a translation of HCV RNA by the cellular ribosomes starts. Translation is the whole process by which the nucleotide sequence of an mRNA is used to order and to join the amino acids in a polypeptide chain [76]. HCV RNA acts as an mRNA to synthesize the polyprotein. The ribosomes physically move along an viral RNA, thereby catalyzing the assembly of amino acids into polypeptide chains [76]. The translation of the HCV open reading frame (ORF) generates a large polyprotein that undergoes co- and post-translationally cleavage events processed by both cellular and viral proteases into the structural and non-structural proteins (Figure 1.1) [148], [32]. The structural proteins are processed by the endoplasmic reticulum (ER) signal peptidase and the non-structural proteins are processed by the two viral proteases, NS2 – 3 protease and the NS3 – 4 serine protease [32]. Like other plus-strand RNA viruses, HCV replication starts with the synthesis of a complementary minus-strand RNA using the genome as a template

---

<sup>1</sup>Cholesterol esters form from the chemical reaction between an alcohol and an acid. They often used for transport and storage than free cholesterol. Read more: <http://www.livestrong.com/article/136499-definition-cholesterol-esters>

and after the minus-strand is synthesized, a genomic plus-strand RNA is produced from a minus-strand RNA template, where the both steps are catalyzed by the NS5B RdRp polymerase [32]. Co-precipitation and immunostaining studies have demonstrated that the newly synthesized HCV genome exists in association with a membraneous web [142], [137]. Expression of all structural and non-structural protein in the context of the entire HCV polyprotein has been observed to induce membrane changes [36]. This study revealed that HCV NS4B protein can induce the formation of the membraneous web alone.

Morphologically similar structures have been identified by the electron microscopy within the hepatocytes of HCV-infected chimpanzees [113]. All HCV non-structural proteins are found to be associated with the membraneous web in Huh-7 cells harboring subgenomic HCV replicons [121]. Experimental studies have shown that the viral RNA and proteins exist within detergent resistant, lipid-raft membranes [153, 128]. Furthermore, a membrane separation analysis has revealed that the HCV NS proteins exist in both in ER and Golgi apparatus, however, the viral RNA was found to primarily occur in the Golgi fraction [8].

What kind of role these membranes play in the HCV replication is not well understood. There are some speculations about the possible functions of them in the HCV RNA synthesis (reviewed in [32]). It has been hypothesized that they may physically support and organize RNA replication complex [79], compartmentalize the viral products [136], tether the viral RNA during unwinding, provide the lipid constituents important for replication and protect the viral RNA from double-strand RNA-mediated host defences.

Hepatitis C virus (HCV) encodes a large polyprotein and all viral proteins are produced in equimolar amounts regardless of their function [120]. Analysis on the Huh-7 cells harboring full-length HCV genomes or subgenomic replicons revealed about 1000-fold excess of HCV proteins over plus and minus-strand RNA [120]. To examine whether all nonstructural protein copies are involved in RNA synthesis, the authors isolated the active HCV replication complexes from the replicon cells and examined them for their content of viral RNA and proteins before and after a treatment with the protease and the nuclease. As the result, they demonstrated that almost the entire minus- and plus-strand RNA were resistant to the nuclease treatment, whereas  $< 5\%$  of the nonstructural proteins were protected from the protease digest but accounted for the full *in vitro*<sup>1</sup> replicase activity. In consequence, only a minor fraction of the non-structural proteins was actively involved in RNA

---

<sup>1</sup>In lab conditions, the components of organism have been isolated to allow more profound analysis than in whole organism. It is also known as “test tube experiments“.

synthesis. Comparing the nuclease-resistant viral RNA to the protease-resistant viral proteins, they estimated that the active HCV replicase complex consists of one minus-strand RNA, two to ten plus-strand RNA, and several hundred nonstructural protein copies. Since only the minor fraction of viral proteins participate in RNA synthesis, the role of other proteins is not well understood. It is believed that several factors may account for the production of the large amount of NS proteins that are not part of replication complexes [146]. They could be the by-products of polyprotein synthesis that produces large amounts of structural proteins; they might play active roles in counteracting innate immunity or cell cycle regulation; they may participate in the formation of infectious virions.

In the following, we will review some cellular factors which have been found to associate with HCV replication. As reviewed in [146], several studies identified that many cellular cofactors regulate HCV replication *in vitro*, including an miRNA [59, 109], the distinct host proteins that interact with nonstructural proteins, and the components of the RNA interference pathway [122]. Recent studies demonstrated that a cyclosporin A (CsA) inhibits HCV replication *in vitro* and a cyclophilin<sup>1</sup> B is a target of CsA action [60]. Recently, it has been reported that the cyclophilin A not the cyclophilin B is the key mediator of inhibition of HCV replication by the cyclosporin A [63], [162].

Some studies suggested that the hyperphosphorylation<sup>2</sup> of NS5A reduces an interaction with the human vesicle associated membrane protein A (hVAP-A) [38], [32]. hVAP-A has been reported to direct nonstructural proteins to lipid rafts that may be involved in viral replication [43]. In cell culture, HCV RNA replication is stimulated by saturated and mono-unsaturated fatty acids, and inhibited by polyunsaturated fatty acids or inhibition of fatty acid synthesis [32]. These results indicate that a membrane fluidity is important for the function of the replication complex. Recent studies [124] on human kinome using a siRNA screen revealed that a kinase, phosphatidylinositol-4 kinase III alpha (PI4KIII $\alpha$ ) are required for HCV replication. The authors reported that an enzymatic activity is critical for HCV replication and the viral non-structural protein NS5A was found to interact with PI4KIII $\alpha$  and stimulate its kinase activity and the absence of PI4KIII $\alpha$  activity induced a dramatic change in the structural morphology of the membranous HCV replication complex. The data of another study also supported the importance of this host

---

<sup>1</sup>Cyclophilins (CyPs) are a family of cellular enzymes possessing the peptidyl-prolyl isomerase activity [162]

<sup>2</sup>Hyperphosphorylation means when a biochemical with multiple phosphorylation sites is fully saturated

factor in regulating HCV replication [72].

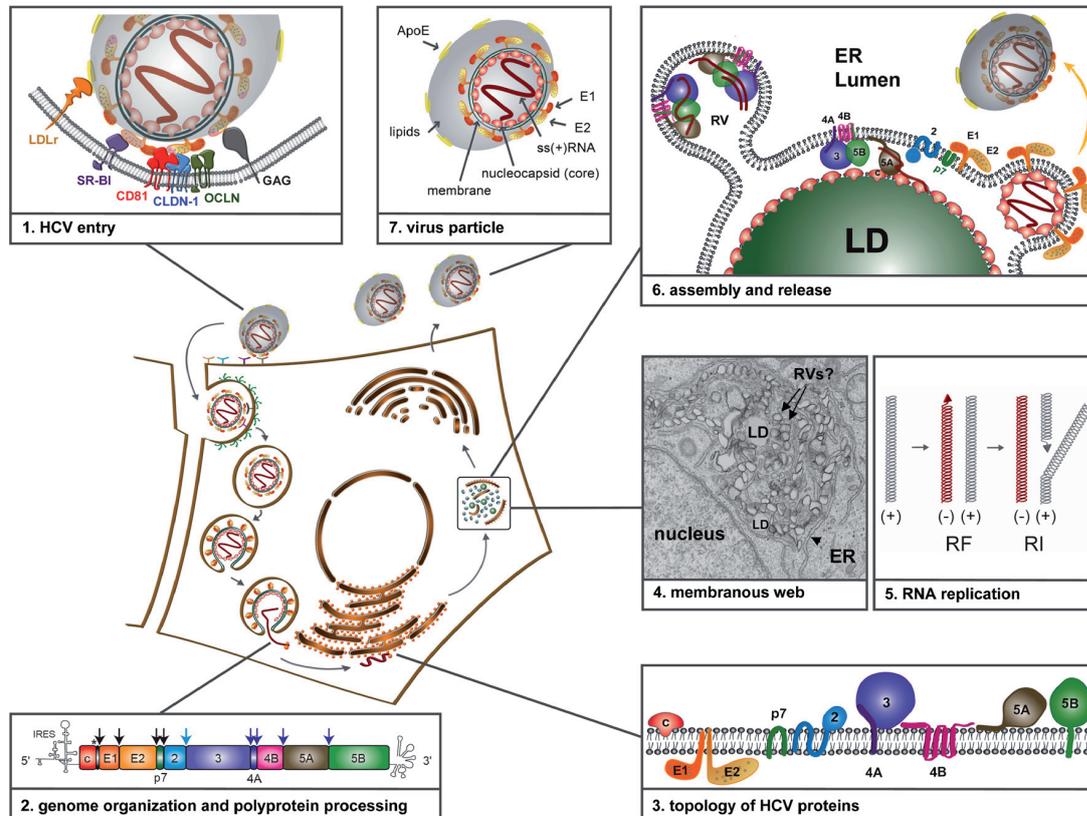


Figure 1.2: *Hepatitis C Virus replication cycle (Adapted from [15]). The process starts by an attachment of the HCV particle to the cell surface receptors. The virus enters the cell by a fusion where the viral genome is released into the cell. The viral genome is then translated by the host ribosomes into a viral polyprotein which is cleaved into a core, structural and nonstructural proteins. The viral genome then replicates within the membranous web by synthesizing a complementary minus-strand RNA using its genome as a template. The newly produced minus-strand RNA is served as a template for the synthesis of a new plus-strand RNA. Eventually, the viral genome participates in assembly and particle formation.*

### 1.2.5 HCV assembly

The last step of HCV lifecycle consists of packaging, assembly and particle release. Currently, a little information is available about the late steps of the viral lifecycle, which indicates that this area is still in its infancy. It has been shown that the mutations in NS5A, NS2 and NS3 block the production of infectious viruses which suggested a role of these non-structural proteins in the virus assembly and the particle release (reviewed in [146]). Cellular host factors like lipoproteins were

detected to participate in HCV assembly [54], [90] and HCV assembly was shown to occur on lipid droplets [94]. Virions is believed to form by budding into ER derived compartment and leave the cell through the secretory pathway [32].

## 1.3 Differential equations

A differential equation is an equation which involves an unknown function and one or more of its derivatives. Many of the fundamental laws of science are formulated in terms of differential equations. When modeling a complex process, one usually ends up with a complex model consisting of several equations which include non-linearity terms. In the following, we will start discussing differential equations by giving some facts about non-linear ordinary differential equations of the first order from [110].

### 1.3.1 Nonlinear systems

Let us consider an initial value problem given by the linear systems of ordinary differential equations

$$\dot{x} = Ax, \quad x(0) = x_0 \quad (1.1)$$

where  $x \in \mathbf{R}^n$ ,  $A \in \mathbf{R}^{n \times n}$  has a solution through each point  $x_0 \in \mathbf{R}^n$ , which is  $x(t) = e^{At}x_0$ . This solution is unique and defined for all  $t \in \mathbf{R}$ .

In the case of the non-linear equations, a unique solution exists only under certain conditions.

Lets consider a non-linear systems of ordinary differential equations given by

$$\dot{x} = f(x) \quad (1.2)$$

and for simplicity assume that they are autonomous. In general, this equation has a solution if the function  $f$  is continuous for all  $x \in \mathbf{R}^n$ . However, in contrast to the linear problem, a continuity of  $f$  is not sufficient to guarantee the uniqueness of the solution of (1.2).

**Definition 1.3.1.** Assume that  $f \in C(E)$  where  $E$  is an open subset of  $\mathbf{R}^n$ . Then  $x(t)$  is a solution of the differential equation (1.2) on an interval  $I$  if  $x(t)$  is differentiable on  $I$  and for all  $t \in I$ ,  $x(t) \in E$  and

$$x'(t) = f(x(t))$$

and given  $x_0 \in E$ ,  $x(t)$  is a solution of the initial value problem

$$\dot{x} = f(x), \quad x(t_0) = x_0$$

on an interval  $I$  if  $t_0 \in I$ ,  $x(t_0) = x_0$  and  $x(t)$  is a solution of (1.2) on the interval  $I$ .

For the existence and uniqueness of the solution, one needs to show that  $C^1$  functions are locally Lipschitz.

**Definition 1.3.2.** Let  $E$  be an open subset of  $\mathbf{R}^n$ . A function  $f : E \rightarrow \mathbf{R}^n$  is said to satisfy a Lipschitz condition on  $E$  if there is a positive constant  $K$  such that for all  $x, y \in E$

$$|f(x) - f(y)| \leq K|x - y|$$

The function  $f$  is said to be locally Lipschitz on  $E$  if for each point  $x_0 \in E$  there is an  $\epsilon$ -neighborhood of  $x_0$ ,  $N_\epsilon(x_0) \subset E$  and a constant  $K_0 > 0$  such that for all  $x, y \in N_\epsilon(x_0)$

$$|f(x) - f(y)| \leq K_0|x - y|$$

$\epsilon$ -neighborhood of  $x_0$  means an open ball of radius  $\epsilon$  given by

$$N_\epsilon(x_0) = \{x \in \mathbf{R}^n, |x - x_0| < \epsilon\}.$$

The following lemma indicates the conditions under which the function  $f$  is locally Lipschitz.

**Lemma 1.3.3.** Let  $E$  be an open subset of  $\mathbf{R}^n$  and let  $f : E \rightarrow \mathbf{R}^n$ . Then if  $f \in C^1(E)$ ,  $f$  is locally Lipschitz on  $E$ .

The following theorem shows the existence and uniqueness of the solution of nonlinear ordinary differential equation.

**Theorem (The Fundamental Existence-Uniqueness Theorem) 1.3.4.** Let  $E$  be an open subset of  $\mathbf{R}^n$  containing  $x_0$  and assume that  $f \in C^1(E)$ . Then there exists a positive  $a > 0$  such that the initial value problem

$$\dot{x}(t) = f(x(t)) \quad x(0) = x_0 \tag{1.3}$$

has a unique solution  $x(t)$  on the time interval  $[-a, a]$ .

This theorem shows that the initial value problem (1.3) has a unique solution defined on some interval. The following theorem shows that (1.3) has a unique solution on a maximal interval of existence  $(\alpha, \beta)$ .

**Theorem 1.3.5.** Let  $E$  be an open subset of  $\mathbf{R}^n$  and suppose that  $f \in C^1(E)$ . Then for each point  $x_0 \in E$ , there is a maximal open interval  $(\alpha, \beta)$  on which the initial value problem (1.3) has a unique solution,  $x(t)$ ; i.e., if the initial value problem has a solution  $y(t)$  on an interval  $I$  then  $I \subset (\alpha, \beta)$  and  $y(t) = x(t)$  for all  $t \in I$ .

### 1.3.2 Dynamical systems

We first start with defining dynamical systems. We assume that the temporal behavior of a system is given as a function  $\Phi(x(0), t)$  of the initial state  $x(0)$  and the time  $t$ . Moreover, we assume that  $x(t)$  satisfies an initial value problem of the form

$$\dot{x}(t) = f(x(t)), x \in E, x(0) = x_0 \quad (1.4)$$

where,  $E$  is an open subset of  $\mathbf{R}^n$  and the function  $f \in C^1(E \rightarrow \mathbf{R}^n)$  is a continuously differentiable function. That is, all partial derivatives of  $f$  with respect to  $x_j$ ,  $\partial f_i / \partial x_j$ , with  $i, j = 1, \dots, n$ , exist and are continuous. This guarantees the existence of a unique solution  $x(t)$  in a time interval  $[-a, a]$  (see theorem 1.3.4).

The solution of system (1.4) is related to a dynamical system, which provides a functional description of the solution of the system. Here, we will give a formal definition of a dynamical system and detailed information can be found in [110].

**Definition (Dynamical system) 1.3.6.** A dynamical system on  $E$  is a  $C^1$ - map

$$\Phi : \mathbf{R} \times E \rightarrow E \quad (1.5)$$

where  $E$  is an open subset of  $\mathbf{R}^n$ , and if  $\Phi_t(x) := \Phi(t, x)$  then  $\Phi_t$  satisfies

- $\Phi_0 = x$  for all  $x \in E$  and
- $\Phi_t \circ \Phi_s(x) = \Phi_{t+s}(x)$  for all  $s, t \in \mathbf{R}$  and  $x \in E$ .

$\Phi(t, x_0)$ , for fixed  $x_0 \in E$  corresponds to the solution of the initial value problem in Theorem 1.3.4. The first property in Definition 1.3.6 assures that the initial condition  $x(0) = x_0$  is fulfilled. The second property states that the evolution of

the system is uniquely determined for every  $t' \in \mathbf{R}$  if the state  $x$  of the system at any  $t$  is known. This means that the solution curves in the state space cannot intersect, since otherwise the time evolution of the system would not be unique at the intersection point. More clearly, we state the following relation between the dynamical system and the initial value problem: If  $\Phi(t, x)$  is a dynamical system defined on  $E \subseteq \mathbf{R}^n$ , then

$$f(x) = \frac{d}{dt}\Phi(t, x)|_{t=0} \quad (1.6)$$

Defines a  $C^1$ -vector field on  $E$ , and for each  $x_0 \in E$ ,  $\Phi(t, x_0)$  solves the initial value problem (1.4). Furthermore, a solution of the initial value problem (1.4) exists for every  $t \in \mathbf{R}$ , meaning that for each  $x_0 \in E$ , the maximal interval of existence of  $\Phi(t, x_0)$  is the time interval  $I(x_0) = (-\infty, \infty)$ . Thus, each dynamical system is related to a  $C^1$ -vector field  $f$ , and the dynamical system describes the solution set of the differential equation defined by this vector field. Conversely, given a differential equation  $\dot{x} = f(x)$ ,  $x \in E$  with  $f \in C^1(E)$  and  $E$  an open subset of  $\mathbf{R}^n$ , the solution  $\Phi(t, x_0)$  of the initial value problem (1.4) with  $x_0 \in E$  will be a dynamical system on  $E$  if and only if for all  $x_0 \in E$ , the maximal interval of existence  $I(x_0)$  of  $\Phi(t, x_0)$  is  $(-\infty, \infty)$ . In this case, we say that  $\Phi(t, x_0)$  is the dynamical system on  $E$  defined by the differential equation  $\dot{x} = f(x)$ .

## 1.4 First replication model

First replication model for HCV was developed by Dahari *et al.* [48]. Their model is based on the assumptions that the translation of HCV polyprotein occurs in the cytoplasm, HCV RNA synthesis occurs in the vesicular membrane structures (VMS)<sup>1</sup> and the strategy of replication involves a double-stranded RNA intermediate. Using the model, the authors tried to address the question if what kind of mechanisms or factors can explain the steady-state observations published in [120]. Here, we will discuss their results more in details. The model is based on the HCV replication scheme depicted in Figure 1.3. They used the biological facts that the translation of the HCV polyprotein takes place in the cytoplasm by involving host ribosomes and the replication of HCV RNA occurs in the vesicular membrane structures (VMS). The model equations describing the translation process in cytoplasm are described

---

<sup>1</sup> in our case, we use the definition of replication vesicles (RV).

as follows:

$$\begin{aligned}
\frac{dR_P^{cyt}}{dt} &= k_2 T_c + k_{Pout} R_P - k_1 R_{ibo} R_P^{cyt} - k_{Pin} R_P^{cyt} - \mu_P^{cyt} R_P^{cyt} \\
\frac{dT_c}{dt} &= k_1 R_{ibo} R_P^{cyt} - k_2 T_c - \mu_{T_c} T_c \\
\frac{dP}{dt} &= k_2 T_c - k_c P \\
\frac{dE^{cyt}}{dt} &= k_c P - k_{Ein} E^{cyt} - \mu_E^{cyt} E^{cyt}
\end{aligned} \tag{1.7}$$

where  $R_P^{cyt}$  represents the numbers of plus-strand RNA,  $T_c$  represents the numbers of translation complexes formed by an interaction of plus-strand RNA with the host ribosomes,  $P$  represents the numbers of viral polyprotein,  $E^{cyt}$  represents the numbers of viral polymerase, NS5B, formed as the result of polyprotein cleavage. Parameter  $k_1$  is the interaction rate of plus-strand RNA,  $R_P^{cyt}$ , with the host ribosomes,  $R_{ibo}$ , to form a translation complex,  $T_c$ , which degrades at rate  $\mu_{T_c}$ . The constant  $\mu_P^{cyt}$  is the degradation rate of free plus-strand RNA.  $R_{ibo}$  represents a complex of several ribosomes that interacts with  $R_P^{cyt}$  to initiate the translation. To keep the model simple, they assumed that 10 ribosomes simultaneously translate the same HCV mRNA. The viral polyprotein,  $P$ , is translated by the effective rate  $k_2$  and the ribosomes involved in translation dissociate when the synthesis of polyprotein is complete. Newly formed polyprotein is cleaved at rate  $k_c$  into several structural and non-structural proteins. One of the non-structural proteins NS5B,  $E^{cyt}$ , is important for HCV replication and degrade at rate  $\mu_E^{cyt}$ . Furthermore, NS5B molecules are transported into the replication vesicles (RV) at rate  $k_{Ein}$ .  $R_{ibo}^{tot}$  is a fraction of the total pool of cellular ribosomes and the number of free ribosome complexes in translation is given by:  $R_{ibo} = R_{ibo}^{tot} - T_c$ . Once the translation is complete, the plus-strand RNA,  $R_P^{cyt}$ , reappear at rate  $k_2$ . Some fraction of free plus-strand RNA is lost due to a transport from cytoplasm into VMS by rate  $k_{Pin}$  and is gained by a transport out at rate  $k_{Pout}$ . The equations that describe the replication kinetics

are:

$$\begin{aligned}
\frac{dR_P}{dt} &= -k_3 R_P E + k_{4p} R_{Ids} + k_{Pin} R_P^{cyl} - (k_{Pout} + \mu_P^{cyl}) R_P \\
\frac{dR_{ds}}{dt} &= k_{4m} R_{Ip} + k_{4p} R_{Ids} - k_5 R_{ds} E - \mu_{ds} R_{ds} \\
\frac{dE}{dt} &= k_{Ein} E^{cyl} + k_{4m} R_{Ip} + k_{4p} R_{Ids} - k_3 R_P E - k_5 R_{ds} E - \mu_E E \\
\frac{dR_{Ip}}{dt} &= k_3 R_P E - k_{4m} R_{Ip} - \mu_{Ip} R_{Ip} \\
\frac{dR_{Ids}}{dt} &= k_5 R_{ds} E - k_{4p} R_{Ids} - \mu_{Ids} R_{Ids}
\end{aligned} \tag{1.8}$$

$R_P$ ,  $R_{ds}$  and  $E$  represent the numbers of plus-strand RNA, double-stranded RNA (dsRNA), and HCV polymerase complexes, respectively. The numbers of plus-strand RNA and dsRNA replicative intermediate complexes given by  $R_{Ip}$  and  $R_{Ids}$ , respectively. The intermediate complex,  $R_{Ip}$ , which is formed at rate  $k_3$  and degrades at rate  $\mu_{Ip}$ , comprises a plus-strand RNA that serves as a template for the newly synthesized minus-strand RNA, the polymerase which plays a machinery role in replication, and the nascent complementary minus-strand RNA. The  $R_{Ids}$ , that is formed at rate  $k_5$  and degrades at rate  $\mu_{Ids}$ , is composed of dsRNA which contains the minus-strand RNA that serves as a template for the newly synthesized plus-strand RNA, the polymerase  $E$ , and the nascent complementary plus-strand RNA,  $R_P$ . It was assumed that once the synthesis of minus-strand RNA has ended, the replication complex,  $R_{Ip}$ , immediately dissociates into the dsRNA,  $R_{ds}$ , and the replication machinery  $E$  at rate  $k_{4m}$ . The same strategy is also applied for minus-strand replication complex  $R_{Ids}$ . So, when the synthesis of the newly produced RNA,  $R_P$ , has ended, the replication complex,  $R_{Ids}$ , immediately dissociates into the new plus-strand RNA, the polymerase  $E$ , and the dsRNA at rate  $k_{4p}$ . Both the free plus-strand RNA and the dsRNA degrade with rate  $\mu_p$  and  $\mu_{ds}$ , respectively. Additionally, the polymerase complex,  $E$ , degrades at rate  $\mu_E$ .

Importantly, any activity of the cellular factors that could limit the replication process was neglected by assuming that they are abundant.

### 1.4.1 Model predictions

In the following, we will summarize the results obtained by the replication model from [48]. The experimental data which will be discussed in this Subsection are from [120] and a short overview about the details is given in Section 3.3.

A model analysis showed that the main factor which significantly affects the number of NS5B polymerase molecules is the number of ribosomes. The results demonstrated that about 500 to 1000 ribosomes complexes are needed for the synthesis of a million NS5B molecules at steady-state.

Further, the model is able to predict 1 : 1 ratio of the total plus-strand RNA in cytoplasm to the total plus-strand RNA in RVs (VMS), by assuming  $k_{Pin}/k_{Pout} = 1$  which means an equal transport of plus-strand RNA into and out of the RVs. However, for the rest of analysis, they set  $k_{Pin}/k_{Pout} = 0.2$  and  $\mu_{Ids}/k_{Ein} = 10^4$ , because it didn't affect 1 : 1 ratio and it was needed to explain further observations.

Further analysis was done on the asymmetry levels of plus-strand and minus-strand RNA. In [120], it was reported that the asymmetry level reflects that the ratio of plus-strand to minus-strand RNA is about 10 : 1. To obtain an approximate ratio, they needed to assume that the formation rate of  $R_{Ids}$  is about 200-fold faster than the formation rate of  $R_{Ip}$ . Here  $R_{Ids}$  and  $R_{Ip}$  are the minus and plus-strand replicative intermediate complexes, respectively. In terms of kinetic rates, it gives  $k_5/k_3 \approx 200$ . The rate constant  $k_3$  varies between 0.004 and 0.02 h<sup>-1</sup> molecules<sup>-1</sup>. This large ratio also explains the observed 6 : 1 ratio of plus-strand to minus-strand RNA within the RVs [120]. This analysis explains how this asymmetry between the plus-strand and the minus-strand arises.

Further analysis was performed on the level of plus-strand RNA at steady-state. They reported that  $k_{Ein}$  and  $\mu_P^{cyt}$  are the parameters that mostly affect the steady-state level of total plus-strand RNA. The model analysis showed that the ranges of  $\mu_P^{cyt}$  (0.06 to 15 h<sup>-1</sup>) and  $k_{Ein}$  ( $3.8 \times 10^{-6}$  to  $6.0 \times 10^{-5}$  h<sup>-1</sup>) allow a steady state with the observed plus-strand RNA level.

Moreover, they used the model to analyze the time to reach the steady state. After the ratios were ensured, the remaining parameters [ $k_c, k_1, k_3, \mu_E, \mu_{T_c}$  and  $R_P^{cyt}(0)$ ] were adjusted to allow the system to reach the steady-state at about 48 hours. The analysis revealed that by keeping a large ratio of  $k_5/k_3$ , both the plus- and minus-strand RNA attain the steady state at 48 hours, when  $R_P^{cyt}(0) \approx 500$  copies/cell,  $\mu_{T_c} < 0.02$  h<sup>-1</sup> and  $k_1 \approx 80$  h<sup>-1</sup>. Moreover, the model analysis showed that there is an inverse correlation between  $R_P^{cyt}(0)$  and time to reach the steady state.

## 1.4.2 Limitations of the model

Due to a lack of experimental data on the numbers of plus-strands at the time of transfection and the lack of detailed kinetic information about the growth rate of plus-strand RNA, minus-strand RNA and proteins from transfection to steady

state, they could not estimate the parameter values precisely. As we discussed in Subsection 1.4.1, the model parameters were adjusted by using the steady-state data, whereas for a proper estimation the kinetic data is required. At a first glance, we examined whether the model calibrated with the steady state data can explain the kinetic measurements (for kinetic data see Section 3.3). We fitted the total plus-strand RNA, the minus-strand RNA and the polyprotein (luciferase) to the corresponding kinetic data, which highlight the replication dynamics from transfection to 72 hours. Initially, we restricted the model calibration to the original parameter ranges (Table 1.1). As a result, we have seen that the model is not able to capture the replication dynamics in the plus-, minus-strand RNA and the polyprotein (Figure 1.4). As we have seen from the parameter estimation by keeping them in the original ranges, it was actually impossible to fit the model to the data.

Next, we tried to relax the parameter ranges to check if the model is able to fit. The results show that there is a significant improvement, but still the model fails to explain the dynamics quantitatively (Figure 1.5). Figure 1.5 depicts several fits performed with the different scale factors. For more information about the scale factor, we refer to Section 3.3. This plot reflects how the different scale factor values change the way of the fitting, which mostly affects the dynamics of plus- and minus-strand RNAs, however, it does not affect the fit at the polyprotein curve initially. Additionally, the initial phase of polyprotein is overestimated in all cases. Theoretically, this issue could be resolved by either decreasing  $k_1$  or increasing  $\mu_P^{cyt}$ , however within the given ranges, tuning these parameters is insufficient to decrease the level of polyprotein initially and it leads to underestimation of a later phase of polyprotein. For small values of the scale factor, the prediction for the minus-strand RNA and as well as plus-strand RNA after 8 hours is underestimated (Figure 1.5 A). Obviously, the polyprotein prediction is overestimated comparing to the observation. When the scale factor gradually increases, the quality of the prediction for the minus-strand RNA and the later phase of plus-strand RNA is slightly improved, whereas the initial phase of the latter one is getting worse (Figure 1.5 B, C, D). Thereby, within the parameter ranges given in [48], it was not possible to find any proper fit. These results demonstrated that the model lacks some vitally important reactions and indicates a need for a more comprehensive model to make quantitative predictions.

Table 1.1: Parameter estimates of the first replication model [48].

Parameter name	Parameter definition	Parameter values
$k_1$	$T_c$ formation	1-100 ( $\text{h}^{-1}$ molecule $^{-1}$ )
$k_2$	polyprotein translation	100* ( $\text{h}^{-1}$ )
$k_c$	polyprotein cleavage	0.2-1 ( $\text{h}^{-1}$ )
$k_{Pin}$	$R_P^{cyt}$ transport into RV	0.2 ( $\text{h}^{-1}$ )
$k_{Pout}$	$R_P$ transport into cytoplasm	0.2 ( $\text{h}^{-1}$ )
$k_{Ein}$	$E^{cyt}$ transport into RV	4e-6 to 4e-5 ( $\text{h}^{-1}$ )
$k_{Am}$	$R_{ds}$ synthesis	1.7* ( $\text{h}^{-1}$ )
$k_{Ap}$	$R_P$ synthesis	1.7* ( $\text{h}^{-1}$ )
$k_5$	$R_{Ids}$ formation	$k_5/k_3 = 200$ ( $\text{h}^{-1}$ molecule $^{-1}$ )
$k_3$	$R_{Ip}$ formation	0.001-0.02 ( $\text{h}^{-1}$ molecule $^{-1}$ )
$\mu_P^{cyt}$	$R_P^{cyt}$ degradation	0.06-15* ( $\text{h}^{-1}$ )
$\mu_{T_c}$	$T_c$ degradation	0.001-0.02 ( $\text{h}^{-1}$ )
$\mu_E^{cyt}$	$E^{cyt}$ degradation	0.06* ( $\text{h}^{-1}$ )
$\mu_{Ip}$	$R_{Ip}$ degradation	0.01-0.06* ( $\text{h}^{-1}$ )
$\mu_{ds}$	$R_{ds}$ degradation	0.06* ( $\text{h}^{-1}$ )
$\mu_E$	$E$ degradation	0.001-0.06 ( $\text{h}^{-1}$ )
$\mu_{Ids}$	$R_{Ids}$ degradation	$\mu_{Ids}/k_{Ein} = 10^4$ ( $\text{h}^{-1}$ )
$\mu_P$	$R_P$ degradation	0.07* ( $\text{h}^{-1}$ )
$R_{ibo}^{tot}$	ribosome complexes	500-1000 molecules
$R_P^{cyt}(0)$	transfected plus-strand RNA	10-1000 molecules

\*, from literature

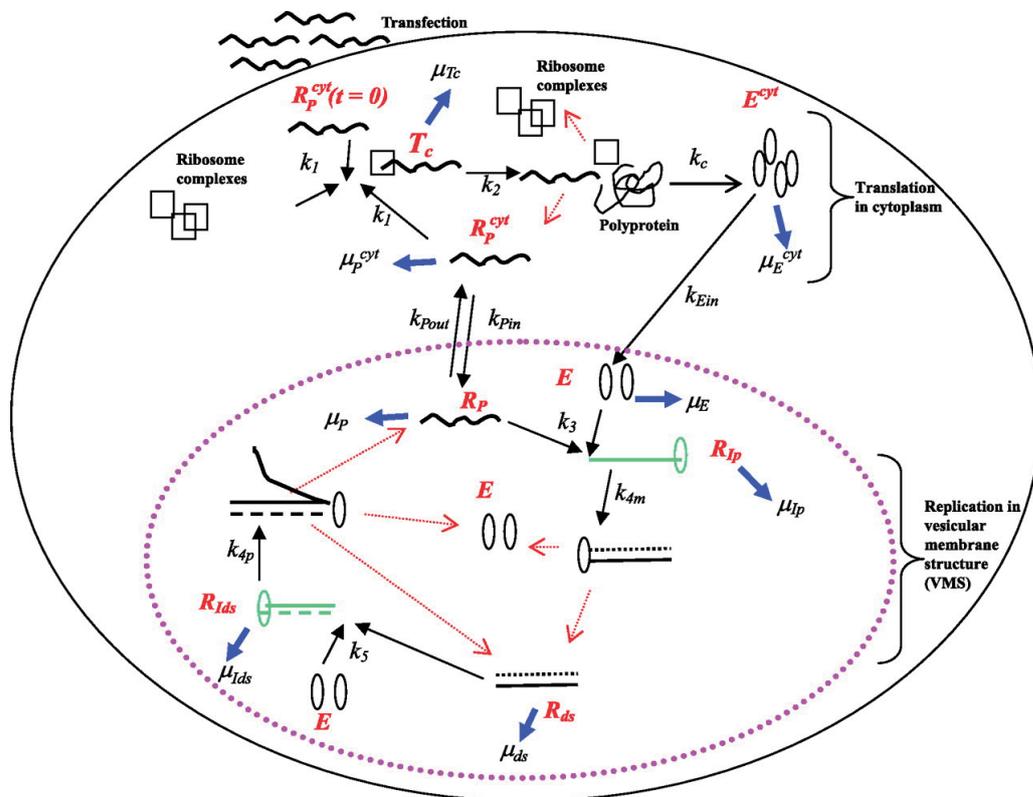


Figure 1.3: Model for Hepatitis C Virus replication cycle (Adapted from [48]). See text for details.

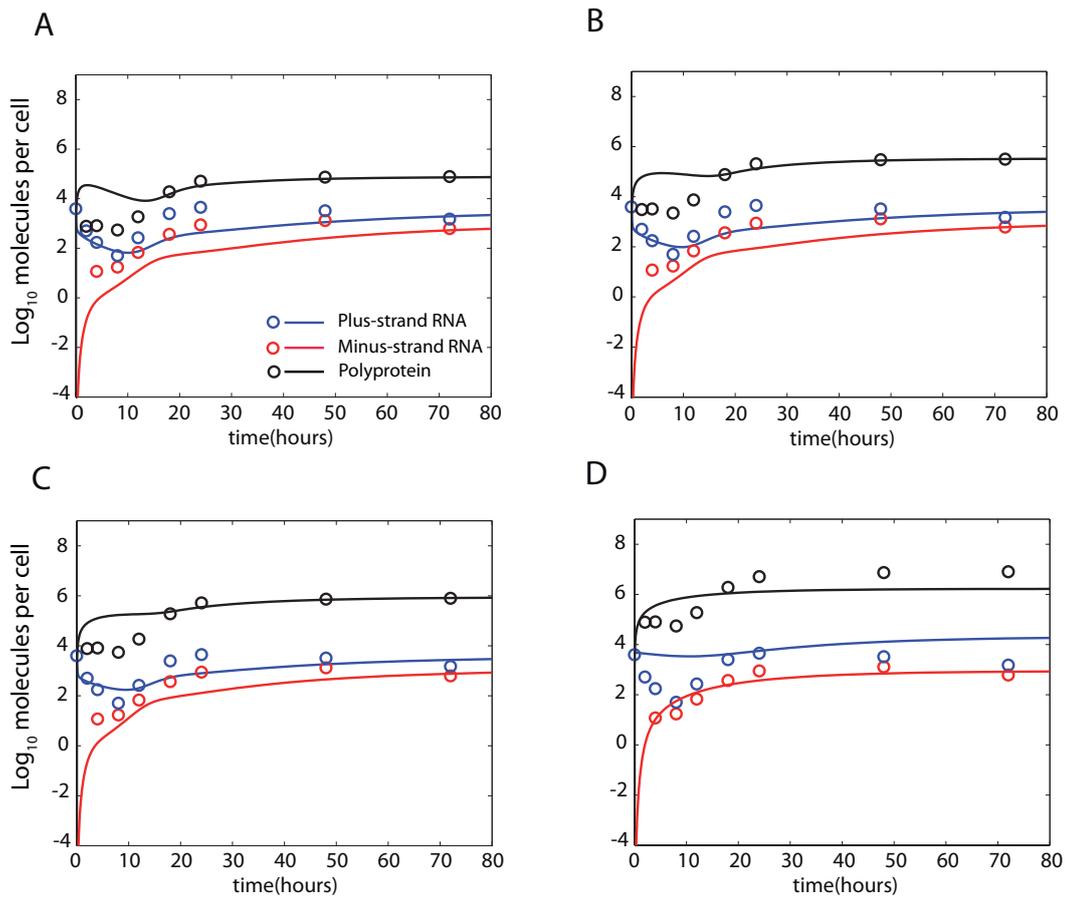


Figure 1.4: The fitting with the original model within the parameter ranges given in Table 1.1 with respect to the different scale factors. Experimental data for numbers of the total plus-, minus-strand RNAs and the polyprotein are denoted by blue, red and black circles, respectively (obtained from high permissive Huh7 Lunet cells). The data points indicate mean values. The corresponding model simulations are denoted by red, blue and black curves, respectively. **(A)** scale factor = 0.001, **(B)** scale factor = 0.004, **(C)** scale factor = 0.01, **(D)** scale factor = 0.1

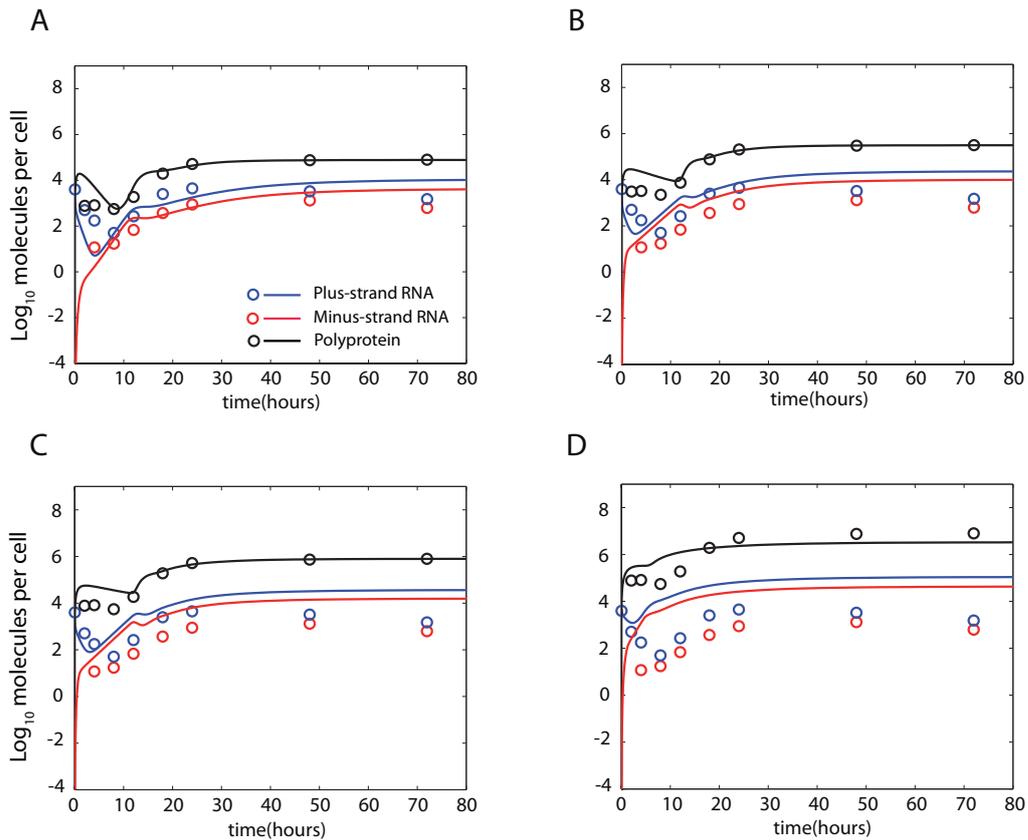


Figure 1.5: The fitting with the original model using the extended ranges with respect to the different scale factors. Experimental data for numbers of the total plus-, minus-strand RNAs and the polyprotein are denoted by blue, red and black circles, respectively (obtained from high permissive Huh7 Lunet cells). The data points indicate mean values. The corresponding model simulations are denoted by red, blue and black curves, respectively. (A) scale factor = 0.001, (B) scale factor = 0.004, (C) scale factor = 0.01, (D) scale factor = 0.1

# Chapter 2

## Modeling HCV Replication

As we already discussed in Subsection 1.4.2, the model developed by Dahari *et al.* [48] can reproduce the steady state data, but failed to explain the kinetic data. The measurements on HCV kinetics demonstrated that the initial stage of replication is highly dynamic. Additionally, these measurements revealed significantly different replication courses in the Huh7 cell clones. We found out that this model failed to even explain the individual measurements performed in these cell lines. All these indicate a strong need for a new model which can elucidate the highly dynamic replication kinetics and predict a source of the different replication dynamics in Huh7 cell lines. In this Chapter, we will discuss the development of a new replication model using mass action kinetics<sup>1</sup>.

### 2.1 Translation Process

Translation is a process by which the nucleotide sequence of an mRNA is used to order and to join the amino acids in a polypeptide chain. In this process, HCV RNA acts as an mRNA to synthesize the polyprotein [76]. Ribosomes physically move along a viral RNA, thereby catalyzing the assembly of amino acids into polypeptide chains. Polypeptide also known as polyprotein is cleaved into a core and several structural and non-structural proteins. In a modeling process, we ignore the role of core and structural proteins, because in the experimental studies the regions that encode for these proteins were removed so that the process can be explained only with non-structural proteins and a viral RNA. Importantly, the whole process takes place in the cellular cytoplasm.

---

<sup>1</sup>for more information see Appendix A

### 2.1.1 Processing of a transfected plus-strand RNA

In the transfection experiments, a large amount of viral genome is transfected into the cell. How effectively these transfected genomes can replicate is not known. It was observed that a high variation in HCV replication may indicate that a significant fraction of viral genomes appear to be defective [85], [17]. It also may be explained from the fact that in vitro conditions, the viral genome lacking a coat protein is used. It is believed that viral genome without the coat protein is more likely to degrade faster than the viral genome which enters the cell in a natural way. Simply, viral genomes which include defects do not contribute to the replication, and those which are not defected contribute. Further, it might be that an adaptation to the cell environment affect the replication process causing defects in the viral genome. Here, we include an initial step which accounts for this kind of effects.



where, the newly transfected plus-strand RNA,  $R_P^{unp}$ , is assumed to be unprocessed, meaning that it comprises all plus-strands which are capable of translating and those which are not. This unprocessed free plus-strand RNA process at rate  $k_0$  to the processed free plus-strand RNA,  $R_P^{cyt}$  and degrades at an effective rate,  $\mu_P^{unp}$ . Using mass action kinetics, we can convert this reaction to the following equation:

$$\frac{dR_P^{unp}}{dt} = -k_0 R_P^{unp} - \mu_P^{unp} R_P^{unp} \quad (2.2)$$

the solution of which nothing else but a typical exponential function:

$$R_P^{unp}(t) = R_P^{unp}(0)e^{-(k_0 + \mu_P^{unp})t} \quad (2.3)$$

where,  $R_P^{unp}(0)$  is an initial number of transfected plus-strand RNA molecules. As we see, the processing step is independent of other processes and can be computed directly. The role of this additional step will be discussed in Chapter 4.

### 2.1.2 Translation

Once processed, the plus-strand RNA can directly interact with the cellular ribosomes, a complex molecule machine composed of both RNA and protein, to form a translation complex:



This reaction happens at rate  $k_1$  and the viral plus-strand RNA degrades at rate  $\mu_P^{cyt}$ .  $T_c$  is a translation complex that comprises plus-strand RNAs and ribosome complexes,  $R_{ibo}$ , and degrades at rate  $\mu_{T_c}$ . A protein bound RNAs were found to be more stable against the host nuclease, therefore, we assume  $\mu_{T_c} < \mu_P^{cyt}$ . This condition indicates that free the plus-strand RNA,  $R_P^{cyt}$  degrade faster than the plus-strand RNA within the translation complex,  $T_c$ . During the translation, the ribosome assembles and links together amino acids in the precise order dictated by the viral plus-strand RNA (serves as a messenger RNA), which leads to the final product, a polyprotein. This process can be simply modelled as:



The synthesis of the polyprotein,  $P$  is done at rate  $k_2$ . Once the polyprotein production is complete, the translation complex dissociates into the plus-strand,  $R_P^{cyt}$  and the ribosome complex,  $R_{ibo}$  at the same rate  $k_2$ .

The viral polyprotein,  $P$ , is then cleaved by host protease into several non-structural (NS) proteins. The NS proteins were found to play a major role in the replication and the role of them are being extensively studied. The key protein which plays a major role in RNA synthesis is an NS5B protein having a polymerase activity. This polymerase can directly copy a complementary RNA from a plus-strand RNA template. Therefore, to keep the model simple, we address the role of this protein in our model. In reaction terms,



The cleavage of polyprotein,  $P$ , is done at rate  $k_c$  or simply, we can say,  $E^{cyt}$  is produced from  $P$  at rate  $k_c$ . In turn, the polymerase,  $E^{cyt}$  degrades at rate  $\mu_E^{cyt}$ .

Once the polymerase is cleaved from the polyprotein, it can be used in the formation of so called replication complexes inside the replication vesicles (RV) (vesicular membrane structures, (VMS)), within which the replication undergoes. Recent experimental studies have revealed that the viral replication happens inside these structures. The formation of these structures (RV) can be induced by viral proteins. Viral plus-strand RNA can only enter the RV together with the viral protein NS5B, translated from itself. This property is called a cis-process where the viral plus-strand RNA is first translated to produce the polyprotein, and after can be used for a complementary RNA synthesis.

First, we assume that NS4B induces the formation of the replication vesicles. As

a first step after RV has been formed, we assume the formation of the intermediate replicative complex, which we define as  $R_{Ip}$ , which contains a plus-strand RNA and a polymerase. This complex is called a plus-strand intermediate replicative complex. Here, we consider that the formations of the replication vesicles and the replication complexes are independent processes, therefore, we assume only the effect of NS5B in the process. The formation of RV happens with physical conformation of cellular components which might participate in this process. Experimental studies have revealed several factors that interact with viral components at the replication stage of viral RNA. Here, we assume that a putative cellular factor participates at the formation of RV. Adding this term to the model, we obtain:



This represents the formation of the replication complex from the plus-strand RNA, the polymerase and the cellular host factor. Here, assume that the ribosomes within the translation complexes dissociate at the same rate  $k_{Pin}$ .

Additionally, in order to describe the dynamics of polyprotein marker (luciferase), we include an additional reaction:



This polyprotein marker is used to quantify the polyprotein dynamics and we assume that it degrades at rate  $\mu_L$ .

Combining all processes that have been described above, we can write the equations for the translation process:

$$\begin{aligned}
\frac{dR_P^{unp}}{dt} &= -k_0 R_P^{unp} - \mu_P^{unp} R_P^{unp} \\
\frac{dR_P^{cyt}}{dt} &= k_0 R_P^{unp} + k_2 T_c - k_1 R_{ibo} R_P^{cyt} - \mu_P^{cyt} R_P^{cyt} \\
\frac{dT_c}{dt} &= k_1 R_{ibo} R_P^{cyt} - k_2 T_c - k_{Pin} T_c H_F E^{cyt} - \mu_{T_c} T_c \\
\frac{dP}{dt} &= k_2 T_c - k_c P \\
\frac{dE^{cyt}}{dt} &= k_c P - k_{Pin} T_c H_F E^{cyt} - \mu_E^{cyt} E^{cyt} \\
\frac{dH_F}{dt} &= -k_{Pin} T_c H_F E^{cyt} + k_{4m} R_{Ip} \\
\frac{dR_{ibo}}{dt} &= -k_1 R_{ibo} R_P^{cyt} + k_2 T_c + \mu_{T_c} T_c + k_{Pin} T_c H_F E^{cyt} \\
\frac{dL}{dt} &= k_2 T_c - \mu_L L
\end{aligned} \tag{2.9}$$

## 2.2 Replication in replication vesicles (RV)

Recent studies revealed that HCV amplifies its RNA inside a membranous web. This web harbors the viral RNA and the polymerase from being degraded by host nuclease and protease. We assume that, the already available plus-strand RNA serves as a template for a synthesis of new minus-strand RNA, which in turn, serves as a template for the formation of new plus-strand RNA.

The family of viruses to which HCV belongs, has a feature of forming the intermediate replication complexes which initiate the production of new viral RNA.

The mechanism of HCV RNA replication inside the RV is not clearly understood. Most well known fact is that the viral RNA and the polymerase participate in this process. Once RV is formed, the first component which serves as a starting point is the plus-strand intermediate replicative complex. The polymerase within this complex can copy a complementary-strand, which is the minus-strand RNA, from the plus-strand template. This process can simply be written as:



Here, the minus-strand RNA,  $R_{ds}$  is synthesized at rate  $k_{4m}$  from the intermediate complex,  $R_{Ip}$ . Once the production of  $R_{ds}$  is complete,  $R_{Ip}$  dissociates into the plus-

strand RNA,  $R_{ds}$ , the polymerase,  $E$  and the cellular factor,  $H_F$ . Here, we assume that the new minus-strand RNA exists in a double-stranded form. Therefore,  $R_{ds}$  consists of both plus-strand and minus-strand RNA. During the computation this term will account for both variables.

After the minus-strand synthesis is complete, the minus-strand RNA,  $R_{ds}$  interacts with the polymerase,  $E$  to form a minus-strand intermediate replicative complex at rate  $k_5$ :



This intermediate complex comprises double-strand RNA and polymerase. Within this complex, minus-strand RNA serves as a template for the production of new complementary plus-strand RNA. Then this intermediate complex leads to the formation of a new plus-strand RNA at an effective rate  $k_{4p}$ :



After the production of new plus-strand RNA,  $R_P$  is done, the intermediate complex dissociates into the double-stranded RNA and the polymerase.

The newly produced plus-strand RNA then interacts with the polymerase,  $E$  to participate in the minus-strand synthesis or is exported at rate  $k_{Pout}$  into the cytoplasm, where it will be translated by the host ribosomes,  $R_{ibo}$  thereby leading to a replication cycle, which lasts till a significant number of viral RNA accumulates in the cell:



This reaction gives a positive feedback to the viral RNA in the cytoplasm.

All reactions developed in this section can be converted to the following system of ordinary differential equations:

$$\begin{aligned} \frac{dR_{Ip}}{dt} &= k_{Pin} T_c H_F E^{cyt} - k_{4m} R_{Ip} + k_3 R_P E - \mu_{Ip} R_{Ip} \\ \frac{dR_{ds}}{dt} &= k_{4m} R_{Ip} + k_{4p} R_{Ids} - k_5 R_{ds} E - \mu_{ds} R_{ds} \\ \frac{dE}{dt} &= k_{4m} R_{Ip} + k_{4p} R_{Ids} - k_5 R_{ds} E - k_3 R_P E - \mu_E E \\ \frac{dR_{Ids}}{dt} &= k_5 R_{ds} E - k_{4p} R_{Ids} - \mu_{Ids} R_{Ids} \\ \frac{dR_P}{dt} &= k_{4p} R_{Ids} - k_3 R_P E - k_{Pout} R_P - \mu_P R_P \end{aligned} \quad (2.14)$$

Eventually, all processes in the cytoplasm and the RV can be integrated into a whole model which is given in Figure 2.1. Similarly, all equations can be combined into a complete system of ordinary differential equations:

$$\begin{aligned}
\frac{dR_P^{unp}}{dt} &= -k_0 R_P^{unp} - \mu_P^{unp} R_P^{unp} \\
\frac{dR_P^{cyt}}{dt} &= k_0 R_P^{unp} + k_2 T_c + k_{Pout} R_P - k_1 R_{ibo} R_P^{cyt} - \mu_P^{cyt} R_P^{cyt} \\
\frac{dT_c}{dt} &= k_1 R_{ibo} R_P^{cyt} - k_2 T_c - k_{Pin} T_c H_F E^{cyt} - \mu_{Tc} T_c \\
\frac{dP}{dt} &= k_2 T_c - k_c P \\
\frac{dE^{cyt}}{dt} &= k_c P - k_{Pin} T_c H_F E^{cyt} - \mu_E^{cyt} E^{cyt} \\
\frac{dH_F}{dt} &= -k_{Pin} T_c H_F E^{cyt} + k_{4m} R_{Ip} \\
\frac{dR_{ibo}}{dt} &= -k_1 R_{ibo} R_P^{cyt} + k_2 T_c + \mu_{Tc} T_c + k_{Pin} T_c H_F E^{cyt} \\
\frac{dL}{dt} &= k_2 T_c - \mu_L L \\
\frac{dR_{Ip}}{dt} &= k_{Pin} T_c H_F E^{cyt} - k_{4m} R_{Ip} + k_3 R_P E - \mu_{Ip} R_{Ip} \\
\frac{dR_{ds}}{dt} &= k_{4m} R_{Ip} + k_{4p} R_{Ids} - k_5 R_{ds} E - \mu_{ds} R_{ds} \\
\frac{dE}{dt} &= k_{4m} R_{Ip} + k_{4p} R_{Ids} - k_5 R_{ds} E - k_3 R_P E - \mu_E E \\
\frac{dR_{Ids}}{dt} &= k_5 R_{ds} E - k_{4p} R_{Ids} - \mu_{Ids} R_{Ids} \\
\frac{dR_P}{dt} &= k_{4p} R_{Ids} - k_3 R_P E - k_{Pout} R_P - \mu_P R_P
\end{aligned} \tag{2.15}$$

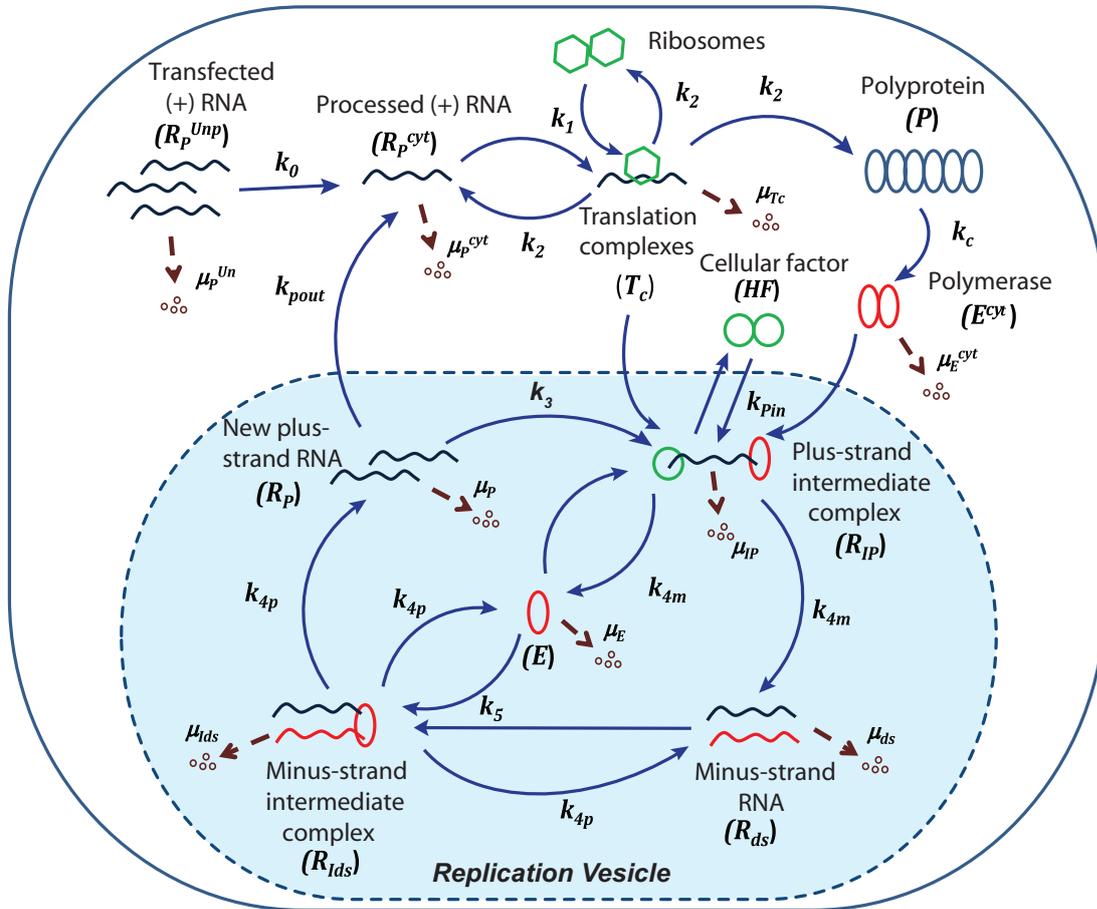


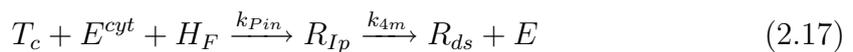
Figure 2.1: Mathematical model of Hepatitis C Virus replication cycle. HCV replication starts once the plus-strand RNA enters the cell during the transfection. We assume that they should be processed first by rate  $k_0$  and during processing degrade with rate  $\mu_P^{unp}$ . Once processed, the viral plus-strand RNA interacts with the host ribosome complexes to form the translation complex with the rate  $k_1$ . Once the translation complex is formed, a viral polyprotein is produced at rate  $k_2$  and subsequently cleaved into nonstructural proteins with rate  $k_c$ . Viral proteins induce the formation of replication vesicles, and the translated plus-strand RNA, the viral polymerase and the cellular factors enter the vesicle with rate  $k_{pin}$ , by forming intermediate plus-strand replication complexes inside. The complementary minus-strand is then formed with rate  $k_{4m}$ , and the intermediate complex is dissociated into dsRNA and NS5B polymerase. Here, we assume that free minus-strand RNA exists in the form of double-stranded RNA (dsRNA). Finally, when the minus-strand RNA is present, the formation of the dsRNA replicative intermediate occurs at rate  $k_5$  and it leads to the synthesis of new plus-strand RNA with rate  $k_{4p}$ . Once the full plus-strand RNA is synthesized, the intermediate complex dissociates into dsRNA and polymerase. The newly synthesized plus-strand RNA then participates in the intermediate complex formation at rate  $k_3$  or leaves the vesicle to participate in the translation process at rate  $k_{pout}$ . In addition, the corresponding degradation rates of the variables are shown.

## 2.3 Host factor assumption

In this section, we will discuss about a particular case of the model. Generally, the property of the host factor which may limit the replication is not known. There are two possible assumptions that we can make about the host factor. First, it may have an enzymatic activity and second, it may be consumed during the replication. In both cases, the host factor will limit the replication. In the model, which has been described above, we assumed that the host factor plays an enzymatic role in the replication process:



In this reaction, the host factor,  $H_F$  participates in the process by playing a machinery role without being consumed. In the case of consumed host factor, we can describe the process as follows:



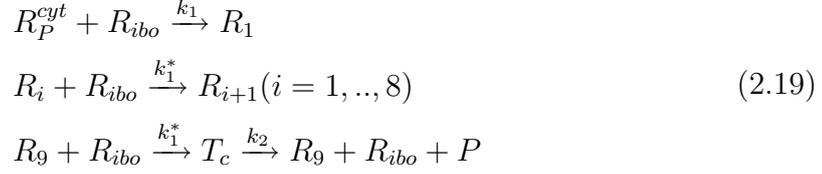
In this case, there will be no feedback from the replication complex,  $R_{Ip}$  to  $H_F$  and the host factor,  $H_F$  is getting consumed in the formation of replication complex,  $R_{Ip}$ . Then the corresponding differential equation for the host factor  $H_F$  looks as follows,

$$\frac{dH_F}{dt} = -k_{Pin} T_c E^{cyt} H_F \quad (2.18)$$

## 2.4 Complex model

In our model, we used a simplified approach which considers a simultaneous binding of 10 ribosome molecules to an HCV RNA. More realistic picture of RNA binding to ribosomes can be obtained by considering a sequential binding. A sequential binding model has been developed and discussed in [48]. The elongation rate in eukaryotes has been estimated at three to eight amino acids per second per ribosome [77], [107]. At these rates, the subgenomic HCV polyprotein, which is 2000 amino acids long, is translated at rate range 5.4 to 14.4 polyproteins per hour per ribosome, which gives a mean rate of 10. Wang *et al.* [158] have shown that at least eight ribosomes were present on an efficient replication of HCV RNA during the translation. If we consider approximately that 10 ribosomes are required for an efficient replication,

then the complex model can be rewritten as:



Here,  $R_i$  represents the number of polysomes consisting of  $i$  ribosomes attached to an HCV RNA. The free plus-strand RNA molecules,  $R_P^{cyl}$ , are converted at rate  $k_1$  into  $R_1$  because of ribosome attachment. A gradual attachment of ribosomes leads to  $R_i + R_{ibo} \xrightarrow{k_1^*} R_{i+1}$  where  $(i = 1, \dots, 8)$ .  $R_i$  and  $R_{i+1}$  differ by the ribosome molecules attached to the HCV RNA. Once the last ribosome molecule attached to the  $R_9$ , the translation complex,  $T_c$  is formed:



Then the translation complex leads to the production of a polyprotein



After the synthesis of polyprotein is complete,  $T_c$  dissociates into  $R_9$  and  $R_{ibo}$ . We further assume that once the plus-strand RNA attaches to a ribosome and become a polysome, it will remain in this state until its degradation with an average rate of  $\mu_{T_c}$ . Under these assumptions, the model can be converted into the following differential equations:

$$\begin{aligned}
\frac{dR_1}{dt} &= k_1 R_{ibo} R_P^{cyl} - k_1^* R_{ibo} R_1 - \mu_{T_c} R_1 \\
\frac{dR_{n+1}}{dt} &= k_1^* R_{ibo} R_n - k_1^* R_{ibo} R_{n+1} - \mu_{T_c} R_{n+1} \\
\frac{dR_9}{dt} &= k_1^* R_{ibo} R_8 + k_2 T_c - k_1^* R_{ibo} R_9 - k_{Pin} T_c E^{cyl} H_F - \mu_{T_c} R_9
\end{aligned} \tag{2.22}$$

Adding and modifying these equations to the system of equations from the simplified model, we get a complex system of ordinary differential equations for HCV RNA replication described by the model of sequential binding. Throughout the thesis, we will concentrate on the simplified model and the complex model will be considered only as a special case.

## 2.5 Experimentally determined parameter values

Totally, the simplified model contains 20 kinetic parameters and 3 initial values. Out of all parameters, 5 are experimentally determined. From the available kinetic data, we have determined one initial value which is the number of transfected plus-strand RNA. Here, we will give a short overview from the literature about some parameters which have been experimentally determined.

### HCV polyprotein elongation rate

We have already discussed in Section 2.4 that in eukaryotes the elongation rate has been estimated as at three to eight amino acids per second per ribosome [77], [107]. This gives 5.4 to 14.4 polyproteins per hour per ribosome. A mean rate, in this case, is 10 polyproteins per hour per ribosome. It has been reported that at least 8 ribosomes were present on an efficient replicon HCV RNA during translation [158]. In our simple model, as it was applied in [48], we assumed that 10 ribosome molecules simultaneously translate one molecule of HCV RNA. Thus, we have the polysome size of 10 ribosomes per HCV RNA yielding a subgenomic HCV polyprotein translation rate,  $k_2 = 100$  polyproteins per hour per ribosome.

### HCV plus- and minus-strand RNA synthesis rate

Experimentally, HCV RNA has been estimated to be synthesized at about 150 nucleotides (nt) per minute by the HCV recombinant NS5B purified from *Escherichia Coli* [106]. In Huh-7 cells, it has been estimated about 180 nt/min [82]. HCV RNA consists of about 6300 nt, and a respective synthesis rate gives  $k_{4p} = 1.7$  RNA molecules per hour per replicative intermediate complex [154]. Since in our model, minus-strand RNA exists with dsRNA, we assume that there is no difference in the synthesis rate when the minus-strand or the plus-strand serves as a template for replication,  $k_{4m} = k_{4p}$ .

### NS5B polymerase degradation rate in cytoplasm

A protease treatment in vitro revealed that about 95% of NS5B were sensitive and found not to be involved in a replicase activity [120], [95]. This suggests that a vast majority of polymerase molecules reside in the cytoplasm, whereas only a minor fraction reside in the RV. We assumed that reported half life of 12 hours of NS5B [114], which was measured in Huh-7 cells in vitro corresponds to the half life of polymerase in the cytoplasm. From this, we can calculate the corresponding degradation

rate, which is  $\mu_E^{cyt} = 0.06 \text{ h}^{-1}$ .

### HCV plus- and minus-strand RNA degradation rates

A nuclease<sup>1</sup> treatment of replicon cells revealed that some fraction of plus-strand RNAs as well as minus-strand RNAs were resistant against this treatment. The corresponding half-life of both variables were detected to be about 10 and 11.5 hours, respectively. This gives the corresponding degradation rates  $0.07 \text{ h}^{-1}$  and  $0.06 \text{ h}^{-1}$ , respectively. We set this rate as a lower boundary for the cytoplasmic HCV RNA degradation. From the kinetic data, we calculated that the transfected plus-strands degrade with the half life about 1 hour which corresponds to the degradation rate,  $0.7 \text{ h}^{-1}$ . We used this information when estimating the degradation rate of transfected plus-strand RNA.

### Degradation rate of polyprotein marker

The polyprotein kinetics have been measured using a luciferase reporter. It has been reported that the half life of this polyprotein marker is about 2 – 4 hours [150], [69]. In our case, we assume that the half life is equal to 2 hours, which gives the corresponding degradation rate of  $\mu_L = 0.35 \text{ h}^{-1}$ .

---

<sup>1</sup>Nuclease is an enzyme that cleaves the bonds between nucleotide subunits of nucleic acids. They degrade RNA and DNA into their mononucleotide building blocks [76].

Table 2.1: *Definitions of the model parameters.*

Parameters	Definitions
$k_0$	processing rate of transfected plus-strand RNA
$k_1$	formation rate of translation complex
$k_2$	polyprotein translation rate
$k_c$	polyprotein cleavage rate
$k_{Pin}$	formation rate of the plus-strand replicative intermediate complex
$k_{Pout}$	transport rate of new plus-strand RNA into cytoplasm
$k_{Am}$	minus-strand RNA synthesis rate
$k_{Ap}$	plus-strand RNA synthesis rate
$k_5$	formation rate of the minus-strand replicative intermediate complex
$k_3$	formation rate of the minus-strand replicative intermediate complex
$\mu_P^{unp}$	degradation rate of transfected plus-strand RNA
$\mu_P^{cyt}$	degradation rate of processed plus-strand RNA
$\mu_{Tc}$	degradation rate of translation complex
$\mu_E^{cyt}$	degradation rate of NS5B protein
$\mu_{Ip}$	degradation rate plus-strand replicative intermediate complex
$\mu_{ds}$	degradation rate of minus-strand RNA
$\mu_E$	degradation rate of active polymerase in replication vesicles
$\mu_{Ids}$	degradation rate of minus-strand replicative intermediate complex
$\mu_P$	degradation rate of nascent plus-strand RNA
$\mu_L$	degradation rate of polyprotein (luciferase)
$R_P^{unp}(0)$	initial number of transfected plus-strand RNA
$R_{ibo}(0)$	initial number of ribosome complexes
$H_F(0)$	initial number of cellular host factor
$p_{scale}$	scaling factor for polyprotein levels (luciferase)



# Chapter 3

## Model Calibration

Calibration is a process by which parameters in a model are adjusted to match a model performance to experimental data [9]. Model calibration has a potential to yield correct and weakly determined values. When learning the specific process kinetically, the rate of the certain process is of great importance. Calibrating the models against the kinetic data gives an excellent method to estimate the kinetic rates of interest. In this Chapter, we will discuss about the model calibration process by explaining individual steps in details.

### 3.1 Inverse problem

After we obtained the ordinary differential equation model for HCV replication, the next task for us is to learn essential parameters in the model which are unknown. In the present case, except the parameters which have been experimentally observed, the remaining parameters should be estimated from the experimental data, which leads to an inverse problem. Given a complete description of a physical system, we can predict the outcome of some measurements. This problem of predicting the result of measurements is called *the modelization problem, the simulation problem, or the forward problem* [147].

The inverse problem consists of using the actual result of some measurements to infer the values of the parameters that characterize the system. The forward problem has a unique solution, while the inverse problem does not.

Lets assume an equation of the form,

$$K(x) = y \tag{3.1}$$

Where  $K$  is an operator describing the explicit relation between the observation  $y$

and the parameter  $x$ . One can formulate the *direct problem* as the evaluation of the operator  $K$ , given the parameter  $x$ . Similarly, the *inverse problem* can be formulated as to solve (3.1) for the parameter  $x$ .

The inverse problem is ill-posed, while the direct problem is well-posed in the sense of Hadamard [50]. According to [65], a definition of a well-posed problem can be given as follows:

**Definition (well-posedness) 3.1.1.** Let  $X$  and  $Y$  be normed spaces,  $K : X \rightarrow Y$  a linear or nonlinear mapping. The equation  $Kx = y$  is called well-posed if the following conditions hold:

- Existence: For every  $y \in Y$  there is at least one  $x \in X$  such that  $Kx = y$ .
- Uniqueness: For every  $y \in Y$  there is at most one  $x \in X$  with  $Kx = y$ .
- Stability: The solution  $x$  depends continuously on  $y$ , *i.e.*, for every sequence  $(x_n) \subset X$  with  $Kx_n \rightarrow Ky$  ( $n \rightarrow \infty$ ), it follows that  $x_n \rightarrow x$  ( $n \rightarrow \infty$ ).

If one of these properties does not hold then the problem is called ill-posed.

Inverse problems are usually ill-posed, whereas the direct problems are well-posed. The following theorem implies that the linear equations of the form  $Kx = y$  with the compact operators  $K$  are always ill-posed.

**Theorem 3.1.2.** Let  $X$  and  $Y$  be normed spaces and  $K : X \rightarrow Y$  be a linear compact operator with nulspace  $N(K) := \{x \in X : Kx = 0\}$ . Let the dimension of the factor space  $X/N(K)$  be infinite. Then there exists a sequence  $(x_n) \subset X$  such that  $Kx_n \rightarrow 0$  but  $(x_n)$  does not converge. In particular, if  $K$  is one-to-one, the inverse  $K^{-1} : Y \supset K(X) \rightarrow X$  is unbounded.

One method to solve an equation  $Kx = y$  is a least squares method. In the following, we will shortly overview this method. For more information see [65].

Let a finite dimensional subspace  $X_n \subset X$  be given. Determine the solution  $x_n \in X_n$  such that

$$\|Kx_n - y\| \leq \|Kz_n - y\| \quad \text{for all } z_n \in X_n. \quad (3.2)$$

Existence and uniqueness of  $x_n \in X_n$  follow from the fact that  $X_n$  is finite dimensional and  $K$  is one-to-one. The solution  $x_n$  of this least squares problem is defined

as

$$(Kx_n, Kz_n) = (y, Kz_n) \quad \text{for all } z_n \in X_n. \quad (3.3)$$

If one chooses a basis  $\hat{x}_j, j = 1, \dots, n$  of  $X_n$ , then (3.2) reduces to a finite dimensional system.

$$\sum_{j=1}^n \alpha_j (K\hat{x}_j, K\hat{x}_i) = \beta_i = (y, K\hat{x}_i), \quad i = 1, \dots, n. \quad (3.4)$$

If we denote  $A_{ij} = (K\hat{x}_j, K\hat{x}_i)$ , then we have  $A\alpha = \beta$ , where  $A \in \mathbf{K}^{n \times n}$  and positive definite, because  $K$  is one-to-one. If  $\mathbf{K} = \mathbf{R}$  then  $A$  is symmetric. Now we assume that the right hand side of equation (3.3) is perturbed. Let  $x_n^\delta$  be the solution of the following equation for continuous perturbations,

$$(Kx_n^\delta, Kz_n) = (y^\delta, Kz_n), \quad \forall z_n \in X_n \quad (3.5)$$

where  $y^\delta \in Y$  is the perturbed right hand side with  $\|y^\delta - y\| \leq \delta$ . The same can be applied to the finite system (3.4). In this case, we consider that  $\beta \in \mathbf{K}^n$  is substituted by  $\beta^\delta \in \mathbf{K}^n$  with  $|\beta^\delta - \beta| \leq \delta$ , where  $|\cdot|$  is an Euclidean norm in  $\mathbf{K}^n$ . By doing this, one obtains the following finite system of equations:

$$\sum_{j=1}^n \alpha_j^\delta (K\hat{x}_j, K\hat{x}_i) = \beta_j^\delta, \quad i = 1, \dots, n. \quad (3.6)$$

Because  $A$  is positive-definite, this system of equations has a unique solution. Since the least squares method is convergent, any optimization algorithm can be applied to minimize  $\|Kx - y\|$ . In the following Section 3.2, we will use the optimization algorithms based on Multiple-Shooting and Genetic Algorithm to find an optimal  $x$  which minimizes the value of objective function.

## 3.2 Parameter estimation

Prior to estimation, we specified some parameter values based on the literature data, which we discussed in Section 2.5. Except those parameters which values were available from the literature, the remaining parameters are estimated from the experimental data by minimizing an objective function. The parameter estimation has been performed using the multiple shooting and genetic algorithms. A complete list of parameter values is given in Table 3.1. The detailed description of the optimization algorithms, we will give below. We will start with the formulation of the parameter estimation problem.

Lets assume that the replication system (2.15) is given in a general form

$$\dot{y}(t) = f(t, y(t), p), \quad t \in [t_0, t_f] \quad (3.7)$$

where  $y(t)$  denotes the vector given by the concentrations of the 13 species,

$$y(t) = (R_P^{unp}, R_P^{cyl}, T_c, P, E^{cyl}, H_F, R_{Ip}, R_{ds}, E, R_{Ids}, R_P, R_{ibo}, L)$$

$p$  is a unknown parameter and has to be determined by parameter estimation.  $t \in [t_0, t_f]$  is the time.

Lets assume that the experiments have been carried out at the given times  $t_j$ ,  $j = 1, \dots, m$  yielding the measurements  $\eta_{ij}$ ,  $i = 1, \dots, m_j$ ,  $j = 1, \dots, m$  of the observation function  $h_{ij}$  which depends on the variables  $y(t)$  and the parameter  $p$ ,

$$\eta_{ij} = h_{ij}(t_j, y(t_j), p) + \varepsilon_{ij}, \quad (3.8)$$

which are subject to the measurement errors  $\varepsilon_{ij}$ . This formulation also includes the case when several species are measured at a time  $t_j$ .

The parameters are estimated by minimizing the difference between the model and the data,

$$\|F_1(y(t_1), \dots, y(t_m), p)\| = \|(\eta_{ij} - h_{ij}(t_j, y(t_j), p))w_{ij}\| \quad (3.9)$$

In our case, the objective function is chosen as a least squares sum of the differences between the measurement  $\eta_{ij}$  and the model predictions  $h_{ij}$  weighted with the factor  $w_{ij}$ ,

$$\min_{y(t), p} \sum_{ij} ((\eta_{ij} - h_{ij}(t_j, y(t_j), p))w_{ij})^2 \quad (3.10)$$

In the case of independent and normally distributed measurement errors with zero mean,

$$\varepsilon_{ij} = \eta_{ij} - h_{ij}(t_j, y(t_j), p) \in \mathcal{N}(0, \sigma_{ij}^2) \quad (3.11)$$

and choosing  $w_{ij} = \frac{1}{\sigma_{ij}}$ , (3.10) yields a maximum likelihood estimate.

Moreover, the additional knowledge about the state variables and the parameters in the model can be considered as the equality and inequality constraints at time  $t_j$ ,

$$F_2(y(t_j), p) = 0, \quad F_3(y(t_j), p) \geq 0, \quad (3.12)$$

summing up, the parameter estimation problem can be formulated as

$$\begin{aligned} & \min_{y(t), p} \sum_{ij} ((\eta_{ij} - h_{ij}(t_j, y(t_j), p))w_{ij})^2 \\ & \text{s.t. } (y(t), p) \text{ solves } \dot{y}(t) = f(t, y(t), p), \quad t \in [t_0, t_f] \\ & F_2(y(t_j), p) = 0, \quad F_3(y(t_j), p) \geq 0 \end{aligned} \quad (3.13)$$

The unknown parameters in our case are,

$$\begin{aligned} p = & (k_0, k_1, k_c, k_{Pin}, k_{Pout}, k_5, k_3, \mu_P^{unp}, \mu_P^{cyt}, \mu_{Tc}, \\ & R_{ibo}(0), \mu_{VMS}, \mu_L, H_{F,high}(0), H_{F,low}(0), p_{scale}) \end{aligned} \quad (3.14)$$

The experimental measurements have been performed within the time interval 0 and 72 hours. Hence, the time interval is chosen as  $t \in [t_0, t_f] = [0, 72]$ .

In our case, the measurement function  $h_{ij}$  is defined as follows:  
For the plus-strand RNA measurements, the measurement function is given by

$$R_P^{tot} = R_P^{unp} + R_P^{cyt} + T_c + R_P + R_{Ip} + R_{ds} + R_{Ids} \quad (3.15)$$

For the minus-strand RNA measurements

$$R_m^{tot} = R_{ds} + R_{Ids} \quad (3.16)$$

and for the polyprotein measurements, it is given by

$$L_{pol} = p_{scale} L \quad (3.17)$$

Additionally, we have the information about the state variables at steady state and the measurement functions for them are defined as:

The ratio of plus-strand RNA to minus-strand RNA,

$$R_P^{tot}/R_m^{tot} = (R_P^{unp} + R_P^{cyt} + T_c + R_P + R_{Ip} + R_{ds} + R_{Ids})/(R_{ds} + R_{Ids}) \quad (3.18)$$

The ratio of plus-strand RNA to minus-strand RNA in the replication vesicles,

$$R_P^{RV}/R_m^{tot} = (R_P + R_{Ip} + R_{ds} + R_{Ids})/(R_{ds} + R_{Ids}) \quad (3.19)$$

The ratio of plus-strand RNA in the cytoplasm to plus-strand RNA in the repli-

cation vesicles,

$$R_P^{CYT}/R_P^{RV} = (R_P^{unp} + R_P^{cyl} + T_c)/(R_P + R_{Ip} + R_{ds} + R_{Ids}) \quad (3.20)$$

### 3.2.1 Multiple shooting parametrization

The problem (3.13) is an infinite-dimensional optimization problem, since the functions  $y$  have been defined as the optimization variables, which have to fulfill an infinite dimensional equality constraint. To reduce the optimization problem to a finite dimension, we used a multiple shooting method [22], [23], [24]. This method has many applications in different areas such as chemical engineering (the denitrogenization of pyridine, [22]), biophysics (the photosynthesis process, [14]) and civil space flight (satellite orbit determination, [71]). Bock *at el.* [24] reported that this method is numerically stable and usage of this method enables to decrease the nonlinearity of the problem. Dividing the integration interval limits an error propagation and allows to solve optimization problems for unstable or chaotic systems [24], [13], [61].

We can now briefly describe the procedure of this method. For detailed overview, we refer to [24] and [71].

One discretizes the time interval where the measurements are given by choosing a suitable grid of multiple shooting nodes  $\tau_k$ ,

$$t_0 = \tau_0 < \tau_1 < \dots < \tau_M = t_f, \quad (3.21)$$

The value of the state variables,  $s_k$  are chosen as additional unknowns and  $M$  relaxed initial value problems at each grid point (see Figure 3.1).

Initial value problem,

$$\dot{y}(t) = f(t, y(t), p), \quad y(\tau_k) = s_k \quad (3.22)$$

are solved on each interval  $[\tau_k, \tau_{k+1}]$  giving a solution  $y(t_k, s_k, p)$ . To ensure that  $y(t)$  is continuous in the solution of the optimization problem, the matching conditions are imposed,

$$s_{k+1} - y(\tau_{k+1}; s_k, p) = 0, \quad k = 0, \dots, M - 1 \quad (3.23)$$

Inserting the computed values  $y(t_k, s_k, p)$  into (3.13), one obtains a constrained

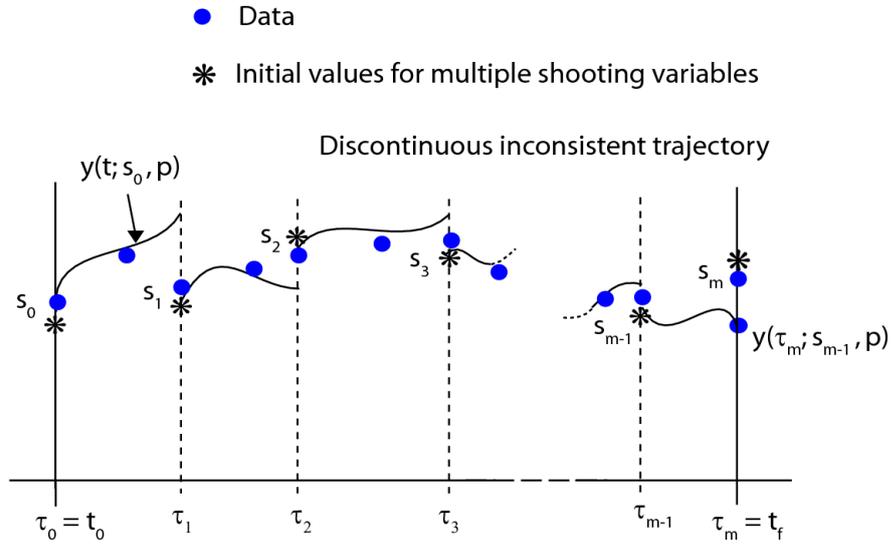


Figure 3.1: Multiple shooting approach. Adapted from [24]. The constrained problem (3.24) is solved as a boundary value problem within the interval of  $[\tau_k, \tau_{k+1}]$ .

problem in the variables  $(s, p) = (s_0, \dots, s_M, p)$ ,

$$\begin{aligned} \min_{s,p} \sum_{i,j} ((\eta_{ij} - h_{ij}(t_j, y(t_j), p))w_{ij})^2 \\ \text{s.t. } F_2(s, p) = 0, \quad F_3(s, p) \geq 0 \\ s_{k+1} - y(\tau_{k+1}; s_k, p) = 0, \quad k = 0, \dots, M-1 \end{aligned} \quad (3.24)$$

Setting  $M = 0$  and omitting the matching conditions will give a single shooting method.

New constrained problem (3.24) is solved using the Gauss-Newton method which is described in Subsection 3.2.2.

### 3.2.2 Generalized Gauss-Newton method

A nonlinear constrained problem simply can be written as

$$\min_x \|F_1(x) = 0\|_2^2 \quad \text{s.t.} \quad F_2(x) = 0, \quad F_3(x) \geq 0 \quad (3.25)$$

For solving this least squares problem, a generalized Gauss-Newton method is used (refer to [71], [24]).

The generalized Gauss-Newton method is an iterative method. Given an iteration,

$$x^k \rightarrow x^{k+1} = x^k + t^k \Delta x^k, \quad 0 < t^k \leq 1 \quad (3.26)$$

which starts from an initial guess  $x^0$  for the optimization variables. Here,  $\Delta x^k$  is determined as the solution of a linearized constrained least-squares problem which is derived from (3.25):

$$\begin{aligned} & \min_{\Delta x^k} \|F_1(x^k) + J_1(x^k)\Delta x^k\|_2^2 \\ \text{s.t.} \quad & F_2(x^k) + J_2(x^k)\Delta x^k = 0, \\ & F_3(x^k) + J_3(x^k)\Delta x^k \geq 0 \end{aligned} \quad (3.27)$$

where  $J_i(x^k) = \left. \frac{\partial F_i(x)}{\partial x} \right|_{x=x^k}$  is the Jacobian matrix of the function  $F_i(x)$  with  $i = 1, 2, 3$ . The  $t^k \in [0, 1]$  in (3.26) is a stepsize.

Using the following notations,

$$F(x^k) = \begin{pmatrix} F_1(x^k) \\ F_2(x^k) \\ F_3(x^k) \end{pmatrix} \quad \text{and} \quad J(x^k) = \begin{pmatrix} J_1(x^k) \\ J_2(x^k) \\ J_3(x^k) \end{pmatrix} \quad (3.28)$$

the solution of the linearized problem can be written as

$$\Delta x^k = -J^+(x^k) \begin{pmatrix} F_1(x^k) \\ F_2(x^k) \\ F_3(x^k) \end{pmatrix} \quad (3.29)$$

where  $J^+(x^k)$  is a generalized inverse of  $J(x^k)$ .

In order to improve the convergence behavior of Gauss-Newton method, the damped iterations  $x^k \rightarrow x^{k+1} = x^k + \lambda^k \Delta x^k$ , with the damping factor  $\lambda^k \in (0, 1]$  was used [25].

### 3.2.3 Statistical analysis

For the solution of the optimization problem, a statistical analysis was performed which is very important for parameter estimation problems. This statistical analysis is based on the covariance matrix  $C$  [24], which is computed from the generalized inverse by

$$C = \beta_c^2 J^+(x^k) \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} (J^+(x^k))^T \quad (3.30)$$

where  $I$  denotes the identity matrix and  $\beta_c^2$  is a common factor, which is computed by  $\beta_c^2 = \frac{\|F_1(x^k)\|_2^2}{l_2}$ , where  $l_2$  denotes the number of degrees of freedoms, i.e. the number of optimization variables minus the number of equality constraints. Estimated standard errors for all optimization variables are calculated by the square roots of the diagonal elements of the covariance matrix.

A confidence interval for the probability  $\alpha$  for the optimization variable  $x_i$  can be obtained by

$$x_i = [x_i - \delta x_i, x_i + \delta x_i] \quad (3.31)$$

where  $\delta x_i = \sqrt{C_{ii}} \sqrt{l_1 F_{1-\alpha}(l_1, l_2)}$ , where  $l_1$  is the number of least squares conditions minus  $l_2$ .  $F_{1-\alpha}(l_1, l_2)$  is a quantile of the  $F$  distribution for error probability  $1 - \alpha$ .

In our case, we observed that during the parameter estimation, the optimization violated the biologically plausible bounds for the parameters  $k_5$  and  $k_c$ . To avoid this issue, we fixed the parameters to their upper bounds and did the covariance analysis for remaining parameters.

### 3.2.4 Software

The numerical methods for the solution of nonlinear constrained least squares problems and the statistical analysis are implemented in the software PARFIT [22],[23], which was used for the parameter estimation of our model. The initial value problem solutions and their derivatives are computed using the solver METANB, which is incorporated in PARFIT.

### 3.2.5 Genetic algorithm

Genetic algorithms are search algorithms based on the mechanics of natural selection and natural genetics [46]. We summarize this algorithm from [86] as following: The genetic algorithm solves optimization problems by mimicking the principles of

biological evolution, repeatedly modifying a population of individual points using rules modeled on gene combinations in biological reproduction.

We will begin by discussing the genetic algorithm and its properties. A detailed description about the genetic algorithm can be found in [46], [159].

These algorithms are a subclass of evolutionary algorithms where the elements of the search space are binary strings or arrays of other elementary types. Genetic algorithms have been developed by John Holland and his colleagues at the University of Michigan [53].

In the genetic algorithm, an objective function is referred to *fitness functions*. The search space of genetic algorithms is referred to *genome* and its elements are called *genotypes*. Genotypes encompass the whole hereditary information of an organism encoded in the DNA. The DNA is a string of base pairs that encodes the phenotypical characteristics of the creature it belongs to. In genetic algorithm, the genomes are strings, like natural prototypes. Due to a linear structure, these phenotypes are also called *chromosomes* [159]. The position where a specific gene is located in a chromosome is called a *locus*.

**Definition (String Chromosome) [159] 3.2.1.** A string chromosome can either be a fixed-length tuple or variable-length list.

In the case of fixed-length tuple, the loci  $i$  of the genes  $g_i$  are constant and, hence, the tuples may contain elements of different types  $G_i$ ,

$$G = \{\forall(g[1], g[2], \dots, g[n]) : g[i] \in G_i, \forall i \in 1, \dots, n\} \quad (3.32)$$

This is not a case in variable-length string genomes. Here, the positions of the genes may shift when the reproduction operations are applied. Thus, all elements of such genotypes must have the same type  $G_T$ ,

$$G = \{\forall lists, g : g[i] \in G_T, \forall 0 \leq i < len(g)\} \quad (3.33)$$

String chromosomes are normally bit strings, vectors of integer numbers or vectors of real numbers.

**Definition (Intron) [159] 3.2.2.** Parts of a genotype  $g \in G$  that does not contribute to the phenotype  $x = gpm(g)$  are called introns.

### 3.2.6 Genetic algorithm operators

Here, we will discuss the genetic algorithm operators and for detailed information refer to [159] and [46].

**Creation:** Creation of fixed-length string individuals means to create a new tuple of the structure defined by the genome and initialize it with random values. It can be roughly described as:

$$\text{create}_{f(l)} \equiv (g[1], g[2], \dots, g[n]) : g[i] = G_i[\text{random}_u * \text{len}(G_i)], \forall i \in 1, \dots, n$$

Variable-length strings can be created by first randomly drawing a length  $l > 0$  and then creating a list of that length filled with random elements. In this process, the individual strings are copied according to their fitness function values. This means that the strings with higher value contribute more offspring in the next generation with higher probability.

**Mutation:** Mutation is an important method for keeping the diversity of the solution candidates by introducing small, random changes into them. In fixed length string chromosomes, this is done by randomly modifying the value (allele) of a gene (Figure 3.2 (a) and (b)). More general version of this form of mutation where  $0 < n < \text{len}(g)$  locations in the genotype  $g$  are changed simultaneously. In binary coded chromosomes, for instance, these genes are bits which can simply be toggled. For real-encoded genomes, modifying an element  $g_i$  can be realized by replacing it with a number drawn from a normal distribution with an expected value  $g_1$ ,  $g_i^{\text{new}} \sim \mathcal{N}(g_1, \sigma^2)$ . In the case of the string chromosomes with variable length, the set of mutation operations can be extended by two additional methods. First, a couple of genes with randomly chosen alleles at any given position can be inserted into a chromosome (Figure 3.2 (d)). Second, in contrast, the elements can be deleted from the string (Figure 3.2 (e)).

**Permutation:** The permutation is one of mutation methods where the alleles of two genes are exchanged (Figure 3.2 (c)). This makes sense only if all genes have similar data types.

**Crossover:** Crossover is a recombination of two string chromosomes which is performed by swapping parts of two genotypes. In a case of single-point crossover, both parental chromosomes are split at randomly determined crossover points. As a

result, a new child genotype is created by appending the second part of the second parent to the first part of the first parent (Figure 3.3 (a)). In two-point crossover, both parental genotypes are split at two points and a new offspring is created by using parts number one and three from the first, and the middle part from the second parent chromosome (Figure 3.3 (b)). For fixed-length strings, the crossover points for both parents are always identical. In the variable-length string chromosomes, the same crossover operations are available as for fixed-length strings except the strings that are no longer necessarily split at the same points. The length of the new strings resulting from such deletion and insertion operation may differ from the lengths of the parents.

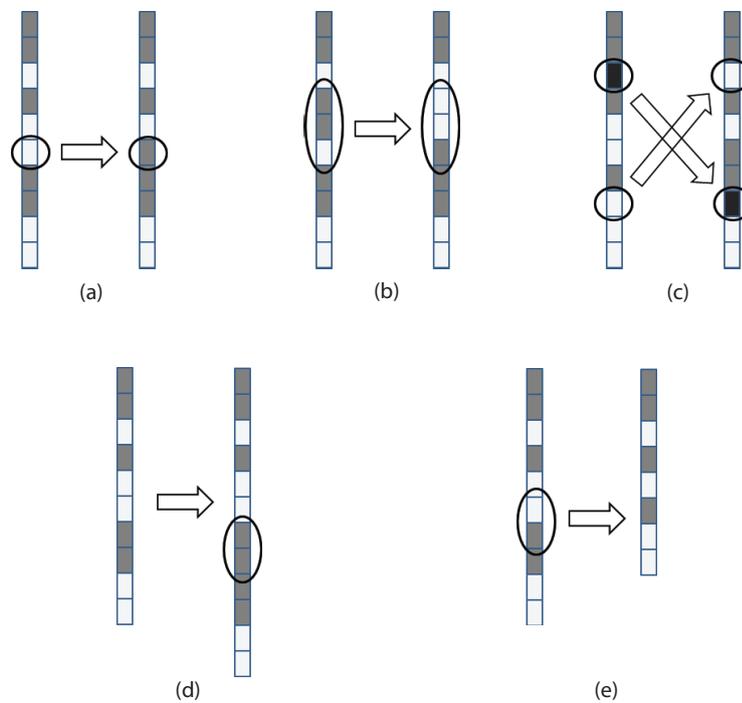


Figure 3.2: Mutation and permutation in genetic algorithm (a) Single gene mutation (b) Multi gene mutation (c) Permutation of genes. Adapted from [159].

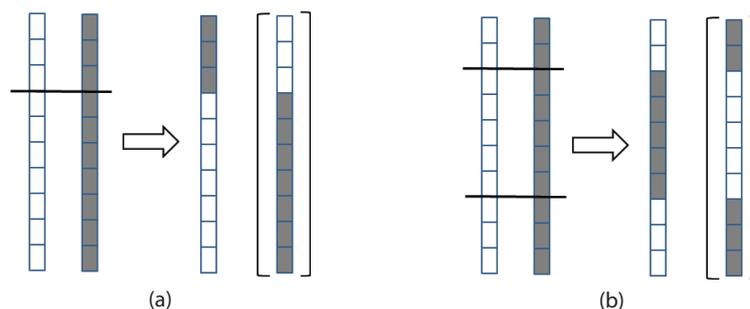


Figure 3.3: Crossover in genetic algorithm (a) Single gene mutation (b) Multi gene mutation (c) Permutation of genes. Adapted from [159].

### 3.2.7 A simple genetic algorithm

Now, we will briefly show how a simple genetic algorithm works. In his work, Mitchell [93] summarized the simple genetic algorithm as follows:

1. Generate a random population of  $n$   $l$ -bit chromosomes i.e. candidate solutions.
2. Calculate the fitness  $f(x)$  of each chromosome  $x$  in the population.
3. The following steps should be repeated until  $n$  offspring have been created:
  - (a) Select a pair of parent chromosomes from the current population. Probability of selection depends on the fitness function. The same chromosome can be selected to become a parent more than once.
  - (b) Cross over the pair with probability  $p_c$  at a randomly chosen point to form two offsprings. If there is no crossover, then two offsprings which are exact copies of their parents are formed.
  - (c) Mutate the two offsprings at each locus with the probability  $p_m$ . If there is a mutation takes place, then there will be now a change in the chromosome. In opposite, if there is 100% mutation, then the whole chromosome is changed. After the mutation integrate the resulting chromosomes into the new population. Mutation often prevents from stacking into a local minima.
4. Replace the current population with the new population.
5. Go to step 2 (these procedure should be repeated with the new population).

Each iteration in GA is referred to a *generation*.

Another important parameter which can improve the optimization is a *population size*. Population size gives information about the number of chromosomes in population [103]. If the population size is too small, only a small part of search space

is explored. In this case, the probability of performing crossover will be dropped. A high population size leads to the generation of many chromosomes which slows down the optimization process.

### 3.2.8 Optimization using genetic algorithm

We performed optimization using *Global Optimization Toolbox* in Matlab. We used Genetic Algorithm (GA) to solve the least squares problem (3.13) defined in Section 3.2. Since there is no structured guidance for using GA operators which maximizes the search for an optimum solution, we set up these operators as following:

Out of total population, we set 10% to be elite children which means that 10% of all individuals with the best fitness value proceed to the next generation.

In Matlab, the parameter which controls the crossover is given in a range of  $[0, 1]$ . 0 means no crossover and 1 means full crossover. We set up this parameter as 0.7, which means that 70% of population which remain after excluding elite children participate in the crossover. The remaining individuals participate in the mutation.

The population size has been chosen between 300-2000 individuals. We found out that the optimization with the high population size gives better results. However, it took longer to converge to the optimum solution. Optimization with the low population size has higher probability to stack into a local minima and usually requires more generations to converge to a global minima.

The number of generations have been chosen between 1000 and 2000. The optimization with the high population size required less generations, whereas the optimization with the low population size needed more generations. If the population size is  $X$ , then the distribution of the population will look as,

$$\begin{aligned}
 0.1 * X &= \text{elite children} \\
 0.7 * (X - 0.1 * X) &= \text{crossover children} \\
 \text{remaining population} &= \text{mutation children}
 \end{aligned}
 \tag{3.34}$$

Table 3.3 demonstrates the GA operators used in our case. The best parameter values obtained by GA are given in Table 3.2.

Table 3.1: *Estimated kinetic parameters and initial values using Multiple Shooting Algorithm.*

Parameter name	Parameter values	90% confidence intervals
$k_0$	$0.00587 h^{-1}$	$(-1.49 \times 10^{-3}, 0.0132)$
$k_1$	$1 h^{-1}molec^{-1}$	fixed
$k_2$	$100 h^{-1}$	experimentally observed
$k_c$	$1 h^{-1}$	fixed
$k_{Pin}$	$2.07 \times 10^{-6} h^{-1}molec^{-2}$	$(-3.23 \times 10^{-6}, 7.36 \times 10^{-6})$
$k_{Pout}$	$0.333 h^{-1}$	$(0.165, 0.502)$
$k_{Am}$	$1.7 h^{-1}$	experimentally observed
$k_{Ap}$	$1.7 h^{-1}$	experimentally observed
$k_5$	$10 h^{-1}molec^{-1}$	fixed
$k_3$	$10^{-4} h^{-1}molec^{-1}$	fixed
$\mu_P^{unp}$	$0.758 h^{-1}$	$(0.464, 1.05)$
$\mu_P^{cyt}$	$0.487 h^{-1}$	$(0.212, 0.762)$
$\mu_{T_c}$	$0.243 h^{-1}$	$(0.106, 0.381)$
$\mu_E^{cyt}$	$0.06 h^{-1}$	experimentally observed
$\mu_{Ip}$	$0.0703 h^{-1}$	$(0.0329, 0.108)$
$\mu_{ds}$	$0.0703 h^{-1}$	$(0.0329, 0.108)$
$\mu_E$	$0.0703 h^{-1}$	$(0.0329, 0.108)$
$\mu_{Ids}$	$0.0703 h^{-1}$	$(0.0329, 0.108)$
$\mu_P$	$0.0703 h^{-1}$	$(0.0329, 0.108)$
$\mu_L$	$0.35 h^{-1}$	experimentally observed
$R_{ibo}(0)$	995 molecules	$(-1030, 3020)$
$H_{F,high}(0)$	88 molecules	$(-30, 206)$
$H_{F,low}(0)$	10 molecules	$(-4, 23)$
$p_{scale}$	$2.80 \times 10^3$	$(-668, 6.28 \times 10^3)$

Table 3.2: *Estimated kinetic parameters and initial values using Genetic Algorithm.*

Parameter name	Parameter values	Status
$k_0$	$0.0036743 h^{-1}$	estimated
$k_1$	$1 h^{-1}molec^{-1}$	fixed
$k_2$	$100 h^{-1}$	experimentally observed
$k_c$	$2.6045 h^{-1}$	estimated
$k_{Pin}$	$2.6408e - 6 h^{-1}molec^{-2}$	estimated
$k_{Pout}$	$0.35673 h^{-1}$	estimated
$k_{4m}$	$1.7 h^{-1}$	experimentally observed
$k_{4p}$	$1.7 h^{-1}$	experimentally observed
$k_5$	$10 h^{-1}molec^{-1}$	fixed
$k_3$	$10^{-4} h^{-1}molec^{-1}$	fixed
$\mu_P^{unp}$	$0.70314 h^{-1}$	estimated
$\mu_P^{cyt}$	$0.36485 h^{-1}$	estimated
$\mu_{Tc}$	$0.2318 h^{-1}$	estimated
$\mu_E^{cyt}$	$0.06 h^{-1}$	experimentally observed
$\mu_{Ip}$	$0.054975 h^{-1}$	estimated
$\mu_{ds}$	$0.054975 h^{-1}$	estimated
$\mu_E$	$0.054975 h^{-1}$	estimated
$\mu_{Ids}$	$0.054975 h^{-1}$	estimated
$\mu_P$	$0.054975 h^{-1}$	estimated
$\mu_L$	$0.35 h^{-1}$	experimentally observed
$R_{ibo}(0)$	500 molecules	estimated
$H_{F,high}(0)$	107 molecules	estimated
$H_{F,low}(0)$	10 molecules	estimated
$p_{scale}$	$1.6 \times 10^3$	estimated

Table 3.3: *The values of genetic algorithm operators used in parameter estimation. These values have been obtained using the formula (3.34). The parameters that have been obtained using these values are given in Table 3.2.*

GA operators	Chosen value
Population size	300
Elite children	30
Crossover fraction	189
Mutation size	81
Generation	1000

### 3.3 HCV replication dynamics

In order to see whether the model explains the replication process, we fitted the model to several kinetic datasets and steady-state observations using the multiple shooting and genetic algorithms which we discussed in Section 3.2. The kinetic measurements have been obtained by measuring the HCV replication in two cell lines, which are the high permissive Huh-7 Lunet cells and the less permissive Huh-7 cells.

The kinetic measurements used in the model calibration are:

- Total number of plus-strand RNAs (molecules)
- Total number of minus-strand RNAs (molecules)
- Polyprotein measurements (luciferase activity)

Experimental settings for the polyprotein measurements differ from that of the plus and minus-strand RNA. Therefore, we introduced a scale factor,  $p_{scale}$  which normalizes the luciferase activity to the molecule numbers. We estimated this parameter during the optimization.

Additionally, we used the following steady-state observations:

- The ratio of total plus-strand RNA to total minus-strand RNA is about 10 : 1 [47, 154, 120].
- The plus-strand to minus-strand RNA ratio in the RVs is about 6 : 1 [120].
- The ratio of plus-strand RNA outside the RVs to that of inside the RVs is about 1 : 1 [120].

We fitted our model to the experimental data from high and low permissive cells simultaneously. The results demonstrated that the model explains both the kinetic and steady state data excellently (Figure 3.4). We revealed that the model can reproduce the kinetic data from high and low permissive cells with the same set of parameters, except the initial value for the host factor,  $H_F(0)$ . All parameters obtained from the model fitting are given in Table 3.1. All figures are shown in this thesis are obtained using the parameter values from Table 3.1.

Further, we will discuss about the calibration process and observed dynamics more in details:

### 3.3.1 Plus-strand RNA kinetics

Since experimentally accessible information about the plus-strand RNA is in total numbers, a model output to describe this will be,

$$R_P^{tot} = R_P^{unp} + R_P^{cvt} + T_c + R_{Ip} + R_{ds} + R_{Ids} + R_P \quad (3.35)$$

which is just the sum of all variables containing the plus-strand RNA. The variables  $R_P^{unp}$ ,  $R_P^{cvt}$ ,  $T_c$ ,  $R_{Ip}$ ,  $R_{ds}$ ,  $R_{Ids}$  and  $R_P$  are the solutions from the system of equations (2.15).

The model fitting is depicted in (Figure 3.4 A, B). From the plot, we can see that the model ( $R_P^{tot}$ ) excellently captures the underlying dynamics. If we look into the plus-strand RNA dynamics more in details, then we can see the main properties in the dynamics of the data which are the lowest peak, the saturation time and the magnitude level of replication. An initial decline in the dynamics reflects the phase where the degradation is dominant over the production. Initially, the decay is mostly due to the degradation of transfected plus-strands,  $R_P^{unp}$ , which we will discuss later in Subsection 4.2.4. The lowest point of replication occurs around  $t = 8$  in the high and  $t = 12$  hours in the low permissive cells. After  $t = 8$  hours the production of new-strands outcompete the degradation, and then the rise in plus-strands starts in the case of high permissive cells. In the low permissive cells, the rise observed later after  $t = 12$  hours. The replication attains a steady state around  $t = 24$  hours in the high permissive cells, and around  $t = 40$  hours in the low permissive cells.

### 3.3.2 Minus-strand RNA kinetics

As mentioned in Subsection 3.3.1, the minus-strand RNA is also given in total numbers. From the model, we can compute the total minus-strand RNA as follows:

$$R_M^{tot} = R_{ds} + R_{Ids} \quad (3.36)$$

$R_{ds}$  and  $R_{Ids}$  are the solutions of the system of equations (2.15).

From the fitting, we see that the model ( $R_M^{tot}$ ) fits to the data with an excellent accuracy (Figure 3.4 A and B). The kinetics of minus-strand RNA starts from 0 molecules, because initially there is no minus-strand available. The minus-strand RNA is synthesized during the replication by using the plus-strand RNA as a template. As depicted in Figure 3.4, the production of minus-strands increases exponentially at the initial phase and slows down after  $t = 24$  hours in the high and  $t = 40$

hours in the low permissive cells. From the plot, it is clear that the minus and plus-strand RNA share some similarities. An initial increase in the minus-strand RNA is highly correlated with the increase in the plus-strand RNA after  $t = 8$  hours in the high and  $t = 12$  hours in the low permissive cells. Furthermore, both the plus and minus-strand RNA saturate approximately at the same time. These similarities in the dynamics indicate a high correlation in the production between plus and minus-strand RNA. As steady-state data indicated, at saturation, there are about 10 fold more plus-strand than the minus-strands available [120]. This can be seen also by comparing the simulated plus and minus-strand RNA levels at steady state (Figure 3.4). Figure 3.4 also depicts that this ratio holds despite the different magnitude levels of plus and minus-strand RNA in the high and low permissive cells. From kinetic data, we can see that initially, the rate of rise of minus-strands is different in both cases. In the high permissive cell case, the rise is faster, while that one in the low permissive case is slower. As it has been seen in the plus-strand case, the saturation time is also affected. The model correctly reproduces the initial phase, the magnitude level as well as the saturation time (Figure 3.4 A, B). A saturation or steady state attainment time is similar to the plus-strand RNA case.

### 3.3.3 Polyprotein kinetics

In this case, the output of the model,  $L$  is fitted to the kinetic data. We can see that the model fits to the data with the excellent accuracy (Figure 3.4 A and B). From the plot, we can observe that the initial rise in the polyprotein dynamics is due to the translation of transfected positive-strands, which is followed by the decline with period of till  $t = 8$  hours in the high permissive cells and  $t = 12$  hours in the low permissive cell case (Figure 3.4 A, B). The decline is caused because the degradation of positive-strands is overwhelming the production. In this case, the synthesis of viral proteins is dominated by the degradation. The equilibrium points at  $t = 8$  and  $t = 12$  hours demonstrate the balance between production and degradation. This state is followed by an exponential-like increase which lasts till replication dynamics attain a steady-state. The model is able to capture both kinetics from high and low permissive cells with the excellent accuracy. Interesting thing is that from the same plot, we can observe that the lowest peak of polyprotein coincides with the lowest peak of plus-strand RNA, which indicates a strong correlation between them. At  $t = 8$  hours, both the plus-strand RNA and the polyprotein dynamics attain the lowest peak in the high permissive cells. The same can be observed from the low permissive cells, which occurs at  $t = 12$  hours. Another observation is that the

saturation time in the plus-strand RNA, minus-strand RNA and polyprotein occurs approximately at the same time both in the high and low permissive cells.

### 3.3.4 Steady state observations

The steady state data have been integrated into the optimization in terms of constraints.

First, a ratio  $R_P^{tot}/R_M^{tot}$  which obtained from (3.35) and (3.36) satisfies 10 : 1 ratio by giving an excellent fit in two cell lines (Figure 3.4 C).

Total plus-strand RNA in the replication vesicles, we computed by summing up all components which contain the plus-strands,

$$R_P^{RV} = R_{Ip} + R_{ds} + R_{Ids} + R_P \quad (3.37)$$

Resulting ratio of  $R_P^{RV}/R_M^{tot}$  is also in excellent agreement with the experimental data giving 6 : 1 ratio. (Figure 3.4 C middle bar).

Lastly, we compared the plus-strand RNA levels both in the cytoplasm and RV. Since we have already defined the total plus-strands in the replication vesicles, we will define the total numbers in the cytoplasm. This can be defined as,

$$R_P^{CYT} = R_P^{unp} + R_P^{cyt} + T_c \quad (3.38)$$

The resulting ratio  $R_P^{CYT}/R_P^{RV}$  from the simulation can excellently reproduce an observed value of 1 : 1 ratio (Figure 3.4 C third bar).

Eventually, we conclude that all steady state ratios reproduced from the model are in close agreement with the data.

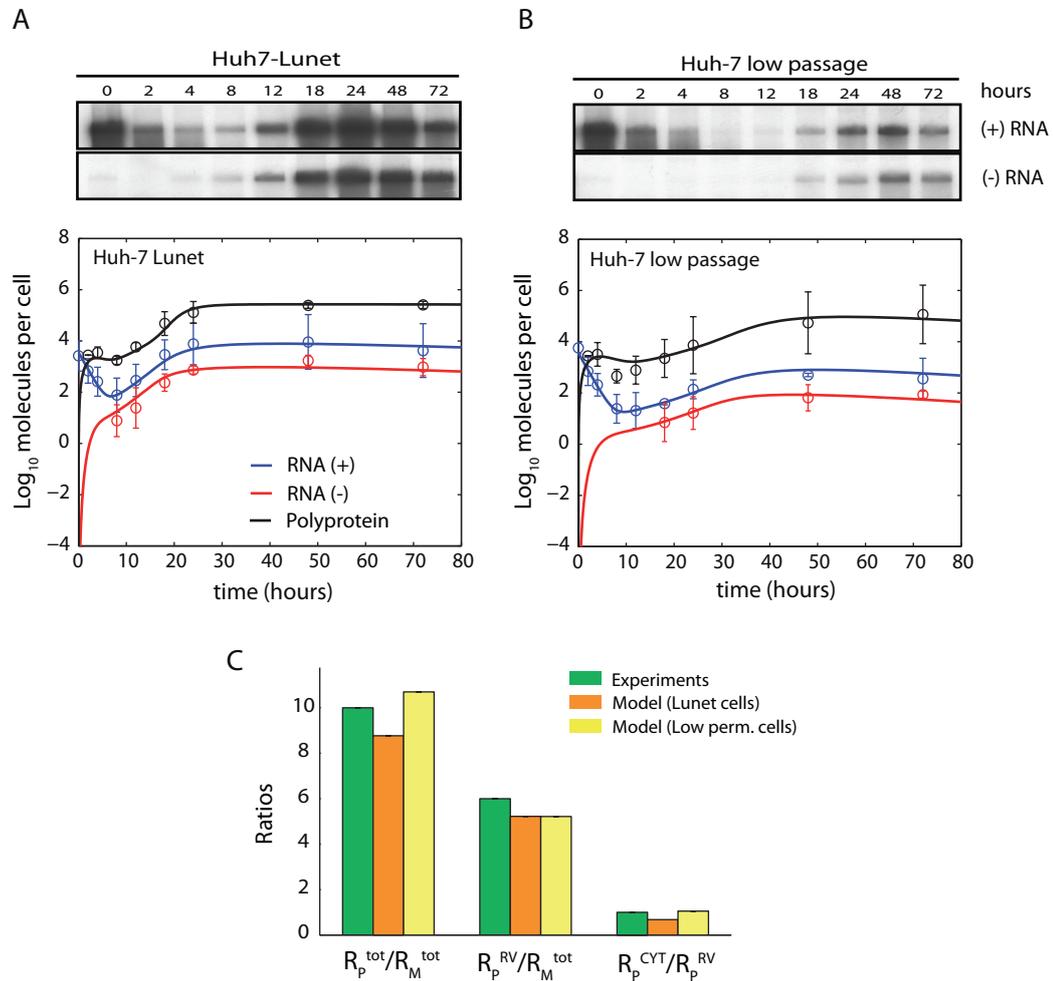


Figure 3.4: Model calibration against experimental data. The data has been normalized by the total number of transfected cells. The quantities represent the mean molecule numbers in a single cell. The circles and the error bars represent the mean and the two standard deviations of the experimental data (mean  $\pm 2*SD$ ). **(A)** The model fitting to the kinetic data from the high permissive Huh7 Lunet cells. **(B)** The model fitting to the kinetic data from the low permissive Huh7 cells. **(C)** The model fitting to the data reflecting the ratios at steady state.  $R_p^{tot}/R_M^{tot}$  ratio represents the ratio of total plus to minus-strand RNA per cell.  $R_p^{RV}/R_M^{tot}$  ratio represents the ratio of plus to minus-strand RNA in the RV.  $R_p^{CYT}/R_p^{RV}$  ratio shows the ratio of plus-strand RNA in the cytoplasm to plus-strand RNA in the RV.



# Chapter 4

## Model Validation and Predictions

### 4.1 Model validation

Model validation is a test of how well model predictions match the set of independent observations. In other definition, the model validation is the process of evaluating model performance against the primary design goal [9]. Model validation aims at increasing our confidence in the constructed model and this step is required before proceeding with the model prediction and analysis [29]. We are interested to check whether our model sufficient to explain the experimental data. In other words, we asked whether the model can also explain an independent data which is not used in the model calibration. In this section, we performed our model validation with the experimental data from two independent processes with the best parameter values obtained from the model calibration (Table 3.1).

#### 4.1.1 Replication deficient virus dynamics

In order to validate our model, we used the kinetic data obtained from a replication deficient virus (Figure 4.1 A). This data is obtained from the JHF1-based virus which lacks an GDD motif of NS5B polymerase [78],[19] that is not able to replicate. GDD motif is responsible for the active site of NS5B polymerase. When this active site is removed, the virus is not able to replicate and its replication dynamics follows a decay. Particularly, an inactive viral polymerase is not able to catalyse the reaction responsible for a synthesis of new viral RNAs. Such an observation has been measured in the high and low permissive cells. Importantly, this data indicates that there is no difference in the permissiveness in Huh-7 cells [78]. This information is used further, when we will discuss the permissiveness later more in details.

We examined this process by using our model. Using the model calibration, we

estimated the kinetic rates, which best explain the experimental data. We assume that these parameters represent the true dynamics of viral replication and further examine the dynamics of replication deficient virus by using these parameters. Since the viral polymerase, in this case, is inactive, we set the kinetic rate,  $k_{Pin} = 0$ . By doing this, we neglect the processes inside the RV and see that the replication dynamics is totally determined by the processes in the cytoplasm (Figure 4.1). In this case, the processes are governed by  $R_P^{unp}$ ,  $R_P^{cyt}$ ,  $T_c$ ,  $P$  and  $E^{cyt}$ .

Because the polymerase is inactive, there is no synthesis of minus-strand RNA undergoes. As we see from the Figure 4.1 A, because of no production of new RNA, the plus-strand RNA level decreases exponentially. In contrast, initially there is a fast increase in the polyprotein numbers, which follows by an exponential-like decrease. Since the production stops at the replication stage, initially transfected plus-strand RNA participate in the polyprotein translation, which can be seen from the initial increase in the polyprotein levels. Once the numbers of plus-strands eliminate due to degradation, then the degradation becomes dominant and the decline in the polyprotein levels is observed. Comparing the model dynamics with the experimental data demonstrated that the model can excellently explain the replication dynamics (Figure 4.1). Additionally, this supports how well the model parameters have been estimated and how correctly the model exhibits the true dynamics.

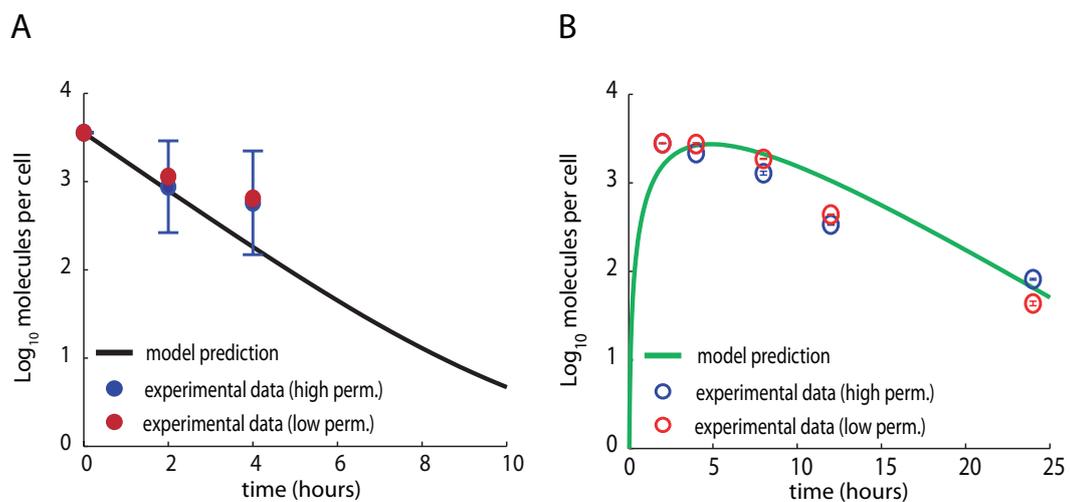


Figure 4.1: Model validation using a replication deficient virus. The circles and the error bars represent the mean and the two standard deviations of the experimental data (mean  $\pm 2*SD$ ). (A) Plus-strand RNA dynamics. (B) Polyprotein dynamics. The solid lines indicates the model predictions with the parameters that have been obtained from the model calibration using only the replication competent virus. The model captures the experimentally observed dynamics in both the plus-strand RNA and the polyprotein by setting the formation rate of the replication vesicles,  $k_{Pin}$  to zero. Note that both the experimental data in (A) and (B) show no significant difference between the high and low permissive cells, indicating that differences between the two cell lines occur post protein translation.

### 4.1.2 Initiation hampered HCV RNA dynamics

In this section, we discuss about the validation of our model with inhibited HCV dynamics.

The 5' nontranslated region (NTR) and the X tail in the 3' NTR are the least parts of the HCV genome and play important role in the initiation of RNA synthesis. In [19], by using the subgenomic replicons of the HCV isolates Con1 (genotype 1) and JFH1 (genotype 2), the authors characterized the genotype specificities of the replication signals contained in the nontranslated regions (NTRs). They demonstrated that the replacement of the JFH1 5' NTR and X tail with the corresponding Con1 sequence resulted in a significant decrease in a replication efficiency. Specifically, the exchange of the X tail reduced the minus-strand synthesis, while a substitution of the 5' NTR impaired the synthesis of progeny plus strands. Moreover, the authors analyzed recombinant NS5B polymerases of both isolates and found some genotype specific template preference for the 3' end of plus-strand RNA in vitro. In order to address a genotype specificity, they constructed a series of intergenotypic replicon chimeras and observed that the NS5B recognizes the genotype specific signals in nontranslated regions. The NTRs of plus-strand RNA viruses contain the signals important for the initiation of RNA synthesis. The 3' X region of the HCV genome has been shown to contain signals for the initiation of minus-strand synthesis. Nascent plus-strand synthesis is thought to be regulated by the 3' end of the minus-strand RNA. The reduction of replication efficiency by heterologous NTRs is thought to be due to a specific impairment of the initiation of minus-strand synthesis in the case of the 3' X tail and of plus-strand synthesis in the case of the 5' NTR. To verify this hypothesis, the authors transfected Luc JFH replicons (polyprotein) with an authentic or a Con1-derived 5' NTRs and/or 3' X tails into Huh7 cells and analyzed the plus- and minus-strand synthesis at different time points after transfection. They observed that the replication efficiency was highest for a replicon harboring authentic NTRs (5' X JFH), whereas the replication efficiency in other replicons harboring heterologous NTRs was impaired (Figure 4.2 A).

We were curious to examine whether our model can reproduce this observation. To perform this task, we used the parameter values obtained from the model calibration. In a case of 3' X tail exchange, the parameters  $k_{Pin}$  and  $k_3$  are responsible for the initiation of the plus-strand synthesis. Here, an expectation will be whether a change in this parameters can reproduce the inhibited dynamics which was observed experimentally. Therefore, to verify this hypothesis, by setting  $k_{Pin}$ ,  $k_3$  and a scale factor for polyprotein free, we estimated these parameters by fitting the

model to the experimental data (Table 4.1). Here, the scale factor is used to convert luciferase activities to the molecule numbers. Because the luciferase activities may change from one experiment to another, we assumed that this scale factor is different from that which has been used in the model calibration. As a result, we see that the model nicely captures the kinetic data reflecting the dynamics of polyprotein marker (luciferase) in a case of 3' X tail exchange (Figure 4.2 B (magenta)). From the fitting, we found out that the estimated parameters are significantly smaller than the parameters in control case (see Table 4.1). Simply,

$$k_{Pin}^{3'X} < k_{Pin}^{wild} \quad \text{and} \quad k_3^{3'X} < k_3^{wild}. \quad (4.1)$$

In the case of 5' NTR exchange, we followed the similar procedure. In this case, one would expect that the change in the minus-strand intermediate complex formation rate  $k_5$  is responsible for the reported observation. We performed the fitting of the model to the kinetic data, where only  $k_5$  and the scale factor have been estimated. Results demonstrated that the model can reproduce the data (Figure 4.2 B (cyan)). Model suggests that the replication efficiencies which are presented in terms of polyprotein marker can be explained by tuning this parameter alone, given the best fit values from the Table 3.1. In a case of 5' NTR exchange, the model predicts that the following condition is required (see Table 4.1),

$$k_5^{5'NTR} < k_5^{wild} \quad (4.2)$$

where, the parameters  $k_5^{5'NTR}$  and  $k_5^{wild}$  are referred to the 5' NTR exchanged and wild type cases, respectively. In this case, fulfilling this condition is sufficient to explain the inhibited replication dynamics.

Further, fulfilling the conditions (4.1) and (4.2), explains the experimentally measured steady state observations (Figure 4.2 B). When 3' X tail has been exchanged, about 2 fold increase in the ratio of plus-strand to minus-strand was observed. Fulfilling (4.1) excellently reproduces kinetic as well as steady state observations. In the case, when 5' NTR has been exchanged, a profound decrease about 2 fold in the ratio of plus-strand to minus-strand was observed. As we mentioned already, fulfilling (4.2) excellently reproduces kinetic measurements and correctly predicts the changes in the ratio.

Table 4.1: The parameter values in the case of inhibited HCV replication. The wild type parameters are from Table 3.1 that have been obtained in Parameter Estimation section.

Parameter	Wild type value	Estimated value
$k_{Pin}$	2.07e-6	2.62e-7
$k_3$	1e-4	1e-8
$k_5$	10	0.61
$p_{scale}$	2800	1600

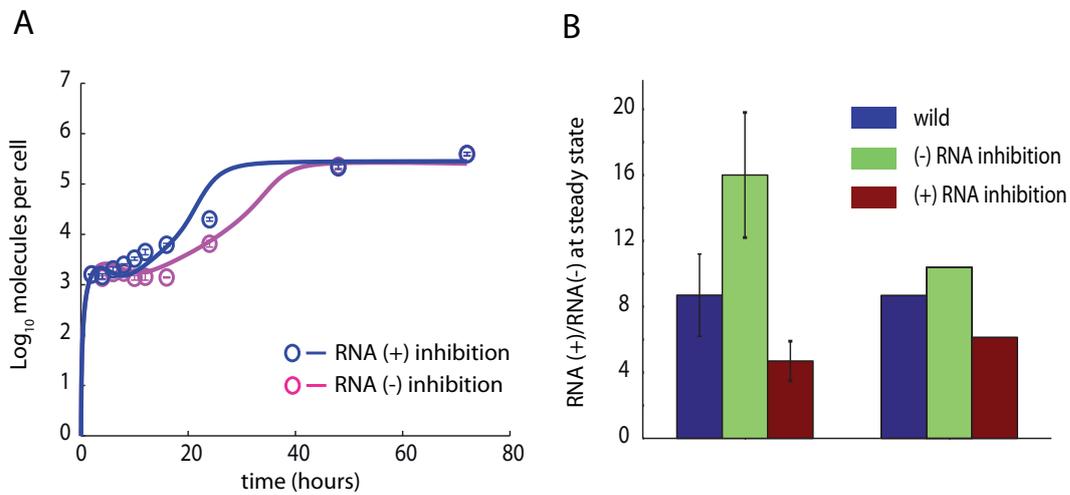


Figure 4.2: Model validation by the synthesis inhibited virus dynamics. **(A)** The model fitting to the polyprotein data from JFH replicons. The data shows that the replication is inhibited when NTRs regions are exchanged. The circles and the error bars represent the mean and the two standard deviations of the experimental data (mean  $\pm$  2\*SD). **(B)** The results demonstrate that the model is able to reproduce the observed changes in the ratio of plus-strand RNA to minus-strand RNA (data is obtained from [19]).

## 4.2 Model predictions

In this section, we will discuss about the predictions derived from the model analysis. First, we will start by overviewing the hypotheses that are used in our analysis.

First important issue, we are going to address is the permissivity of Huh-7 cells for HCV replication, which we will discuss in Subsection 4.2.1. The interesting observation is that the genetically identical Huh-7 cells reflect a different replication course for viral replication. Many studies reported different cellular factors which may have role in HCV replication (reviewed in [32], [148]). However, whether they limit the viral replication or not it is not known. Further, how the interactions of cellular factors with the virus determine a course of viral replication is poorly understood.

In Subsection 4.2.2, we will examine the hypothesis about a property of replication vesicles (RV). It was reported that the HCV RNA and the non-structural proteins (NS) that are associated with the RV are resistant against nuclease and protease treatment [120]. This leads to the conclusion that the RVs could have a protective property. In the light of this observation, we are going to address a question if how this protective property of RV associated with the HCV replication?

In Subsection 4.2.3, we are going to discuss the steady state predictions derived from the model. We are going to discuss these predictions in comparison to the experimental observations from [120].

In Subsection 4.2.4, we will discuss about an influence of transfected HCV RNA on the initial HCV replication dynamics.

### 4.2.1 Permissiveness of HCV RNA replication in different cells

The time course kinetic data measured in clonal cells demonstrated that HCV replication exhibits different dynamics (Figure 4.3). By clonal cells, we mean the high permissive Huh7 Lunet and the low permissive Huh7 cells. This clearly shows the cell permissiveness for HCV replication. Here, one of the central challenges is to understand how the cells respond to virus invasion and what processes lead to the observed difference in virus replication. Binder *et al.* [20] suggested that the differential expression in the cellular factors might cause the observed difference in virus replication. Recent developments in HCV research resulted in a huge number of possible candidates which have been shown to interact with this pathogen (reviewed in [32], [148], [99]). These candidates were reported to interact from the cell receptors

till the cellular proteins which needed for virus to form particles. Much is known in an interaction level, however, less information is available at the functional level. How does the interaction of virus and host define the replication course? Where and when do possibly virus and host interact? What is possible underlying mechanisms for virus and host interactions? Addressing these questions may help us to elucidate a mystery of HCV replication and its interaction with the host. Use of mathematical modeling can help us to address these issues. Therefore, we use our established model to examine this problem.

From the Figure 4.3, we can see that the course of replication dynamics in both cells is different. Simply, what we see is that the replication in the high permissive cells is more efficient than that in the low permissive cells. In all cases, we observe that in the high permissive cells the saturation time is shorter and the magnitude of replication is higher than that in the low permissive cells. This shows that the virus within the first cell replicates faster and is produced more in numbers. In contrast, the low permissive cells are less preferable for virus which can be seen from the delayed saturation time and the reduced steady state level. The number of possible hypotheses can be drawn to explain this observation.

The first hypothesis which can be derived is about a cell to cell variability. Even the clonal cells may vary in a content of cellular proteins which can play a key role in the virus replication. Since the HCV strain used in these studies is the same virus, JFH1, the difference in replications may only stem from the cellular host factors which are important in replication. When we mean a host factor, then it can be any host factor which can participate in the virus replication. Therefore, when constructing a mathematical model we assumed a role of two host factors, the ribosomes and the cellular factor which interact with the virus at the translation and formation stages of replication vesicles, respectively. Any of these factors may limit the virus replication and cause a difference in replication. To check whether this is the case, we fitted our model to the data by assuming the different initial conditions for the ribosomes and the host factor at the formation stage of replication vesicles. For the high permissive cells, we assume  $R_{ibo}^{high}(0)$  and  $H_F^{high}(0)$ , and for the low permissive cells,  $R_{ibo}^{low}(0)$  and  $H_F^{low}(0)$ . Results demonstrated that the difference replication is best explained by the different conditions for the host factor which participates in the replication vesicles formation (Figure 4.3). The simulations showed that the difference in a ribosome content cannot explain the diverse course of replication (Figure 4.4 A). Use of replication deficient data suggested that this difference in replication cannot stem from the translation process, and as we see our model simulations support this observation (Figure 4.4 A).

Additionally, we examined the other possibilities which may explain the difference to ensure if the explanation with the host factor is unique.

One possible hypothesis in this case is the different degradation due to a variation in a cellular environment. Therefore, we assumed different degradation rates in two cells. We checked this also for different degradation rates in the cytoplasm and the replication vesicles. As a result, we showed that neither of these factors can explain the observed difference in the replication (Figure 4.4 B and C).

Another hypothesis which could be addressed is whether the difference in the transfection setting can explain the permissivity in clonal cells. It is usually the case in transfecting the virus into the cell, where the numbers of viral strains,  $R_P^{unp}(0)$  cannot be controlled and the difference in numbers may cause the difference in replication. However, we found out that the different values of  $R_P^{unp}(0)$  cause the difference in the saturation time whereas they did not affect the magnitude levels (Figure 4.4 D).

It appears that the cellular host factor which participates in the formation of replication vesicles is the only factor which can explain the observed difference in HCV replication. Importantly, the host factor difference excellently explains the difference in all three variables (Figure 4.3). The model predicts that about 9-10 fold difference in the host factor amount explains the cell permissivity for replication best (Table 3.1).

Further, we examined how this cellular host factor affects the virus concentration. Therefore, we analyzed if how the replication course changes with respect to the host factor amount. Results revealed that there is a linear correlation between the plus, minus-strand RNA and the host factor (Figure 4.5). It demonstrates that the relation between them can be quantified with a simple relation, in the form of  $f(x) = ax + b$ . It shows an easy controllability of the virus when we know this host factor. This also shows that this host factor specifically limits the level of viral RNA, whereas the viral protein levels are less affected (Figure 4.5). The polyprotein and the host factor levels have a nonlinear-like relation, which shows that the polyprotein levels saturate at the less amount of host factor compared to viral RNA. From Quinkert *et al.* [120], we know that about 50% of plus-strand RNA and 95% of viral proteins reside in the cytoplasm. Since the host factor which explains the difference in replication participates in the replication vesicles formation, about half of plus-strand RNA and only  $< 5\%$  of viral protein are directly dependent on the host factor. The difference in dependencies cause the linear relation for the plus-strand RNA and the nonlinear relation for the viral proteins.

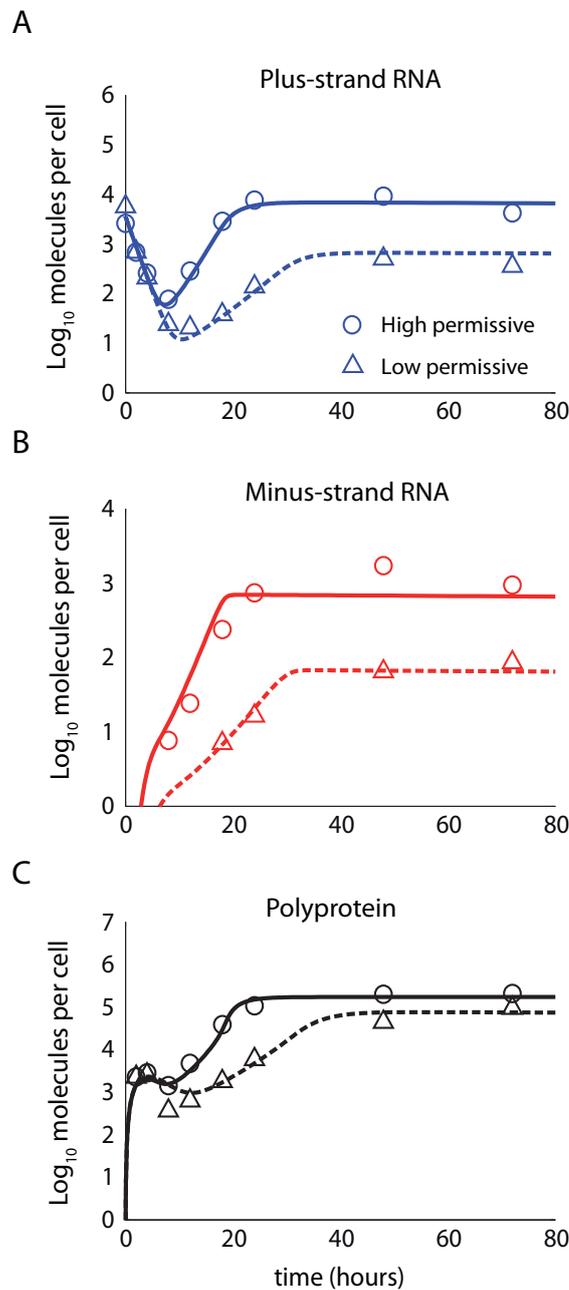


Figure 4.3: The difference in the replication dynamics in Huh7 clonal cells can be explained by the differential expression level of the cellular host factor. The circles and the triangles denote the experimental data from the high and low permissive cells, respectively. (A) Total plus-strand RNA (B) Total minus-strand RNA (C) Polyprotein. The results have been obtained by fitting the model to the experimental data simultaneously only with the different host factor values. The parameter and different host factor values are given in Table 3.1.

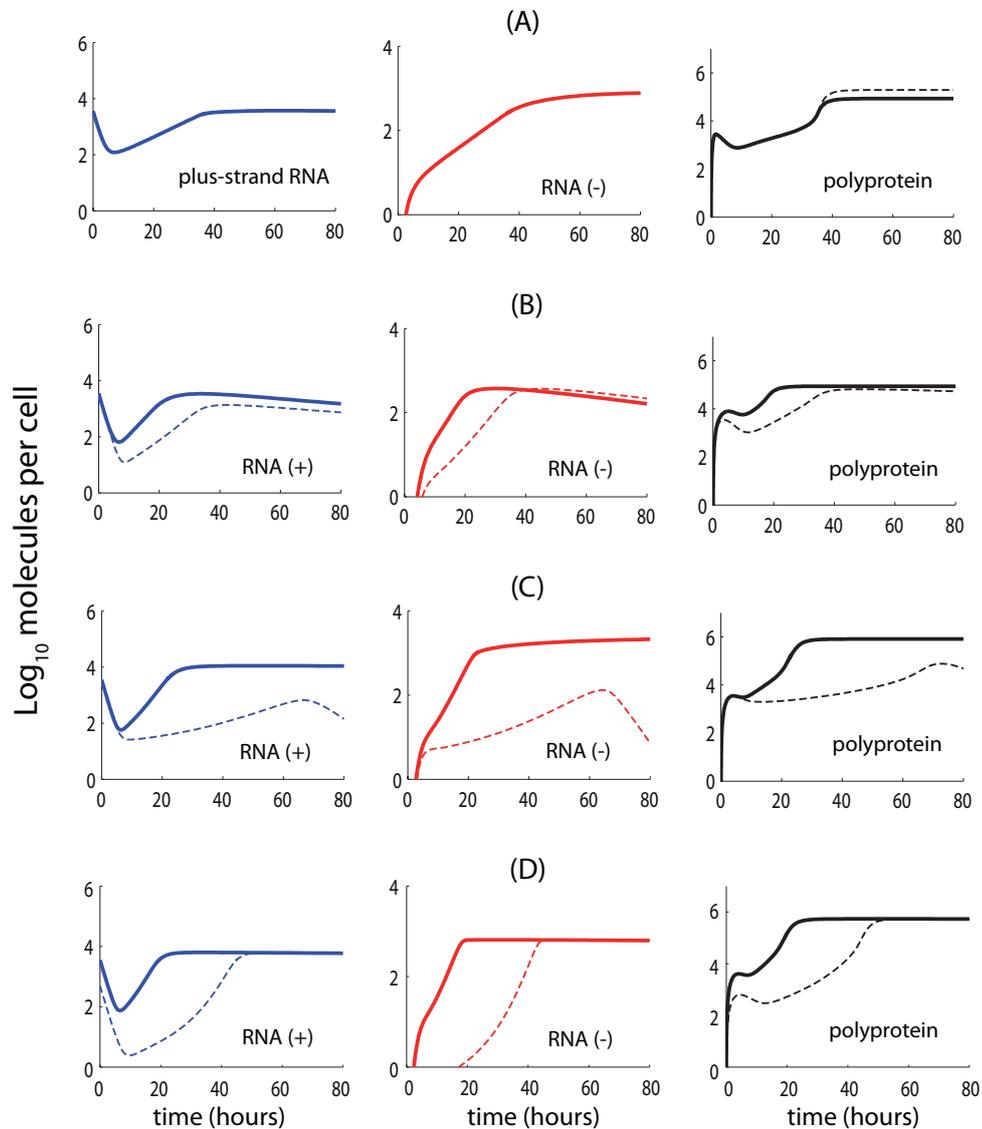


Figure 4.4: Effect of different factors on the plus-strand RNA, minus-strand RNA and polyprotein. Results have been obtained by fitting the model to the data with the different values of the ribosomes, the degradation rates in the cytoplasm and the replication vesicles, and the transfected HCV RNA. **(A)** Effect of the different ribosome numbers. It is clearly seen that the ribosomes affects the polyprotein, whereas the plus and minus-strand RNA are not affected. **(B)** Effect of the different degradation rates in the cytoplasm. The difference in the cytoplasmic degradation affects the replication only initially, while the later dynamics are less affected. **(C)** Effect of the different degradation rates in the replication vesicles (RV). The difference caused by the degradation rates in the RVs is profound, but the later dynamics are not at steady state anymore. **(D)** Effect of the different transfected HCV RNA numbers on the replication dynamics. Only the saturation time is severely affected, while the magnitude level is not influenced.

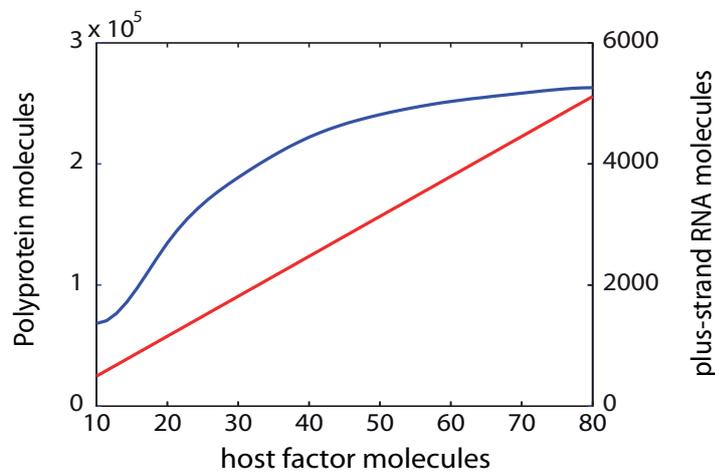


Figure 4.5: The correlation of the virus and the host factor which explains the replication permissiveness. This correlation is calculated at steady state. The blue curve indicates the polyprotein and the red curve indicates the plus-strand RNA with respect to the host factor amount. In the range between 10 and 80 host factor molecules, the plus-strand RNA steady state levels respond linearly to concentration changes of the host factor. The polyprotein levels show a bi-phasic steady state behavior, with an exponential response up to increasing host factor amounts to about 70 molecules, by showing a saturation thereafter.

### 4.2.2 Degradation inside the replication vesicles

The role of replication vesicles in viral RNA synthesis is not well understood. It is currently hypothesized that it may include: the physical support and organization of the RNA replication complex [79]; the compartmentalization and local concentration of viral products [136]; tethering of the viral RNA during unwinding; provision of lipid constituents important for replication; and the protection of the viral RNA from double-strand RNA-mediated host defences or RNA interference.

The results published in [120] have demonstrated that the HCV plus- and minus-strand RNAs are highly resistant to the nuclease treatment. Additionally, the same studies revealed that  $< 5\%$  of the NS5B polymerase molecules are the protease resistant. These findings indirectly indicate the protective role of replication vesicles. We were interested to see how this protective behavior of the replication vesicles is linked to the HCV replication. Simply, we ask whether this linkage can be elucidated using our model. Since the replication vesicles remain to be a black box, meaning that they are still not accessible for the experiments, useful information or a prediction can be drawn from the modeling. Therefore, in this case, we used replication model which is calibrated and validated with the experimental data to analyze this problem.

For simplicity, we fixed the same parameter values for the degradation of the viral components in replication vesicles as we did in the model calibration.

$$\mu_{VMS} = \mu_{Ip} = \mu_{ds} = \mu_{Ids} = \mu_E = \mu_P \quad (4.3)$$

After we introduce a quantity  $s$ , which in turn describes the ratio of the degradation rates of viral components in the cytoplasm and replication vesicle or simply,

$$s = \frac{\mu_{cyt}}{\mu_{VMS}} \quad (4.4)$$

Here,  $\mu_{cyt}$  accounts for the degradation rates  $\mu_P^{cyt}$  and  $\mu_{Tc}$ . From the model calibration, we have seen that these kinetic rates differ not so much.

After we analyzed the model behavior w.r.t. the change in the ratio  $s$ . For this purpose, we fixed the estimated value of  $\mu_{cyt}$  and changed  $\mu_{VMS}$  and observed the model output. The best fit has been obtained by returning  $s > 1$  or  $\mu_{cyt} > \mu_{VMS}$ . As the result, we observed that the replication course is successful for the large values of  $s$  (Figure 4.6). Once the ratio was set to a smaller value, the replication started to show a decline. When the ratio  $s$  was set to 1, so the degradation rates are

equal,  $\mu_{cyt} = \mu_{VMS}$ , the model predicts a complete clearance of the minus-strands by  $t = 20$  hours, whereas the plus-strands level get cleared about  $t = 25$  hours (Figure 4.6 A and B). This demonstrates that the minus-strands are more susceptible to the change in the ratio  $s$  and protected by the vesicles from high degradation. The polyprotein marker level was decreased considerably by showing a drastic decline which predicts a complete clearance about  $t = 50$  hours (Figure 4.6 C).

The non-structural protein, NS5B level is less susceptible against  $s$  than the polyprotein levels by getting cleared after  $t = 80$  hours (Figure 4.6 D). From experimental observations, we know that the degradation rate of polyprotein,  $\mu_L$  is higher than the degradation rate of NS5B,  $\mu_E^{cyt}$ . Therefore, the clearance time of NS5B protein is longer than that one of polyprotein marker. Interestingly, in the case when the degradation rates are equivalent, the production of viral components still proceeds but at a slower rate and when the balance between the production and the decay has reached a critical threshold, the decay becomes dominant and the replication started to decline.

With this simple analysis, we showed that how the protective property of replication vesicles can be explained by the different degradation rates in the cytoplasm and the replication vesicles which is consistent with the observed protective property of replication vesicles. These observations demonstrated that for a successful replication, the degradation in the replication vesicles should be much smaller than that in the cytoplasm, thereby supporting the theory about the protective role of replication vesicles. As we see from the plot, all three variables are sensitive to the small changes in the ratio  $s$ . From the parameter estimation, we revealed that the confidence interval for the degradation rate inside the replication vesicles is very narrow, therefore, the small changes in this rate cause severe changes in the replication dynamics.

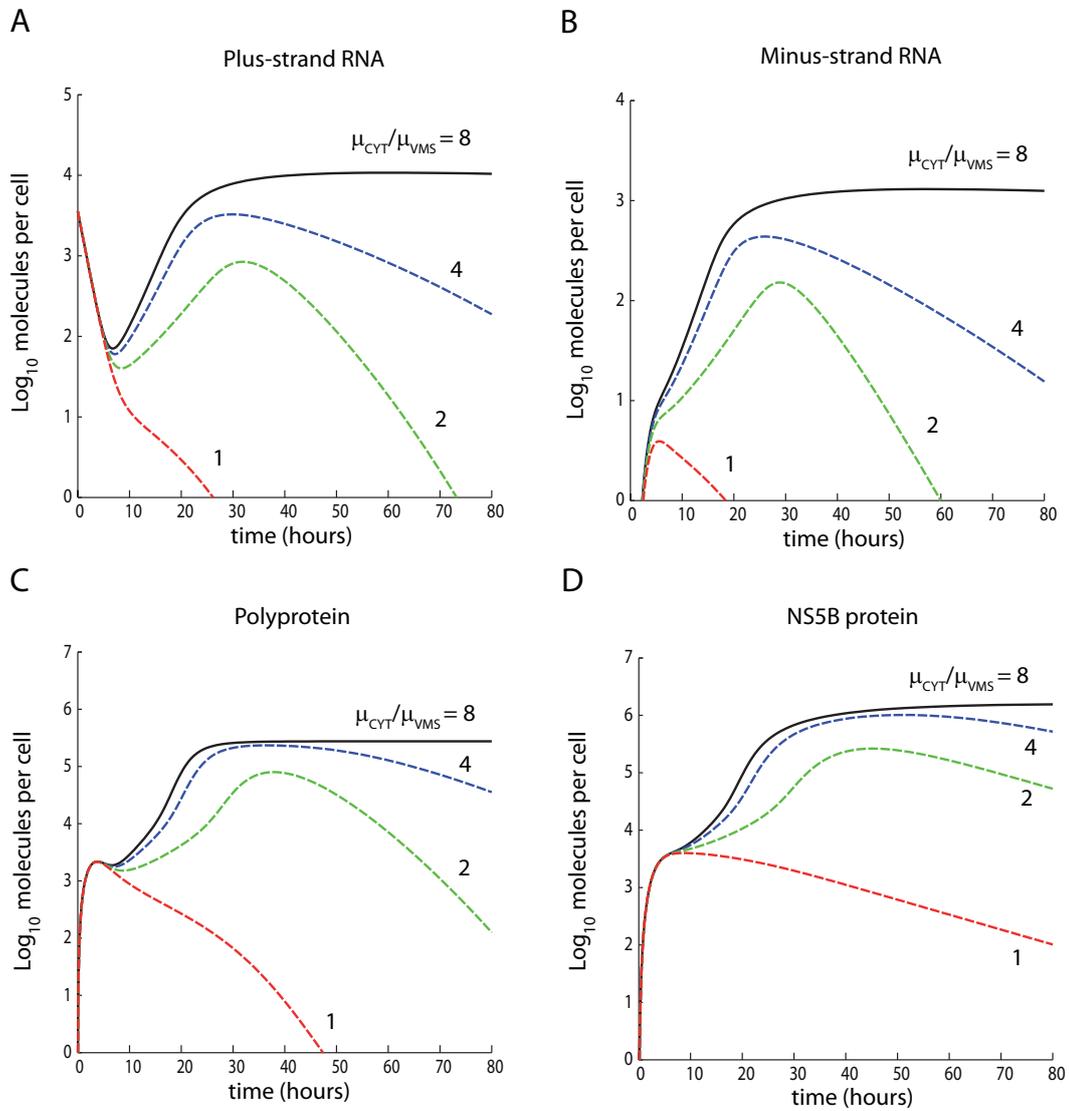


Figure 4.6: The replication dynamics w.r.t. the different values of  $s = \mu_{\text{cyt}}/\mu_{\text{VMS}}$  (representation is in logarithmic scale). This figure highlights how the difference in the degradation rates in the cytoplasm and the replication vesicles affect the replication dynamics. The optimal and successful replication is obtained with  $s = 10$ , which means that the degradation in the cytoplasm is about 10 fold higher than the degradation in the replication vesicles.

### 4.2.3 Steady state predictions

Experimental observations revealed that at steady state the level of viral proteins accumulates up to  $800000 - 2 \times 10^6$  molecules per cell [120]. Available data indicates that there are 2 to 3 orders of magnitude abundant NS5B proteins than the viral RNAs.

We performed a simulation with the parameter values which best explain the experimental data. Particularly, we have calculated the predicted level of NS5B proteins, the plus and minus-strand RNA from the model and have taken the respective ratios at  $t = 72$  hours after transfection. The total amount of NS5B proteins can be calculated by solving and summing the equations for  $E^{cyt}$ ,  $E$ ,  $R_{Ip}$  and  $R_{Ids}$  which give the total amount of NS5B proteins,

$$E^{total} = E^{cyt} + E + R_{Ip} + R_{Ids} \quad (4.5)$$

Further, the derived ratios from the simulation have been compared with the experimentally observed ratios. As a result, we have seen that the model can reproduce the experimentally observed results (Figure 4.7). Figure 4.7 depicts the results from the high permissive cells (HPC). In particular, our model predicts about  $1.5 \times 10^6$  molecules for the HPC, which is consistent with the experimental observations. For the LPC (low permissive cells), model predicts about 3-4 fold less proteins than the HPC. The reason why the model predicts less protein numbers is due to host factor amount. In addition to the cellular ribosomes, the host factor also significantly affects the protein levels. As reported, the polyprotein levels are positively correlated with the host factor amount (Figure 4.5).

Overall, these results demonstrate that the model is correctly structured so that it can reproduce a real data which was not used in the model calibration. In [48], it was reported that the amount of ribosome complexes is the main parameter which significantly affects the NS5B level. In our model, it also preserves its function and its estimated value is consistent with the reported value from [48].

Furthermore, it was reported that at steady state about  $< 0.1\%$  of NS5B proteins are enzymatically active. This shows that only a minor fraction of polymerase is active that can be used as a replication machinery. In order to calculate the active fraction of polymerase, we summed up the polymerase inside the replication vesicles,

$$E_{active} = R_{Ip} + E + R_{Ids} \quad (4.6)$$

When calculated this number from the model with the calibrated parameters, we

found out that they consist about  $< 0.1\%$  which is consistent with the experimentally reported value (Figure 4.7 B). The model predicts that there are more active polymerase in HPC than in LPC. Together, the model predictions performed with the calibrated parameters are in a close agreement with the experimental data, thereby showing a predictive power of the established model.

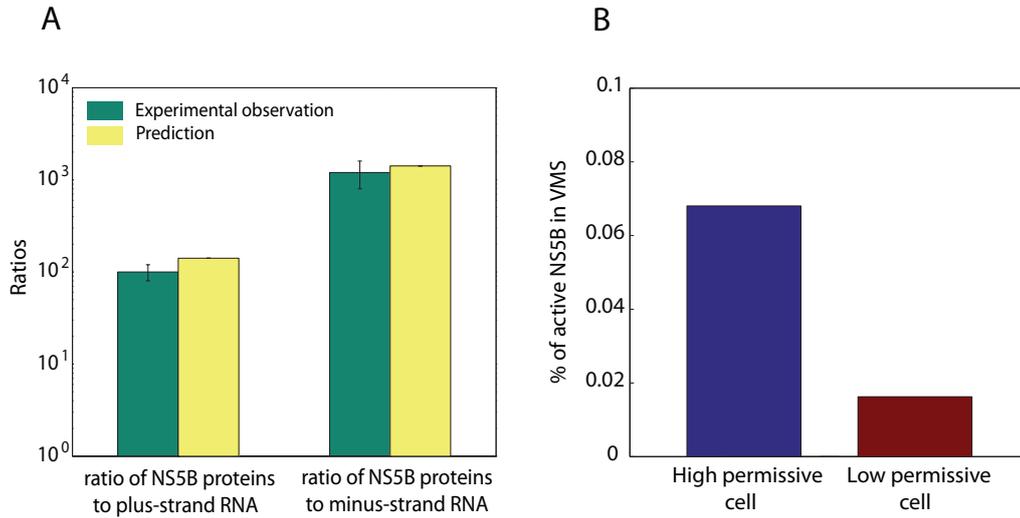


Figure 4.7: *The steady state predictions. (A) The ratio of NS5B proteins to plus and minus-strand RNA at steady state. The model predictions are in close agreement with the experimental observations. (B) The percentage of active NS5B polymerase in the high and low permissive cells. The experimental results showed that the active fraction of NS5B polymerase consists of less than 0.1% of total NS5B. The active fraction of NS5B polymerase is the number of polymerase within the replication vesicles (RV or VMS).*

#### 4.2.4 Transfected plus-strand RNA degradation

After the HCV RNA has been introduced into the cell, it interacts with the host ribosomes to form a viral polyprotein. After transfection, there is approximately an exponential decline in the HCV RNA numbers till  $t = 8$  hours in the HPC and  $t = 12$  hours in the LPC.

We revealed that an inclusion of the additional reaction (2.1) in Subsection 2.1.1, which accounts for the processing of transfected plus-strand RNA is required to explain the kinetic data, because without this step, the model failed to fit. Further, we analyzed the importance of this step in explaining the experimental data. The model results suggested that the decay which has been observed initially is mainly determined by the exponential function of the form

$$R_P^{unp} = R_P^{unp}(0)e^{-(k_0 + \mu_P^{unp})t}. \quad (4.7)$$

Figure 4.8 depicts the simulated dynamics of transfected plus-strand RNA obtained from (4.7). In particular, the model predicts that till the half life time, which is about 1 hour, about 99 % all plus-strand RNA are determined by  $R_P^{unp}$ . About  $t = 3$  hours, still 95 % are determined by the transfected plus-strand RNA. As we know from the data, the production of minus-strand and the new plus-strand RNA starts earlier, however, the degradation is still severely dominant over the production. The situation changes close to the equilibrium point around  $t = 8$  hours, where the contribution of transfected plus-strands decreases up to 10 % for HPC, whereas for LPC, this number shows about 40 %. In the light of these predictions, we can explain this process with few words.

In the experimental measurements, a large number of HCV RNA copies are transfected into the cell. Out of these transfected plus-strand RNA, only a small fraction can successfully replicate. Different factors can affect this observation. It may be that the transfected plus-strands need to adapt to the host cell environment and this adaptation filters them to those which can replicate successfully and those which just degrade. In other words, the viral plus-strand RNA,  $R_P^{unp}$  may need some time to adapt to the cell environment before it is translated by the ribosomes,  $R_{ibo}$ , and at the meantime, degrade with the rate which is higher than the rate of the processed RNA,  $R_P^{cvt}$ . Since the transfected plus-strand RNA has no feedback, it declines until it gets completely cleared from the cell. Since the decline in the transfected RNA follows an exponential decline, we can easily calculate the degradation rate and the corresponding half life. As we reported, the half life of transfected RNA is roughly about  $t = 1$  hour, which coincides with the half life of total plus-strand RNA. By fitting the model to the experimental data, we estimated that the corresponding degradation rate is about  $\mu_P^{unp} = 0.758 h^{-1}$ . In summary, we conclude that the initial decline in the total plus-strand RNA until  $t = 4$  hours is mainly determined by the decay of the transfected plus-strand RNA.

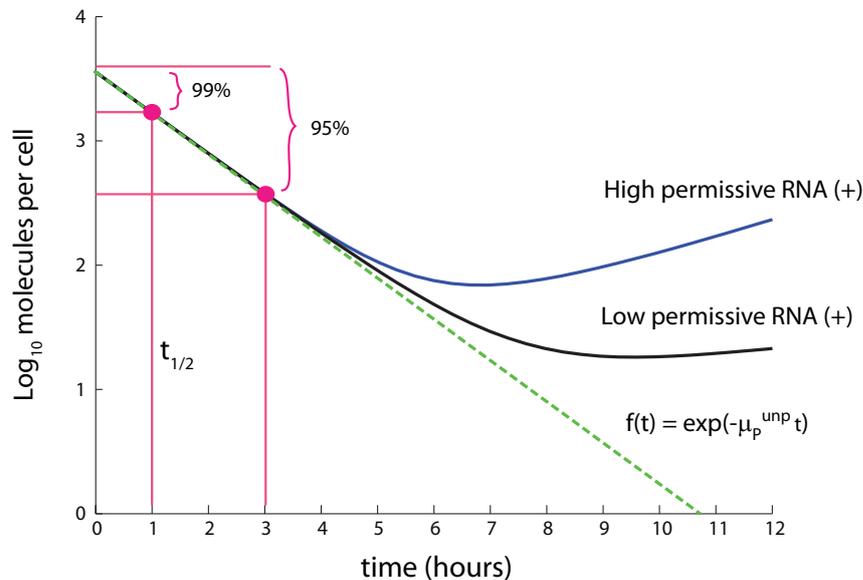


Figure 4.8: The comparison of the transfected and total plus-strand RNA. It is clearly seen that initially, the decline of the total plus-strand RNA,  $R_P^{tot}$  is mainly defined by the decline of the transfected plus-strand RNA,  $f(t) = R_P^{unp}$ .

### 4.3 Sequential binding of ribosomes to HCV RNA

In Section 2.4, we discussed about the model of sequential attachment of ribosomes to the HCV RNA. In this section, we will discuss about the results in comparison to the simplified model where 10 molecules of ribosomes simultaneously attach to the HCV RNA molecule.

The model gives the similar results to the simple case, if the following condition is fulfilled:

$$k_1 < k_1^* \quad (4.8)$$

This condition implies that the rate of the first ribosome to attach the HCV RNA molecule should be lower than that of the subsequent ribosomes. The model simulations predict very low rate for the first ribosome to attach, whereas the rate for the subsequent attachment was maximized (Figure 4.9). Figure 4.9 depicts the dynamics of the polyprotein marker (luciferase) and the plus and minus strand RNA dynamics. In order to predict the experimentally observed level of polyprotein, the model predicts the involvement of more ribosome complexes than that which was observed in the simple case. The simulations demonstrated that about 5 fold more ribosome complexes are required to obtain the experimentally observed level. When the kinetic rate  $k_1$  is slow and the ribosome complexes,  $R_{ibo}(0)$  are comparable to

the value from the simple case, we observe that the magnitude of polyprotein marker decreased by several folds (Figure 4.9, lowest dashed curve (right)).

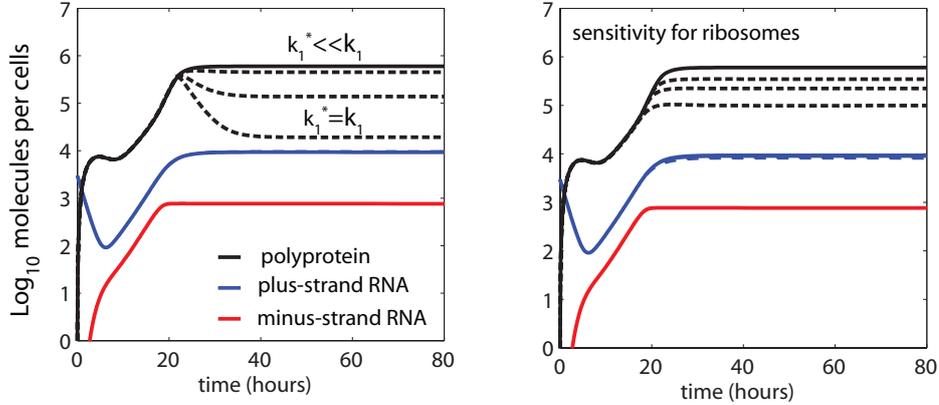


Figure 4.9: Polyprotein (black), plus strand RNA (blue) and minus strand RNA dynamics (red). The dynamics have been obtained for the different values of  $k_1^*/k_1$  (left) and the ribosome complexes,  $R_{ibo}(0)$  (right). The sequential model requires the initial binding rate,  $k_1$  to be much more slower than the rate of the subsequent ribosomes,  $k_1^*$ . Furthermore, the model predicts more ribosome complexes compared to the simple model to obtain the correct level of polyprotein dynamics.

# Chapter 5

## Sensitivity Analysis

The behavior of physical and chemical systems is affected by many parameters that characterize the system. The analysis of how a system responds to changes in the parameters is called *parametric sensitivity* [156]. In most cases, when some parameters are varied slightly, while keeping the remaining parameters fixed, the response of a system also changes slightly. However, other set of parameter combinations can cause the system to respond enormously, even if one or more parameters are varied only slightly. In this case, it is said that the system behaves in a parametrically sensitive manner [156].

Sensitivity Analysis is a method for quantifying uncertainty and its objective is to identify critical inputs, such as parameters and initial conditions of a model and quantifying how the input uncertainty impacts model outcome [84]. In this chapter, we discuss the use of sensitivity analysis to evaluate an influence of parameter variations on the model predictions by local and global sensitivity methods.

### 5.1 Local Sensitivity Analysis

A local sensitivity analysis investigates the impact on the model output based on changes in the parameters only very close to the nominal values [84]. When the input factors such as parameters or initial conditions are known with a little uncertainty, one can examine the partial derivative of the output function with respect to the input factors. In the following, we will give a brief overview about the local sensitivity analysis and for more details refer to [156], [129].

Here, we consider a system described by an ordinary differential equation,

$$\frac{dy}{dt} = f(y, \phi, t) \quad (5.1)$$

with the initial condition  $y(0) = y_0$ . Where  $y$  is the dependent variable,  $t$  is the time, and  $\phi$  represents the vector containing the  $m$  system input parameters. The function  $f \in C^1$  is a continuously differentiable function. That is, all partial derivatives of  $f$  with respect to  $x_j$ ,  $\partial f_i / \partial x_j$ , with  $i, j = 1, \dots, n$ , exist and are continuous. This guarantees that the above equation has a unique solution,  $y = y(t, \phi)$ , which is continuous in  $t$  and  $\phi$ .

Let  $\phi_j + \Delta\phi_j$  denote the change from  $\phi_j$  in the  $j$ th parameter in the parameter vector  $\phi$ . Then the corresponding solution becomes

$$y = y(t, \phi_j + \Delta\phi_j) \quad (5.2)$$

This solution is continuous in  $\phi_j$  and can be expanded into a Taylor series as follows:

$$y(t, \phi_j + \Delta\phi_j) = y(t, \phi_j) + \frac{\partial y(t, \phi_j)}{\partial \phi_j} \Delta\phi_j + \frac{\partial^2 y(t, \phi_j + \theta \Delta\phi_j)}{\partial \phi_j^2} \frac{\Delta\phi_j^2}{2} \quad (5.3)$$

where  $0 < \theta < 1$ . If  $\Delta\phi_j$  is sufficiently small, *i.e.*,  $\Delta\phi_j \ll \phi_j$ , the Taylor series can be truncated after the second term on the right hand side, leading to

$$\Delta y = y(t, \phi_j + \Delta\phi_j) - y(t, \phi_j) = \frac{\partial y(t, \phi_j)}{\partial \phi_j} \Delta\phi_j \quad (5.4)$$

where  $\Delta y$  represents the variation of  $y$  due to the change of the input parameter  $\phi_j$ , given by  $\Delta\phi_j$ . If we divide both sides of the equation (5.4) and consider an infinitesimal variation ( $\Delta\phi_j \rightarrow 0$ )

$$s(y; \phi_j) = \frac{\partial y(t, \phi_j)}{\partial \phi_j} = \lim_{\Delta\phi_j \rightarrow 0} \frac{y(t, \phi_j + \Delta\phi_j) - y(t, \phi_j)}{\Delta\phi_j} \quad (5.5)$$

This defines a *local sensitivity* of the variable,  $y$ , with respect to parameter,  $\phi_j$  [156]. Higher order local sensitivities can be defined using similar procedure.

In order to compare the computed sensitivities between the different input parameters, a *normalized sensitivity* is commonly used. The normalized sensitivity of  $y$  with respect to  $\phi_j$  is defined as follows:

$$S(y, \phi_j) = \frac{\phi_j}{y} \frac{\partial \ln y}{\partial \ln \phi_j} = \frac{\phi_j}{y} s(y; \phi_j) \quad (5.6)$$

Here, the magnitudes of the input parameter  $\phi_j$  and the variable  $y$  are normalized. If the local sensitivity  $s(y; \phi_j)$  is known, the computation of  $S(y; \phi_j)$  is straightforward.

When the system is described by dependent variables of size  $n$ ,

$$\frac{dy_i}{dt} = f(y_i, \phi, t), \quad y_i(0) = y^i \quad (5.7)$$

where  $i = 1, \dots, n$ , the sensitivity measure can be generated by the column sensitivity vector:

$$s(y_i, \phi) = \frac{\partial y_i}{\partial \phi_j} = \left[ \frac{\partial y_1}{\partial \phi_j}, \frac{\partial y_2}{\partial \phi_j}, \dots, \frac{\partial y_n}{\partial \phi_j} \right]^T = [s(y_1; \phi_j), s(y_2; \phi_j), \dots, s(y_n; \phi_j)]^T \quad (5.8)$$

By combining all the row and column sensitivity vectors, we get an  $n \times m$  sensitivity matrix, which comprises sensitivity indices as elements.

$$s(y_i, \phi) = \frac{\partial y_i}{\partial \phi_j} = \begin{bmatrix} \frac{\partial y_1}{\partial \phi_1} & \frac{\partial y_1}{\partial \phi_2} & \dots & \frac{\partial y_1}{\partial \phi_m} \\ \frac{\partial y_2}{\partial \phi_1} & \frac{\partial y_2}{\partial \phi_2} & \dots & \frac{\partial y_2}{\partial \phi_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial \phi_1} & \frac{\partial y_n}{\partial \phi_2} & \dots & \frac{\partial y_n}{\partial \phi_m} \end{bmatrix} \quad (5.9)$$

### 5.1.1 Results

We performed a local sensitivity analysis for 16 kinetic rates and 3 initial values. This analysis has been performed in Matlab using Simbiology Toolbox. Figure 5.1 shows the normalized sensitivity values for the parameters. We arranged the model parameters into five different groups, with respect to the steps in the replication cycle that they participate in:

- Cytoplasmic RNA processing:  $R_P^{unp}(0)$ ,  $k_0$ ,  $\mu_P^{unp}$ ,  $\mu_P^{cyt}$
- Protein translation and cleavage:  $k_1$ ,  $\mu_{Tc}$ ,  $R_{ibo}(0)$ ,  $k_2$ ,  $k_c$ ,  $\mu_E^{cyt}$
- Formation of the replication vesicles:  $k_{Pin}$ ,  $H_F(0)$
- Genome replication:  $k_{4m}$ ,  $k_{4p}$ ,  $k_5$ ,  $\mu_{VMS}$ ,  $k_3$
- Other:  $k_{Pout}$ ,  $\mu_L$

From Figure 5.1, we see that at the cytoplasmic RNA processing, the plus-strand RNA is very sensitive against the degradation of the transfected plus-strand RNA,  $\mu_P^{unp}$ . From the model calibration, we have seen that this parameter plays an important role and the sensitivity analysis supports this observation. The initial

value of plus-strand RNA,  $R_P^{unp}$  and the processing rate,  $k_0$  mainly affect the minus-strand RNA and polyprotein levels.

At the translation and the cleavage, the degradation rate of translation complexes,  $\mu_{T_c}$  has a significant effect on the polyprotein level but having less effect on the plus- and minus-strand RNA. In addition to this, the ribosome numbers,  $R_{ibo}$  and the translation rate,  $k_2$  have a significant effect on the polyprotein level, but also having a less influence on the plus- and minus-strand RNA. Overall, the translation rate  $k_2$  is the most sensitive parameter for the polyprotein levels.

At the replication vesicle formation, the cellular host numbers,  $H_F$  influence the levels of plus and minus-strand RNA significantly by being ranked as a third most sensitive parameter for both variables.

At the genome replication, the synthesis rate of complementary minus-strand RNA,  $k_{4m}$  is found to be very sensitive for the minus-strand RNA by being ranked as a second after the degradation rate inside the replication vesicles,  $\mu_{VMS}$ . The polyprotein levels are very sensitive against the synthesis rate of plus-strand RNA,  $k_{4p}$ . The degradation rate inside the replication vesicles,  $\mu_{VMS}$  is one of the most sensitive parameters which affects the level of all three variables. It is ranked as a second for the plus-strand RNA and the most sensitive for the minus-strand RNA. It has also the influential effect on the polyprotein level.

The export rate of new plus-strand RNA from the replication vesicles into the cytoplasm,  $k_{Pout}$  is important for the polyprotein levels. Finally, the polyprotein level is very sensitive against its degradation rate,  $\mu_L$ .

Furthermore, the parameters  $\mu_P^{cyt}$ ,  $k_1$ ,  $k_c$ ,  $\mu_E^{cyt}$ ,  $k_5$  and  $k_3$  are found to be insensitive for all three variables. The parameters  $R_P^{unp}(0)$ ,  $k_0$ ,  $\mu_{T_c}$ ,  $R_{ibo}$ ,  $k_2$ ,  $k_{Pin}$ ,  $k_{Pout}$  and  $\mu_L$  are partially sensitive either for one or two variables. The parameters  $\mu_P^{unp}$ ,  $H_F(0)$ ,  $k_{4m}$ ,  $k_{4p}$  and  $\mu_{VMS}$  are found to be sensitive for all three variables.

The local sensitivity analysis perturbs the parameters only at the vicinity of the nominal values, therefore, these parameters may be insensitive only in the neighborhood of the nominal values. To get more insights, we will perform a global sensitivity analysis and discuss it in Section 5.2.

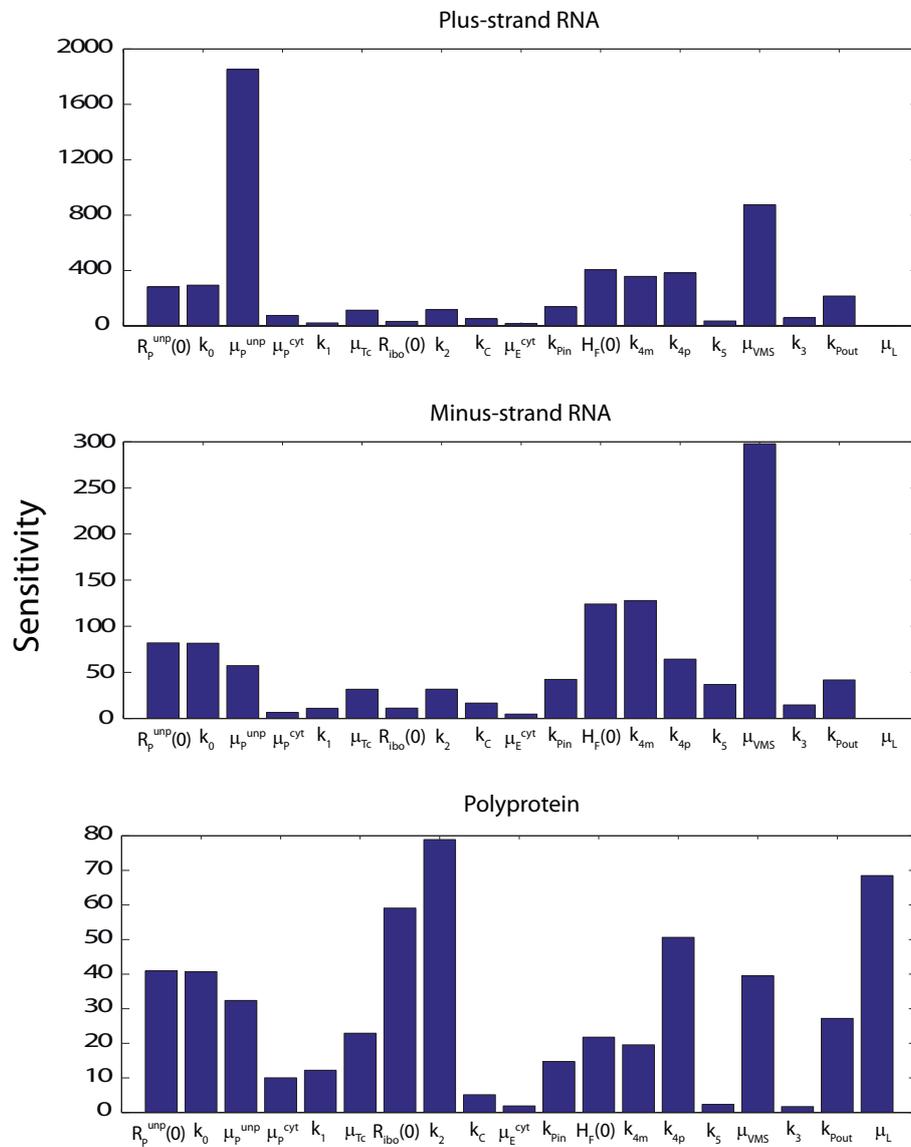


Figure 5.1: Local sensitivity analysis on the model parameters. Shown are the parametric sensitivities for the plus, minus-strand RNA and the polyprotein. The degradation rate of the transfected plus-strand RNA,  $\mu_p^{unp}$  is the most sensitive parameter for the plus-strand RNA, while  $\mu_{VMS}$  is the most sensitive parameter for the minus-strand RNA. The translation rate  $k_2$  is the most sensitive parameter for the polyprotein.

## 5.2 Global Sensitivity Analysis

The local sensitivity methods derive measures of importance by estimating the effects of infinitesimal variations of each factor having on the model output, in the region of a fixed nominal point. The local methods are widely used on steady-state models and on studies dealing with the stability of a nominal point. The local methods can only account for small variations from the nominal values and fail to capture the large variations in a parameter set.

Global sensitivity methods are advantageous when performing a full search of the parameter space, hence providing information independent of nominal points. Furthermore, the global methods can account the total uncertainty in the model output, while all parameters are varied at the same time. In addition, the global sensitivity analysis methods evaluate the effect of a parameter while all other parameters are varied simultaneously, accounting for interactions between parameters without depending on the stipulation of the nominal point. The most widely used methods in global sensitivity analysis are Fourier Amplitude Sensitivity Testing (FAST) method [31], Morris [100], Sobol [139] and Derivative based Global Sensitivity methods [68]. In this section, we are going to overview the Extended Fourier Amplitude Sensitivity Testing (eFAST) and its application to our model.

### 5.2.1 Extended Fourier Amplitude Test (eFAST)

Extended Fourier Amplitude Test was developed by Saltelli *et al.* [132],[130],[131]. Extended Fourier Amplitude Test is based on the original Fourier Amplitude Test developed by Collins and Avissar [28], Schaibly and Shuler [135] and Cukier *et al.* [30]. eFAST allows the computation of the total contribution of each input factor to the output's variance. Here, we will give a short overview of the classical FAST and the eFAST from [84] and [130].

Let us consider the model given by  $y = f(x)$ . The output  $y$  is the function of  $n$  input factors  $x = (x_1, x_2, \dots, x_n)$  through the model  $f$ . Let us assume that the domain of input factors is the unit hypercube given by

$$K^n = (x \mid 0 \leq x_i \leq 1; i = 1, \dots, n) \quad (5.10)$$

Let us assume that  $x$  is a random vector with the probability distribution function (pdf)  $P(x) = P(x_1, x_2, \dots, x_n)$ . A summary statistic is given by the  $r$ th moment of

$y$ :

$$\langle y^{(r)} \rangle = \int_{K^n} f^r(x_1, x_2, \dots, x_n) P(x_1, x_2, \dots, x_n) dx \quad (5.11)$$

Using a multidimensional Fourier transformation of  $f$ , one can perform an ANOVA-like decomposition<sup>1</sup> of the variance of  $y$  as a function of the input  $x$  in order of increasing dimensionality. In this case, it is done by computing the main effects and the interactions of any order. The curve which explores the space  $K^n$  is given by a set of parametric equations,

$$x_i(s) = G_i(\sin \omega_i s), \quad \forall i = 1, 2, \dots, n \quad (5.12)$$

where  $s$  is a scalar variable defined in the range  $-\infty < s < +\infty$ .  $G_i$  are the transformation functions.  $\{\omega_i\}$  is a set of different angular frequencies. As  $s$  varies, all the factors change simultaneously along the curve that systematically explores  $K^n$  with a different frequency  $\omega_i$ . Each  $x_i$  periodically oscillates at the corresponding frequency  $\omega_i$  depending on  $G_i$ . The high amplitude oscillations of  $y$  at the frequency  $\omega_i$  are obtained, if the  $i$ th factor has a strong influence on the output. If the frequencies  $\omega_i$  are linearly independent,

$$\sum_{i=1}^n r_i \omega_i \neq 0, \quad -\infty < r_i < +\infty \quad (5.13)$$

then (5.11) can be computed by an one-dimensional integral

$$y^{(r)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^r(x_1(s), \dots, x_n(s)) ds \quad (5.14)$$

The variance of the model is

$$D = \langle y^{(2)} \rangle - \langle y^{(1)} \rangle^2 \equiv y^{(2)} - (y^{(1)})^2 \quad (5.15)$$

and can be computed by evaluating the one-dimensional integral (5.14). By assuming  $f(s)$  within the finite interval  $(-\pi; \pi)$ , and setting  $T = 2\pi$ , one gets

$$y^r = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^r(s) ds \quad (5.16)$$

and

$$\hat{D} = y^{(2)} - (y^{(1)})^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s) ds - \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \right]^2 \quad (5.17)$$

---

<sup>1</sup>ANOVA (Analysis of Variances) for details see Appendix A.

The function  $f(s)$  can be expanded to a Fourier series

$$y = f(s) = \sum_{j=-\infty}^{+\infty} \{A_j \cos js + B_j \sin js\}, \quad (5.18)$$

where the Fourier coefficients  $A_j$  and  $B_j$  are defined as

$$\begin{aligned} A_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos js ds \\ B_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin js ds \end{aligned} \quad (5.19)$$

and  $j$  is an integer frequency.

A spectrum of the Fourier series expansion is defined as

$$A_j = A_j^2 + B_j^2, \quad j \in Z \quad (5.20)$$

By evaluating the spectrum for the frequency  $\omega_i$  and its higher harmonics  $p\omega_i$ , one can estimate the portion of the output variance,  $D_i$ :

$$\hat{D}_i = \sum_{p \in Z_0} A_{p\omega_i} = 2 \sum_{p=1}^{+\infty} A_{p\omega_i}, \quad (5.21)$$

where,  $Z_0$  is the set of all relative integers except 0.

The total variance is estimated as

$$\hat{D} = \sum_{j \in Z_0} A_j = 2 \sum_{j=1}^{+\infty} A_j, \quad (5.22)$$

The ratio

$$S^{FAST} = \frac{\hat{D}_i}{\hat{D}} \quad (5.23)$$

is the first order sensitivity index which is a fraction of total variance that estimates the main effect of  $x_i$  on  $y$ .

The classical FAST method estimates the first-order effects, while the higher-order effects are neglected. In order to compute these higher-order effects, one needs to consider the frequencies which do not belong to the set

$$p_1\omega_1, p_2\omega_2, \dots, p_n\omega_n, \quad p_i = 1, 2, \dots, \infty, \quad i = 1, \dots, n$$

These frequencies will account for a residual variance  $D - \sum_i D_i$  which include the higher order interactions among the factors that are not considered by the first order

indices. In this case, a frequency  $\omega_i$  is assigned for the  $i$ th factor and a different frequency  $\omega_{i'}$  is assigned to all the remaining factors. By evaluating the spectrum at the frequency  $\omega_{i'}$  and the higher harmonics  $p_i\omega_i$ , one can estimate a partial variance  $D_{-i}$  which includes all the effects of any order except  $i$ .

The total-order variance is computed as

$$D_{T_i} = D - D_{-i} \quad (5.24)$$

This means that  $D_{T_i}$  is calculated as the remaining variance after the contribution of the complementary,  $D_{-i}$  is subtracted.

The total-order sensitivity indices are computed by

$$S_{T_i} = \frac{D_{T_i}}{D} = \frac{D - D_{-i}}{D} = 1 - \frac{D_{-i}}{D} = 1 - S_{-i} \quad (5.25)$$

The first-order indices are computed from (5.23). To get more efficient parameter sampling, eFAST implements a random resampling of search curves, because different search curves produce different combinations of parameters thereby leading to the different sensitivity measures. Additionally, due to the symmetry properties of trigonometric functions, the same samples are repeated.

By using the advantage of repeated measurements, Marino *et al.* [84] proposed a novel method based on dummy parameters to determine the significance of first and total order indexes. This dummy parameter does not appear in the model equations and does not affect the model. The authors used a two-sample t test<sup>1</sup> on the data resulted from the resampling to determine if the sensitivity indexes of a parameter of interest are significantly different from the indexes returned for the dummy parameter.

### 5.2.2 Results

As we discussed in Section 5.1, the local sensitivity gives the information only at the vicinity of nominal values, and this deficiency can be solved by using the global sensitivity analysis. We performed the global sensitivity analysis using eFAST. Detailed information about its implementation can be found in [84]. It allows us to determine the sensitivity of parameter values in more details within the defined interval range. From the introduction of eFAST, we know that it gives the information about the first and total-order sensitivities. Figure 5.2 and 5.3 depict that the first

<sup>1</sup>It is assumed that two samples are from the normal distributions with unknown and unequal variances.

order sensitivities comprise only a small of part of whole sensitivities, while a major contribution comes from the total order sensitivities.

The ranges for the parameters were chosen as

$$(\text{best fit values}/10, \text{best fit values} * 10) \quad (5.26)$$

and the sensitivities were calculated within these ranges. Table 5.1 indicates the best fit values and the parameter ranges defined using (5.26). We found out that it is feasible to divide the time course into the highly dynamic initial phase and the slow dynamic steady-state phase, since there is no significant difference is found out between the individual time points. Therefore, we computed the sensitivities at 4 hours for the initial phase, and at 72 hours as a reference for the steady state.

Performing the analysis revealed that initially at 4 hours (Figure 5.2) at the cytoplasmic RNA processing, the processing rate of transfected RNA,  $k_0$ , and the degradation,  $\mu_P^{unp}$  are one of the parameters which affect the viral replication significantly. They affect all the plus, minus-strand RNA and the polyprotein levels. The model demonstrated that the initial degradation of plus-strand RNA after transfection, is determined mainly by the degradation rate,  $\mu_P^{unp}$ . From the model development, we know that these steps are essential in explaining the kinetic data. The sensitivity analysis supports this observation by indicating them as one of the most influential parameters at the initial stage of replication. However, the effect of these parameters is not influential at the later stages of the replication, at 72 hours (Figure 5.3). This directly indicates the dynamical nature of the replication, that the influence of some processes varies during the time.

The degradation of the processed plus-strand RNA,  $\mu_P^{cyt}$ , is found to be a non-sensitive at the initial stages of replication, however, appears to be important at steady-state for the plus-strand RNA.

The most influential parameter at the translation stage, is the translation rate of polyprotein,  $k_2$ . Since the polyprotein levels are directly dependent on this kinetic rate, it is one of the most important parameters which has a high sensitivity. This parameter is one of the few parameters which affects the level of all three variables at the initial and later stages of replication.

At the replication vesicles formation, the formation rate  $k_{Pin}$  is sensitive only at the initial stages and affects the level of all three variables. But at steady state, it is not sensitive anymore. The parameters in the replication vesicles are found to be the most influential even at the initial stage of replication. This directly highlights the importance of the processes inside the replication vesicles.

Initially, the synthesis rate of new plus-strand RNA,  $k_{4p}$ , affects the level of plus-, minus-strand RNA and polyprotein, while  $k_{4m}$  and  $\mu_{VMS}$  affect mainly the level of plus and minus-strand RNA. At steady state, the degradation rate,  $\mu_{VMS}$  is the most sensitive parameter which affects the viral replication and is found to be the most influential parameter in general.  $k_{4p}$  still shows an impact for the plus-strand RNA and a less sensitivity for the minus-strand and the polyprotein at later stages of replication. The parameter  $k_3$  is insensitive initially and mainly influences the level of the plus-strand RNA at steady state.

The parameter  $k_{Pout}$  is also one of the parameters which significantly affects the viral replication initially. The polyprotein,  $L$  is very sensitive against its degradation rate,  $\mu_L$  at the initial and later stages of replication. Overall, this analysis highlighted that the processes inside the replication vesicles play an important role in the viral replication.

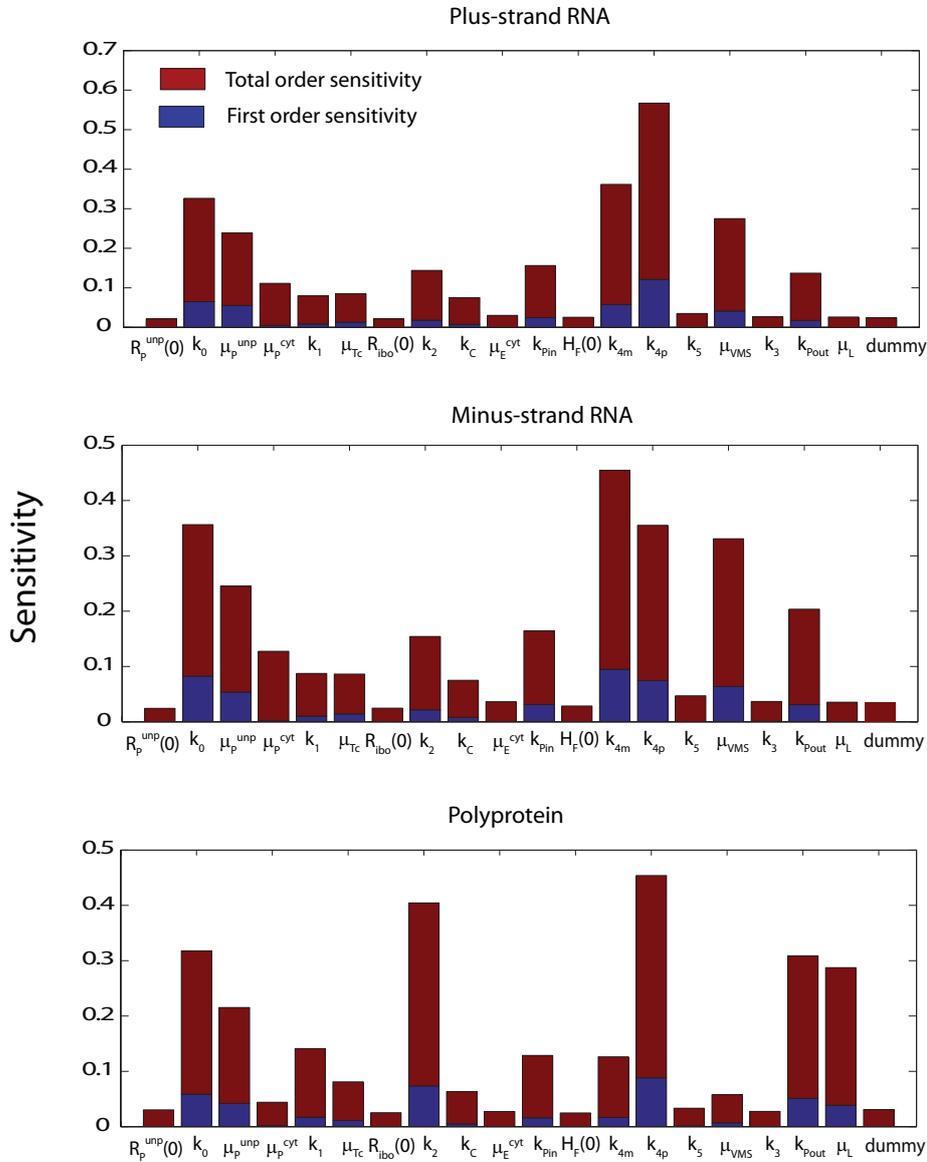


Figure 5.2: The parameter sensitivity at  $t = 4$  hours. Shown are the first and total order sensitivities for the plus, minus-strand RNA and the polyprotein. Initially, the processing rate  $k_0$ , the degradation rate  $\mu_p^{cyt}$ , the synthesis rates  $k_{4m}$ ,  $k_{4p}$  and the degradation rate  $\mu_{VMS}$  are the most influential parameters for the plus and minus-strand RNA. The processing rate  $k_0$ , the degradation rate  $\mu_p^{cyt}$ , the translation rate  $k_2$ , the synthesis rate  $k_{4p}$ , the export rate  $k_{Pout}$  and the polyprotein degradation rate  $\mu_L$  are the most influential parameters for the polyprotein. The dummy parameter is independent of model and is included by eFAST during the simulation. The sensitivities that are different than that of the dummy parameter are considered to be influential.

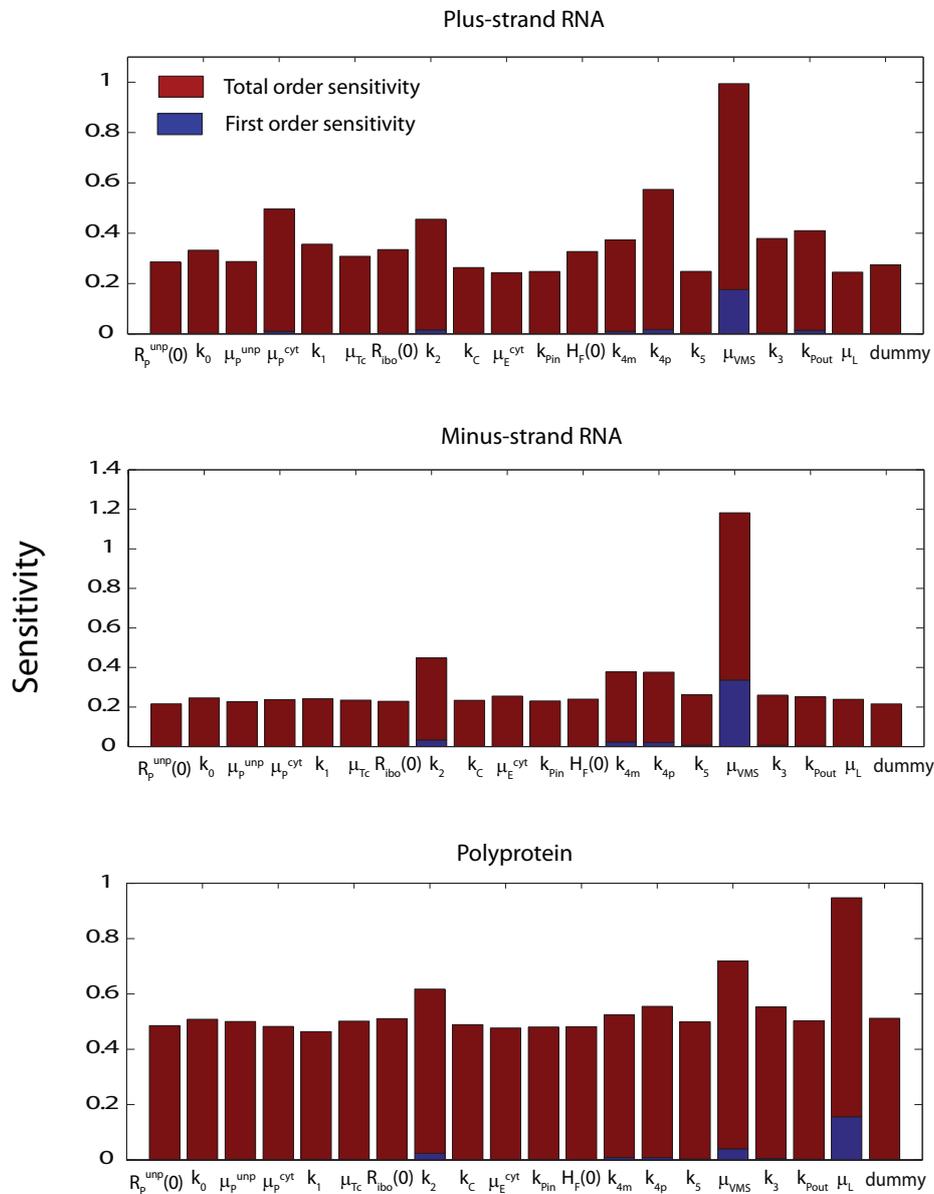


Figure 5.3: The parameter sensitivity at  $t = 72$  hours. Shown are the first and total order sensitivities for the plus, minus-strand RNA and the polyprotein. At steady state, the degradation rate inside the replication vesicles,  $\mu_{VMS}$  is the most influential parameter for the plus, minus-strand RNA and the polyprotein. In addition, the degradation rate of processed free plus-strand RNA,  $\mu_p^{cvt}$ , the translation rate  $k_2$  and the synthesis rate  $k_{4p}$  are also most influential for the plus-strand RNA. The translation rate  $k_2$  is found to be important for the minus-strand RNA. The degradation rate of polyprotein  $\mu_L$  affects the polyprotein level at most. The translation rate  $k_2$  also has a significant effect on the polyprotein.

### 5.2.3 Local sensitivity versus global sensitivity

In this section, we will discuss about the difference and the common points of the local and global sensitivities for the replication model.

At the cytoplasmic RNA processing, the parameters  $\mu_P^{unp}$  and  $k_0$  are found to be significant both in the local and global analysis (Figure 5.1 and Figure 5.2). Of course, from Figure 5.2 it is clear that they have a significant effect at the initial stage of replication. This highlights the importance of these rates in the HCV replication. The parameter,  $\mu_P^{cyt}$  is insensitive according to the local sensitivity analysis, however, found to have an effect on the plus-strand RNA at the later stages of replication (Figure 5.3).

The ribosome numbers,  $R_{ibo}$  is insensitive according to the global sensitivity, whereas, it has an influential effect on the polyprotein level in the local analysis. The translation rate,  $k_2$  in the local analysis is a sensitive parameter only for the polyprotein. However, according to the global analysis, it is one of the most sensitive parameter at the initial and later stages of replication.

At the replication formation, the formation rate of vesicles,  $k_{Pin}$  is sensitive rather than the host factor numbers,  $H_F(0)$  when performing the global analysis.

The synthesis rates,  $k_{4m}$  and  $k_{4p}$  are sensitive both in the local and global sensitivity analysis. In the global analysis, they are sensitive only in the initial phase of replication. Interestingly, in both analyses,  $k_{4p}$  found to affect mostly the polyprotein level (Figure 5.1 and Figure 5.2). The degradation rate inside the replication vesicle,  $\mu_{VMS}$  is also found to be important in both analyses. The global analysis revealed this parameter as the most important at steady state, while the local sensitivity analysis showed it to be the most important for the minus-strand RNA level.

The parameter  $k_{Pout}$  is important for the polyprotein level in the local analysis. According to the global analysis at 4 hours, it is shown to have a significant effect. At steady state, it has an influential effect on the plus-strand RNA. The degradation rate of polyprotein,  $\mu_L$  is sensitive for the polyprotein level according to both analyses.

Table 5.1: The parameter ranges for the global sensitivity analysis. The parameter ranges have been defined as indicated in (5.26). For more information see the text.

Parameters	Best fit values	Parameter ranges for global GSA
$k_0$	$0.00587 h^{-1}$	(0.0005787, 0.0587)
$k_1$	$1 h^{-1}molec^{-1}$	(0.1, 10)
$k_2$	$100 h^{-1}$	(10, 1000)
$k_c$	$1 h^{-1}$	(0.1, 10)
$k_{Pin}$	$2.07e - 6 h^{-1}molec^{-2}$	( $2.07e - 7$ , $2.07e - 5$ )
$k_{Pout}$	$0.333 h^{-1}$	(0.0333, 3.33)
$k_{Am}$	$1.7 h^{-1}$	(0.17, 17)
$k_{Ap}$	$1.7 h^{-1}$	(0.17, 17)
$k_5$	$10 h^{-1}molec^{-1}$	(1, 100)
$k_3$	$1e - 4 h^{-1}molec^{-1}$	( $1e - 5$ , $1e - 3$ )
$\mu_P^{unp}$	$0.758 h^{-1}$	(0.0758, 7.58)
$\mu_P^{cyt}$	$0.487 h^{-1}$	(0.0487, 4.87)
$\mu_{T_c}$	$0.243 h^{-1}$	(0.0243, 2.43)
$\mu_E^{cyt}$	$0.06 h^{-1}$	(0.006, 0.6)
$\mu_{VMS}$	$0.0703 h^{-1}$	(0.00703, 0.703)
$\mu_L$	$0.35 h^{-1}$	(0.035, 3.5)
$R_{ibo}(0)$	995 molecules	(99.5, 9950)
$H_F(0)$	88 molecules	(8.8, 880)



# Chapter 6

## Identifiability Analysis

Modeling biological processes often results with a large number of parameters. An important issue is therefore to examine whether these parameters can be uniquely determined from an input, initial conditions and observed outputs. Moreover, before estimating these parameters from the experimental data, a serious problem to overcome is how to verify whether the model parameters are identifiable based on the measurements of output variables. In the following, we will give a brief overview of identifiability analysis and for more details refer to [91].

Lets assume that a general dynamic system is expressed as follows:

$$\begin{aligned}\dot{x}(t) &= f(t, x(t), u(t), \theta), \\ y(t) &= h(x(t), u(t), \theta),\end{aligned}\tag{6.1}$$

where  $x(t) \in R^m$  is a vector of state variables,  $y(t) \in R^d$  is a measurement vector,  $u(t) \in R^p$  an input vector, and  $\theta \in R^q$  is a parameter vector. For the inverse problem,  $\theta$  is unknown and has to be estimated from the experimental data.

**Definition 6.0.1.** The dynamic system given by (6.1) is identifiable if  $\theta$  can be uniquely determined from the given system input  $u(t)$  and the measurable systems output  $y(t)$ . Otherwise, it is non-identifiable.

**Definition 6.0.2.** A system structure is globally identifiable if for any admissible input  $u(t)$  and any two parameter vectors  $\theta_1$  and  $\theta_2$  in the parameter space  $\Theta$ ,  $y(u, \theta_1) = y(u, \theta_2)$  holds if and only if  $\theta_1 = \theta_2$ .

Note that,  $u(t)$  is called an admissible input if it satisfies all system constraints at any time of interest and a solution of the dynamic system exists.

**Definition 6.0.3.** A system structure is locally identifiable if for any  $\theta$  within an

open neighborhood of some point  $\theta^*$  in the parameter space  $\Theta$ ,  $y(u, \theta_1) = y(u, \theta_2)$  holds if and only if  $\theta_1 = \theta_2$ .

The parameters for which no unique solution exists are termed as *non-identifiable*. These non-identifiabilities can arise from different sources. There are two main sources exist that are the non-identifiabilities due to a model structure and an insufficient amount of data. First, the model structure itself may cause the parameters to be functionally related which leads to the non-identifiability. The branch of identifiability analysis which deals with this kind of problems is called a *prior or structural identifiability*. Second, the model parameters are estimated by fitting the model to the experimental data and the limited amount of data may cause non-identifiabilities in the model parameters. The branch of identifiability analysis dealing with this problem is called a *practical identifiability*. In the following, we will overview the identifiability methods and for more information refer to [52], [143], [44], [91].

## 6.1 Prior Identifiability

A prior identifiability is concerned with the ability to uniquely identify the model parameters from a noise-free experimental data, given a particular input-output experiment [143]. We follow a definition given in [91].

**Definition 6.1.1.** For an admissible input  $u(t)$  in the time range of interest  $[t_0, t_1]$  and a given initial state  $x_0 = x(t_0)$ , which is independent of  $\theta$  and not an equilibrium point, if there exists an open set  $\Theta^0$  within the parameter space  $\Theta$  such that, for any two different parameter vectors  $\theta_1, \theta_2 \in \Theta^0$ , the solutions  $x(t, \theta, u)$  exist on  $[t_0, t_0 + \epsilon]$ , ( $t_0 < \epsilon \leq t_1 - t_0$ ) for both  $\theta_1$  and  $\theta_2$ , and  $y(t, \theta_1, x_0, u(t)) \neq y(t, \theta_2, x_0, u(t))$  on  $[t_0, t_0 + \epsilon]$ , the system structure is locally strongly identifiable.

This definition can be applied in defining the prior or structural identifiability and Xia and Moog [161] introduced the following definition.

**Definition 6.1.2.** Let  $C_u^N[t_0, t_1]$  is the function space expanded by all input functions that are  $N$  times differentiable and let  $M$  denote on open set of initial system states. The system structure is said to be structurally identifiable if there exists open and dense subsets  $M^0 \subset M$ ,  $\Theta^0 \subset \Theta$ , and  $U^0 \subset C_u^N[t_0, t_1]$  such that the system is locally strongly identifiable at  $\theta$  given  $u$  for any  $x_0 \in M^0$ ,  $\theta \in \Theta^0$ , and  $u \in U^0$ .

Prior identifiability analysis can be performed without any experimental observation. This analysis heavily relies on two basic assumptions: model structures are

absolute accurate and measurements are error free [91]. But, in practice these two assumptions are not valid. For example, in biomedical research, both the model uncertainty and the measurement error are usually large. Therefore, the estimates of model parameters might still be unreliable, when even prior identifiability suggests that the model parameters can be uniquely identified.

There are various methods available to detect the structural non-identifiabilities such as a power series expansion [116], a volterra and generating power series approach [70], a similarity transform approach [155] or differential algebraic methods [74], and a mean optimal transformation method [52].

## 6.2 Practical Identifiability

A practical identifiability deals with the accuracy of parameter values that can be estimated from noisy measurements [143]. As already mentioned in previous section, in practice, the assumptions on which prior identifiability analysis relies are not valid. Therefore, it is necessary to evaluate whether structurally identifiable parameters can be reliably estimated from a noisy data.

In this case, we assume that the output model contains measurements errors,

$$y(t) = h(x(t), u(t), \theta) + \varepsilon(t) \quad (6.2)$$

where  $\varepsilon(t)$  is a measurement error and  $\varepsilon \in \mathcal{N}(0, \sigma^2)$ .

Rodriguez-Fernandez *et al.* [126], [127] developed a method for practical identifiability analysis of ODE models by analyzing the correlations between the model parameters. We will shortly review the idea of this approach [91].

Let us assume the parameter estimate  $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q]$  that has been obtained after fitting the model to experimental data. The correlation matrix of the parameter estimates is calculated based on the Fisher Information matrix (FIM)<sup>1</sup> which quantifies a measure of informativeness of noisy measurement data for estimating the model parameters:

$$R = \begin{bmatrix} r_{11}(\hat{\theta}_1, \hat{\theta}_1) & r_{11}(\hat{\theta}_1, \hat{\theta}_2) & \cdots & r_{1q}(\hat{\theta}_1, \hat{\theta}_q) \\ r_{21}(\hat{\theta}_2, \hat{\theta}_1) & r_{22}(\hat{\theta}_2, \hat{\theta}_2) & \cdots & r_{2q}(\hat{\theta}_2, \hat{\theta}_q) \\ \vdots & \vdots & \ddots & \vdots \\ r_{q1}(\hat{\theta}_q, \hat{\theta}_1) & r_{q2}(\hat{\theta}_q, \hat{\theta}_2) & \cdots & r_{qq}(\hat{\theta}_q, \hat{\theta}_q) \end{bmatrix} \quad (6.3)$$

<sup>1</sup>For more details refer to Appendix A.

where  $r_{ij}$  is the correlation coefficient between parameter estimates  $\hat{\theta}_i$  and  $\hat{\theta}_j$  ( $i, j = 1, 2, \dots, q$ ). This correlation coefficient is defined as  $-1 \leq r_{ij} \leq 1$ . If  $r_{ij}$  is close to 1, then it means that there is a strong correlation between the parameters  $\hat{\theta}_i$  and  $\hat{\theta}_j$ . A strong correlation means that the parameters strongly depend on each other and a change in a model parameter  $\hat{\theta}_i$  can be compensated by an appropriate change in a model parameter  $\hat{\theta}_j$ . Thus, the highly correlated parameters are non-identifiable.

If the measurements errors are assumed to be uncorrelated and  $\varepsilon \in \mathcal{N}(0, \sigma^2)$ , then (6.1) is given as

$$FIM = \sum_{i=1}^N \left( \frac{\partial \hat{y}_i}{\partial \hat{\theta}} \right)^T V^{-1} \left( \frac{\partial \hat{y}_i}{\partial \hat{\theta}} \right) \quad (6.4)$$

where  $i$  is the  $i$ th time point of experimental observation,  $N$  is the total number of observations,  $\hat{y}_i$  is the model approximation of observation,  $\hat{\theta}_i$  is the parameter estimate of the model and  $V$  is a matrix of weights on variances which is positive definite. According to the Cramer Rao theorem [123],

$$C = FIM^{-1} \quad (6.5)$$

where  $C$  denotes a covariance matrix.

The elements of the correlation matrix then can be defined as

$$\begin{aligned} r_{ij} &= \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \quad i \neq j, \\ r_{ij} &= 1, \quad i = j. \end{aligned} \quad (6.6)$$

The inverse of  $FIM$  provides a lower bound for the variances of the parameter estimates based on the Cramer-Rao inequality<sup>1</sup> [73]. The lower bounds of the parameter variances are computed by

$$\sigma_{p_i}^2 \geq [FIM^{-1}]_{ii}, \quad (6.7)$$

From this statement, a 95% confidence interval for each parameter is defined as

$$[p_i - 1.96\sigma_{p_i}, p_i + 1.96\sigma_{p_i}] \quad (6.8)$$

A parameter is said to be a practically identifiable when its value is non-zero within the 95% confidence interval. We applied the practical identifiability analysis to our model parameters by fitting the model to the experimental data. For detailed

---

<sup>1</sup>For details see Appendix A.

information, we refer to Section 6.4.

## 6.3 Local Identifiability Analysis

In this Section, we will discuss about a local identifiability analysis which gives information at the vicinity of nominal values. This analysis is also known as a sensitivity-based identifiability analysis. The idea of sensitivity analysis can also be used to examine the identifiability of model parameters.

Sensitivity-based identifiability analysis resembles the structural identifiability analysis approach because it does not require the experimental data and assumes that the measurements are error free. It differs from the structural analysis by the fact that it does not use a model structure which is essential for the structural analysis. This sensitivity based method is different from the practical identifiability analysis in the sense that it does not take the observation error into account and similar in the sense that both require the nominal or actual estimates. This method is considered to be between the structural and practical identifiability analyses. We will discuss this approach in this section and the interested reader is referred to [91], [57].

A sensitivity based approach requires a predetermined nominal value, therefore, parameter identifiability is evaluated with respect to a particular point in the parameter space. The following definitions are introduced to define the concept of *at-a-point identifiability* [74], [119].

**Definition 6.3.1.** Let  $\theta^*$  denote a fixed point in the parameter space  $\Theta$ . A system is globally at-a-point identifiable if, for any admissible input  $u(t)$  and any parameter vector  $\theta \in \Theta$ ,  $y(u, \theta) = y(u, \theta^*)$  implies  $\theta = \theta^*$ .

**Definition 6.3.2.** Let  $\theta^*$  denote a fixed point in the parameter space  $\Theta$ . A system is locally at-a-point identifiable if, for any admissible input  $u(t)$  and any parameter vector  $\theta$  within an open neighborhood of  $\theta^*$ ,  $y(u, \theta) = y(u, \theta^*)$  implies  $\theta = \theta^*$ .

Let us assume the number of time points  $t_1 \leq t_2 \leq \dots \leq t_N$  at which the state variables are measured. Then the sensitivity coefficient at each time point  $t_k$ ,  $k = 1, 2, \dots, N$  for a given nominal parameter vector  $\theta^*$  is given as

$$s_{ij}(t_k) = \frac{\partial y_i(t_k, \theta^*)}{\partial \theta_j}, \quad i = 1, \dots, d \quad j = 1, \dots, q \quad (6.9)$$

where  $y_i$  is the  $i$ th component of  $y$  and  $\theta_j$  is the  $j$ th component component of  $\theta$ .

Then the sensitivity matrix for all time points is defined as

$$S_{dN \times q} = \begin{bmatrix} s_{11}(t_1) & \cdots & s_{1q}(t_1) \\ \cdots & \ddots & \cdots \\ s_{d1}(t_1) & \cdots & s_{dq}(t_1) \\ \vdots & \vdots & \vdots \\ s_{11}(t_N) & \cdots & s_{1q}(t_N) \\ \cdots & \ddots & \cdots \\ s_{d1}(t_N) & \cdots & s_{dq}(t_N) \end{bmatrix} \quad (6.10)$$

The parameter is identifiable if the system output is highly sensitive to a small perturbation of the parameter. The correlation between any two parameters can be evaluated by analyzing the dependence of the sensitivity matrix columns. Jacquez and Greif developed a correlation method to study identifiability for linear models which is also applicable for nonlinear models [58], [57].

Lets consider a Taylor expansion of the first order for the system output in the neighborhood of the nominal parameter vector  $\theta^*$ ,

$$\begin{aligned} y_k(\theta) &= y(x(t_k), u(t_k), \theta) \\ &\approx y(x(t_k), u(t_k), \theta^*) + \left. \frac{\partial y(x(t_k), u(t_k), \theta)}{\partial \theta} \right|_{\theta=\theta^*} \cdot (\theta - \theta^*), \end{aligned} \quad (6.11)$$

where  $k = 1, 2, \dots, N$  denotes the index of the measurements time points. Consider  $r_k$  denote the measurements at  $t_k$  without errors and  $\Delta = \theta - \theta^*$ . We can then obtain the residual sum of squares between the exact measurements and the linear approximation,

$$\begin{aligned} RSS(\Delta\theta) &= \sum_{k=1}^N \left[ r_k - y_k(\theta^*) - \left. \frac{\partial y(x(t_k), u(t_k), \theta)}{\partial \theta} \right|_{\theta=\theta^*} \cdot \Delta\theta \right]^2 \\ &= \sum_{k=1}^N \left[ \left. \frac{\partial y(x(t_k), u(t_k), \theta)}{\partial \theta} \right|_{\theta=\theta^*} \cdot \Delta\theta \right]^2 \end{aligned} \quad (6.12)$$

Note that  $r_k - y_k(\theta^*) = 0$ . Then we can formulate (6.12) in terms of sensitivity matrix,

$$RSS(\Delta\theta) = (S\Delta\theta)^T \cdot S\Delta\theta, \quad (6.13)$$

where  $S$  is the sensitivity matrix defined in (6.10).

The minimum of  $RSS(\Delta\theta)$  is obtained when  $S^T S \cdot \Delta\theta = 0$ .  $\theta = \theta^*$  is the unique solution of  $S^T S \cdot \Delta\theta = 0$ , if  $S^T S$  is of full rank. This indicates that the model parameters  $\theta$  are *locally identifiable* at  $\theta^*$ . If  $S^T S$  is singular, then there at least one nontrivial solution  $\hat{\theta} \neq \theta^*$  exists such that the model parameters are not identifiable at  $\theta^*$ .

It is important to identify those parameters that are not identifiable if  $S^T S$  is not of full rank. This can be done by calculating the correlations between parameters using the sensitivity matrix (6.10). Each column of (6.10) indicates the sensitivity of the system responses at all time points with respect to one specific parameter. Thus, the correlation between two parameters can be calculated as the sample correlation of two columns in the sensitivity matrix (6.10)

$$\text{corr}(S_i^*, S_j^*) = \frac{\text{cov}(S_i^*, S_j^*)}{\sigma(S_i^*)\sigma(S_j^*)} \quad (6.14)$$

where  $S_i^*$  and  $S_j^*$  denote  $i$ th and  $j$ th columns of the sensitivity matrix, respectively.  $\text{cov}(S_i^*, S_j^*)$  denotes the sample covariance between  $S_i^*$  and  $S_j^*$ .  $\sigma(S_i^*)$  and  $\sigma(S_j^*)$  are the sample standard deviations of  $S_i^*$  and  $S_j^*$ , respectively. As mentioned before, if the correlation between any two parameters is close to 1, then these two parameters are non-identifiable.

We are going to apply this idea of local identifiability to our model parameters and the details will be discussed in Section 6.4.

## 6.4 Results

We identified the non-identifiable parameters when performing a model fitting to the experimental data. We iteratively fitted the model to the data and defined the parameters which are non-identifiable. We performed the practical identifiability analysis by fitting the model to the data (for parameter estimation refer to Chapter 3). During the fitting, it has been observed that the parameters  $k_1$ ,  $k_c$ ,  $k_5$  and  $k_3$  have large confidence intervals, which means they cannot be properly estimated from the data. When performing the practical identifiability analysis, Hengl *et al.* [52] suggested to fix the parameter within plausible ranges to that value which belongs to the best fit. In order to render all parameters identifiable, we performed a fixation of these parameters.

The parameter  $k_1$  is insensitive within the reported range [1, 100] in [48], which is accepted to be a plausible range. Therefore, we set  $k_1 = 1$  which is the lowest

range for  $k_1$ . We found out that this parameter is sensitive below the lowest range.

As it was reported in Dahari *et al.* [48], we find out that to ensure the experimentally observed 10:1 ratio, the condition  $k_3 \ll k_5$  should be fulfilled. The model simulations demonstrated that when these parameters are equivalent, then this ratio is not valid anymore. Therefore, we set  $k_5/k_3 = 10^5$  which satisfies the ratio of the plus-to minus-strand RNA. Furthermore, we observed during the parameter estimation for our model that the optimization was trying to push the parameters  $k_5$  and  $k_c$  to the values beyond the upper bounds of the biologically plausible range, which are  $[0, 10]$  and  $[0, 1]$ , respectively. In order to avoid this, we fixed these parameters to their upper bounds,  $k_5 = 10 \text{ h}^{-1}$  and  $k_c = 1 \text{ h}^{-1}$ .

During the optimization, we observed that the condition,  $\mu_{T_c} < \mu_P^{cyt}$  was violated. Simply, the optimization returned almost equal values for both parameters. This condition implies that the free plus-strand RNA,  $R_P^{cyt}$  degrades faster than the plus-strand RNA within the translation complexes,  $T_c$ . Further using this fact, we assumed that  $R_P^{cyt}$  degrades two times faster than  $T_c$ , and set  $\mu_P^{cyt}/\mu_{T_c} = 2$ . This allows us to get rid of one free parameter.

Finally, we performed the optimization for the remaining 11 free parameters and obtained a well-conditioned problem with identifiable parameters (Table 6.1).

We then performed a local identifiability analysis at the best fit values in order to determine non-identifiable parameters using SensSB [125], a software toolbox for the development and sensitivity analysis of systems biology models. This analysis is able to determine the identifiability in a case of unlimited and error free data and can distinguish between the non-identifiability due to the structure of the model or the lack of experimental data. The result of this analysis is depicted in Figure 6.1 which represents the parameter correlation matrix.

The correlation matrix quantifies the relationship between the parameters, where the elements of the correlation matrix are the correlation coefficients between the  $i$ -th and  $j$ -th parameter. As in Section 6.1, a high correlation between two parameters means that the change in the model output caused by the change in the parameter can be compensated by the appropriate change in the other parameter. This then prevents the parameters from being uniquely identifiable despite the output is very sensitive to changes in the individual parameters.

The results demonstrated that most parameters at the neighborhood of best fit values are identifiable (Figure 6.1). The correlation values of these parameters are significantly different from the high and low correlation values. Therefore, this analysis supports the goodness-of-fit and the optimality of the estimated parameters. This analysis revealed that only the parameters  $\mu_E^{cyt}$  and  $\mu_L$  are highly correlated.

Table 6.1: *Practical identifiability analysis on the model parameters.*

Parameter name	Parameter values	Status
$k_0$	$0.00587 h^{-1}$	identifiable
$k_1$	$1 h^{-1}molec^{-1}$	fixed
$k_2$	$100 h^{-1}$	experimentally observed
$k_c$	$1 h^{-1}$	fixed
$k_{Pin}$	$2.07 \times 10^{-6} h^{-1}molec^{-2}$	identifiable
$k_{Pout}$	$0.333 h^{-1}$	identifiable
$k_{Am}$	$1.7 h^{-1}$	experimentally observed
$k_{Ap}$	$1.7 h^{-1}$	experimentally observed
$k_5$	$10 h^{-1}molec^{-1}$	fixed
$k_3$	$10^{-4} h^{-1}molec^{-1}$	fixed
$\mu_P^{unp}$	$0.758 h^{-1}$	identifiable
$\mu_P^{cyt}$	$0.487 h^{-1}$	identifiable
$\mu_{T_c}$	$0.243 h^{-1}$	identifiable
$\mu_E^{cyt}$	$0.06 h^{-1}$	experimentally observed
$\mu_{VMS}$	$0.0703 h^{-1}$	identifiable
$\mu_L$	$0.35 h^{-1}$	experimentally observed
$R_{ibo}(0)$	995 molecules	identifiable
$H_{F,high}(0)$	88 molecules	identifiable
$H_{F,low}(0)$	10 molecules	identifiable
$p_{scale}$	$2.80 \times 10^3$	identifiable

A high correlation means that the change in the output caused by the change in  $\mu_E^{cyt}$  is compensated by the change in  $\mu_L$ . Interestingly, these parameters have been experimentally observed which indicates that even the experimentally observed parameters can be rendered as non-identifiable.

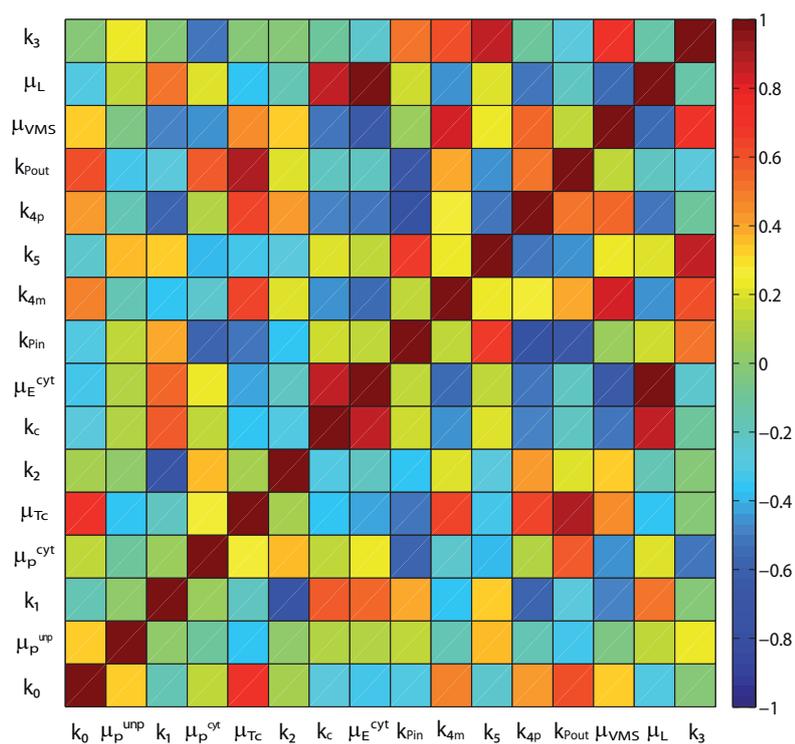


Figure 6.1: The correlation matrix which represents the correlation between the individual parameters. The positive values indicate a positive correlation, whereas the negative values indicate a negative correlation. The diagonal elements have a perfect correlation of 1 which indicates that the parameter perfectly correlates with itself. This correlation matrix indicates that all parameters except  $\mu_E^{cyt}$  and  $\mu_L$  are locally identifiable at the vicinity of best fit values. The parameters  $\mu_E^{cyt}$  and  $\mu_L$  are highly correlated and the correlation value is close to 1.

# Chapter 7

## Discussion

In this chapter, we will discuss essential aspects of this work. In particular, we will give a short summary and a discussion of the results of each chapter. The main statements and results are summarized in Section 7.1. In Section 7.2, our model developed in this thesis will be compared and discussed with the other existing models. In Section 7.3, we will discuss a future perspective and ultimate goals of this work.

### 7.1 Summary

In this thesis, we developed a mathematical model based on the ordinary differential equations with the focus to investigate the mechanisms which elucidate the Hepatitis C Virus replication.

In **Chapter 1**, we discussed the biological insights about the HCV and its life-cycle. We gave a broad overview about the recent status of HCV research from the clinical and basic science perspective. HCV research is inspired from the clinical studies, and became a hot topic for the basic science. The current statistics claims that only about 50-60 % of all treated patients show a positive response to the drug treatment. The limited efficiency of current available drug treatment led to a need to understand HCV lifecycle in more details, and a better understanding of HCV lifecycle will lead to the development of new drugs with a high efficiency. From this chapter, we see that the HCV lifecycle consists of three essential parts, which are the entry, the replication and the particle formation or the assembly.

HCV entry into the host cell is the first step in the virus life-cycle which starts with the attachment of the virus particle to the host cell. This is done by the specific interactions between several receptors on the cell surface and a viral attachment

protein on the surface of the particle. Identification of the factors which are limiting the HCV entry at this stage may prevent a successful virus replication. The application of the electroporation<sup>1</sup> allows a transfection of virus strains into the cell by ignoring the virus entry and focus directly on the replication and the assembly. The development of subgenomic HCV made possible to isolate the virus replication by ignoring the HCV assembly. In subgenomic HCV, the regions in the genomes which encode for structural proteins are deleted, which are essential for the formation of virus particles. HCV replication is the process which consists of the translation where the HCV genome is translated into a viral polyprotein which is cleaved into several non-structural proteins, and the replication which undergoes in the vesicles called “replication vesicles”. Overall, HCV replication process is poorly understood and the development of mathematical models made possible to study the process in a systematic way.

As we discussed, the first attempt to model the HCV replication was done by Dahari *et al.* [48] and this model was calibrated with the data from Quinkert *et al.* [120]. The model can nicely reproduce the steady state data and explain the mechanism underlying the observed ratio between the plus and minus-strand RNA. However, we showed that the model is unable to fit to the kinetic data, thereby showing a complexity of the replication process. This model only considers the effect of cellular ribosomes which is reported to be a limiting factor for the replication. The kinetic data indicates that the replication process is highly dynamic and a large number of virus host interactions imply the high complexity of the process.

In **Chapter 2**, we discussed the development of the new mathematical model for HCV replication. First, the model was represented in terms of kinetic reactions and then using mass action kinetics was converted into the ordinary differential equations. We assumed that whole replication process undergoes in the cytoplasm and the replication vesicles. Initially, after transfection, we assumed that the transfected plus-strand RNA is processed before being translated. This assumption allows us to filter those plus-strands which are able to participate in the translation and the replication. Those which are not able to replicate are considered to be defective and they degrade before being translated. As we already mentioned in Subsection 2.1.1, the viral RNA lacking a coat protein might degrade faster before adapting to the cell environment.

---

<sup>1</sup>The use of high-voltage electric shocks to introduce DNA into cells can be used with most cell types, yields a high frequency of both stable transformation and transient gene expression. It is usually used in molecular biology to introduce some substance into a cell [117]

After the processing, the plus-strand RNA is translated by the host ribosomes into the viral polyprotein. This polyprotein is then cleaved into several non-structural proteins. One of these proteins NS4B induces the formation of replication vesicles, and this process we considered implicitly in our model. Another protein, NS5B participates in the formation of replication complexes within the replication vesicles together with the translated plus-strands and the cellular host factor. As mentioned, a number of cellular factors have been found to be associated with the replication complex. Therefore, we assumed the interaction of the host factor with the virus at the formation of replication complex. As the replication complex here we mean an intermediate plus-strand replicative complex. This complex leads to the synthesis of complementary minus-strand RNA which exists in the form of double-stranded RNA. In this case, a polymerase within the replication complex plays a machinery role by copying the minus-strand RNA from the plus-strand RNA. The minus-strand RNA which exists within the double-strand RNA then interacts with the polymerase to form an intermediate minus-strand replicative complex. Within this complex, the minus-strand RNA serves as a template for the synthesis of new plus-strand RNA. The newly produced plus-strands then interact with the polymerase in the production of the minus-strand RNA or leaves the vesicle to participate in the translation. This export serves as a feedback for the translation which allows the new plus-strand RNA to be translated and further participate in the genome replication.

As we mentioned earlier, many cellular host factors have been found to interact with the virus, however, in the model, we considered only the effect of two host factors, the ribosomes and the cellular factor that participates in the formation of replication vesicles. Due to a limited amount of experimental data, we kept the model simple than actually required.

Further important thing is the number of replication vesicles which are formed during replication. Quinkert *et al.*[120] reported the existence of several hundred replication vesicles formed during a viral amplification. As reported in [48], we assumed the formation of a single replication vesicle, which approximates many vesicles if they are formed at the same time. However, Wölk *et al.* [160] reported the existence of replication vesicles with different sizes. This indicates that these vesicles may be formed gradually as the number of viral RNA increases. This observation also supports that these replication vesicles may grow in size as the virus replicates. In order to include these details, a more complex model will be required, and our model is limited to incorporate these assumptions. In this case, instead ODEs more complex PDE models can be exploited to incorporate above mentioned effects.

In **Chapter 3**, the results obtained from the model calibration with the new model are discussed. The model calibration problem is formulated as an inverse problem where apart solving differential equations we have to estimate the parameter values so that the model predictions match the experimental data. In the optimization problem, we formulated an objective function which represents the difference between the model and the experimental data. The optimization problem was reduced to the finite dimension by using the Multiple Shooting algorithm and the resulting nonlinear constrained least squares function was solved by using the Generalized Gauss-Newton method. The list of the estimated parameter values are given in Table 3.1.

Additionally, we used the evolutionary optimization, the Genetic Algorithm which was performed independently (Table 3.2). Genetic Algorithms are search algorithms based on the mechanics of natural selection and natural genetics [46]. Due to its random nature, the genetic algorithm improves the chances of finding a global solution. It enables to solve unconstrained, bound-constrained, and general optimization problems, and it does not require the functions to be differentiable or continuous [86]. The results showed that the parameter results obtained by this algorithm lie close to that one which were obtained by the Multiple Shooting Algorithm. Overall, the optimization problem is counted as a difficult problem when the number of estimated parameters are high. The gradient-descent method like the Gauss-Newton method suffers from sticking into a local minima by giving only partial information about the optimality of the solution. The evolution based method like the Genetic Algorithm is advantageous by allowing to search a broader space and a chance to find a global minima in this case is higher. Taking together, the combination of both approaches gives the more confident results about the optimality of the solution.

The optimization results show that the model can excellently fit to the experimental data. Interestingly, the model fits to the two kinetic datasets from the high (Huh7 Lunet cells) and low permissive cells (Huh7 cells) simultaneously. In contrast to the existing model, the model can excellently explain the kinetic data even from the both cells. The model also excellently fits to the steady state data from Quinkert *et al.* [120]. The only parameter which is responsible for the difference in the replication dynamics in both cells is the amount of the cellular host factor which participates in the formation of replication complex within the replication vesicles.

**Chapter 4** focuses on the validation of model dynamics and predictions as well as conclusions drawn from the model. As already discussed, initially, we used the

replication deficient JFH virus measurements to validate our model. We demonstrated that by setting the formation rate of the replication vesicles,  $k_{Pin}$  to zero, the model excellently reproduces the plus-strand RNA and the polyprotein dynamics by using the parameters which have been obtained from the fitting. This validation supports the plausibility of the model parameters.

Another model validation was performed using the mutated HCV genome kinetics, where the non-translated regions of the viral genome have been exchanged which leads to the decrease in the replication efficiency. This efficiency was measured in terms of polyprotein kinetics, and at the same time, the steady state changes in the ratio of plus to minus-strand RNA were observed. We show that the model nicely reproduces these kinetic and saturation data, by tuning only those parameters which mimic the changes in the non-translated regions. From the steady state data which represents the ratio of the plus-strand RNA to the minus-strand RNA, we can see the severe changes in the ratio when these regions are mutated (Figure 4.2). From this plot, we easily observe that the mutation in the region which is responsible for the minus-strand synthesis causes a dramatic decrease in the synthesis of minus-strand RNA. The model, however, can explain the observation only qualitatively by predicting only a slight change in the ratio. This shows that the model lacks some detailed mechanisms which may help to explain the changes properly.

Quantitative predictions require a calibration and a validation of the model and after completing these steps, we performed predictions using the model. One of our central predictions is the prediction of the source which is responsible for the replication permissiveness in the high permissive Huh7 Lunet and the low permissive Huh7 cells. We demonstrated that the replication permissiveness can be explained by the differential expression of the cellular host factor which participates in the formation of replication vesicles. We showed the uniqueness of this observation by searching the other possible factors that can explain this observation. It is known that most conformational changes in the cell after the virus infection happens by the formation of replication vesicles. A limited amount of cellular host factor at this stage may limit the formation of these vesicles which are essential for the genome replication. Additionally, we showed that the replication attains a steady state due to the limited amount of the host factor. A large number of cellular factors have been found to be associated with the HCV replication, however, the specific role of them is unknown. This prediction can help to filter these factors according to their properties and lead to the identification of the cellular host factors which may limit the HCV replication. However, in reality, there are several other host factors may exist that can limit the replication. This model may not be appropriate

one to apply if there the other limited factors exist. Our model approximates the interactions of many host factors which interact with the virus at the formation of replication vesicles.

Furthermore, by using a simple analysis, we showed that for the successful replication the degradation inside the replication vesicles should be smaller than the degradation in the cytoplasm. This prediction supports the observation that the replication vesicles protect the viral genome and the protein from the host protease and nuclease. However, in experimental conditions, it is difficult to measure the degradation rate inside the replication vesicles which makes things impossible to verify this prediction.

Further, we showed that the model excellently predicts some observations at steady state such as the protein level which are about a million molecules and the active polymerase levels which are  $< 0.1\%$  of total polymerase and they are in close agreement with the experimental observations. However, in the low permissive cells, the model predicts about 2-3 fold less proteins than the observed one. The experimental data is derived from the high permissive cells, as it was reported in [120]. The situation might be different in the case of the low permissive cells, therefore, the additional measurements are needed to verify this prediction.

We also revealed an importance of the assumption about the processing of transfected viral genome by showing that the initial decrease in the plus-strand RNA mainly determined by the decay in the transfected viral genome. It is difficult to detect if the viral RNA degrades at different rates in the cytoplasm. It might be that the virus degrades at the same rate and thereby uses different mechanisms to sustain a successful replication. However, the assumption of the different degradation rates for the viral RNA helps to explain the experimental data.

**Chapter 5** focuses on the parameter sensitivity analysis where we investigated the important parameters which affect the HCV replication. As already discussed, we performed two types of sensitivity analysis to get a more detailed knowledge about the parameters. The local sensitivity analysis aimed at investigating the sensitivity at the vicinity of nominal values (best fit values). The results demonstrated the various effects of parameters for the plus, minus-strand RNA and the polyprotein levels. The most influential and common parameters for all three variables, plus-, minus-strand RNA and polyprotein are  $\mu_P^{unp}$ ,  $H_F(0)$ ,  $k_{4m}$ ,  $k_{4p}$  and  $\mu_{VMS}$ . It is obviously seen from these results that the kinetic rates associated with the replication vesicles are the most important. This indicates the importance of existence of the replication vesicles.

In order to get more detailed information, we performed a global sensitivity analysis with eFAST (Extended Fourier Amplitude Test). This analysis was performed on the initial and saturation stages of the replication. The parameters  $k_0$  and  $\mu_P^{unp}$  are the most influential parameters at the initial stage of replication. However, these parameters are not sensitive at steady state. The most sensitive parameter is the degradation rate inside the replication vesicles,  $\mu_{VMS}$ . The parameters within the replication vesicles are also found to be sensitive initially. Overall, the sensitivity analysis revealed a strong correlation between the three variables and the common influential parameters. It appeared that in most cases, the influential parameter for one variable is also sensitive for another variable. It also demonstrates a high dependence of the replication processes on each other. In general, this analysis is very useful to detect a vulnerable step in the replication process and to identify a possible target for a drug treatment.

**Chapter 6** discusses about the identifiability of the model parameters. Identifiability is a prerequisite before making proper predictions, because non-identifiabilities in the model can enormously affect an outcome of the prediction. This is inevitable process due to the lack of enough experimental measurements to fully calibrate the model and this problem can be solved by acquiring a large amount of data. We performed a fitting and a fixing approach to decrease the number of the free parameters. If no fixing is done then the confidence intervals of the parameters are wide which indicates the non-identifiabilities in the model parameters.

Further, using a local identifiability analysis, we showed an optimality of the estimated parameters (Figure 6.1). We demonstrated that only the degradation rates  $\mu_L$  and  $\mu_E^{cyl}$  are highly correlated. These parameters are experimentally observed, but still rendered structurally non-identifiable at the vicinity of the best fit values. Structural non-identifiability manifests itself in functionally related parameters and is independent of the accuracy of available experimental data. This analysis is useful to support the parameter estimation problem and to explore the robustness of the system at the vicinity of nominal values by computing the local sensitivities. However, this kind of analysis gives information only in a neighborhood of the best fit values and for a complete analysis a global identifiability is needed.

## 7.2 Comparison with other models

We have already discussed the first replication model in details in Section 1.4. Here, we are going to discuss the other related models from [92] and [89].

Mishchenko *et al.* in [92] reported a model for a suppression of subgenomic HCV RNA replication in the cell culture in the presence of HCV NS3 protease and NS5B inhibitors. In this model, the authors considered a role of human vesicle-associated membrane protein (VAP-A) which is critical for the assembly of the active replicase complex on the membrane. This protein was assumed to play a limiting role in the replication. However, whether this protein limits the replication or not is not known. Since there can be a dozen of such factors which may limit the replication, the experimental validation is needed. Additionally, they did not consider the formation of replication vesicles which is found to be crucial for the virus to amplify efficiently. The protective property of replication vesicles is found to be essential, as we showed using our analysis in Subsection 4.2.2.

Recently, McLean *et al.* [89] published another model for intracellular lifecycle of plus-strand RNA virus which was applied for HCV case. This model is based on the model reported in [48]. In addition to the viral replication, they considered the assembly of viral particles. They studied the alternative strategies for the allocation and reallocation of the viral genomes in HCV lifecycle. Moreover, the authors investigated an inherent trade-off in the replication of plus-strand RNA viruses within cells that results from the translation, the replication and the packaging. They reported using the model that a free allocation of the viral genomes among translation, replication and packaging gives the most productive strategy for the virus. As in [92], they did not consider the formation of replication vesicles. In contrast to our model, they did not take into account the effect of the cellular factor on the viral replication. With our analysis, we demonstrated that the model cannot explain the viral kinetics quantitatively without incorporating a cellular host factor.

Both models described above suffer from a lack of kinetic data, thereby leading to the problems in the parameter estimation. The use of kinetic data allowed us to properly establish a mathematical model, which quantitatively fits to the data and explains the cause of the different HCV replication in the Huh7 clones.

## 7.3 Perspectives for future work

### 7.3.1 Infection model

In transfection, a large number of the viral genome are introduced into the cell and it is usually done in vitro studies. It allows us to model this process by using a deterministic approach. However, in a real situation, it has been observed that only a

small amount of virus can lead to the successful replication in the cell. In this case, a deterministic approach is not valid anymore and instead a stochastic modeling should be applied. Nevertheless, testing a deterministic modeling and applying a stochastic approach is widely used in a practice. This approach has an advantage to estimate the parameters first by using the deterministic model and to perform further simulations using the stochastic approach. Since the multiplicity of infection (MOI) is too low in the natural infection, the processing of transfected plus-strand RNA which accounts for defective and non-defective virus can be neglected. In our case, the HCV entry into the cell was neglected and therefore, this step was necessary to explain the data. In a natural case, the virus infects the cell by attaching to the cell surface receptors, fuses and enters the cell. It means that only the virus which successfully enters the cell can participate in the replication. Under such conditions, the stochastic model may yield qualitatively different behavior, since it allows for inherent fluctuations in the levels of viral constituents [141].

The use of the stochastic analysis provides certain advantages over the deterministic analysis. The random fluctuations might affect reaction dynamics, when modeling the process as a discrete rather than continuous entities [87]. The model parameters have already been estimated through the deterministic modeling and only the stochastic simulations are required by using these parameters. In this case, multiple stochastic simulations are required to obtain reasonable results and an average of these results can be compared with the deterministic analysis to reveal the important differences between these two analyses. Eventually, it will be interesting to see how the infection setting differs from the known deterministic setting. Of course, for the model validation some extra measurements for the infection setting are required.

### 7.3.2 Future application of the model

Despite the recent advances in understanding of HCV biology, many questions remain without answers. The development of mathematical models for HCV replication will certainly enhance our understanding. Therefore, our model may serve as a nice tool to study the complex nature of HCV lifecycle.

Some studies reported a number of cellular factors which are associated with the HCV replication (see [32]). Certainly, not a single, but the complex interactions of these host factors with the virus might shape the outcome of the viral replication. However, our understanding of how these interactions resulted in the successful replication is still in its infancy. In this respect, our model can be extended to

include the known host factors which will help us to analyze the effect and the role of the host factors on the replication process. This is important because some studies suggested to apply the drugs not against the virus itself but on the cellular host factors which found to be associated with the virus.

Our model is based on the experimental settings where the immune pathway was silenced to obtain a successful replication. The model can be extended to study a complex interaction between the virus and the host immune system. The immune pathway which detects HCV genome is an RIG-pathway which leads to a production of type I interferons (IFNs) in the infected cells [163]. It was reported that the HCV escapes this pathway by using unknown mechanisms. In this case, our model can be used in developing a comprehensive model to study the host immune and virus interactions.

Currently available experimental studies with the subgenomic HCV allowed us to model only the replication cycle of HCV. Using the infection setting it is possible to extend the studies to the whole HCV lifecycle. Usage of mathematical model in this case will be of great help.

Today, a number of mathematical models are available to study the effect of drug treatment against the HCV infection [12], [102], [33]. These models consider the virus infection only in a population of cells, whereas the models for a single cell level are limited. The future goal is to integrate the replication model in the single cell level with the population based models and to get a more comprehensive model to improve the efficiency of the drug treatment.

# Appendix A

## Law of Mass Action

Assume a reversible chemical reaction with  $N$  reactants  $S_i$ ,  $i = 1, \dots, N$  and  $M$  products  $P_j$ ,  $j = 1, \dots, M$ , given by



where  $\alpha_i$  and  $\beta_j$  denote the number of the reactants and products molecules involved in a single reaction step. From [4], the kinetic rates of the forward and reverse reactions are given as follows

$$\begin{aligned} [\text{forward rate}] &= k_1 \prod_i^N S_i^{\alpha_i} \\ [\text{reverse rate}] &= k_{-1} \prod_j^M P_j^{\beta_j} \end{aligned} \quad (2)$$

This formula shows that the rate of the reaction is directly proportional to the product of the reactant concentration. It means that the rate of reactions depend on the probability of simultaneous collision of the corresponding number of the reactant molecules.

From (2), for the stationary non-equilibrium state, we can derive

$$[\text{forward rate}] = [\text{reverse rate}]$$

which gives

$$\frac{k_1}{k_{-1}} = \frac{\prod_i^N S_i^{\alpha_i}}{\prod_j^M P_j^{\beta_j}} \quad (3)$$

This ratio is called the *law of mass action*. This means when a reversible reaction has attained an equilibrium, the ratio of molar concentrations of the reactants to those of the products remains constant.

## ANOVA Decomposition

We will define a definition of ANOVA decomposition from [139]. Lets define the output variable as  $y = f(x)$ ,  $x = (x_1, \dots, x_n)$ , defined in the unit hypercube  $0 \leq x_1 \leq 1, \dots, 0 \leq x_n \leq 1$ . The representation of  $f(x)$  in the form

$$f(x) = f_0 + \sum_i f_i(x_i) + \sum_{i < j} f_{ij}(x_i, x_j) + \dots + f_{12\dots n}(x_1, \dots, x_n). \quad (4)$$

is called ANOVA - decomposition if

$$f_0 = \int f(x) ds \quad (5)$$

and

$$\int_0^1 f_{i_1, \dots, i_s} dx_{i_p} = 0 \quad \text{for } 1 \leq p \leq s. \quad (6)$$

The conditions (5) and (6) uniquely define all the terms in (4). To define one-dimensional terms, one needs to integrate (4) over all variables except  $x_i$ .

$$\int f(x) \prod_{p \neq i} dx_p = f_0 + f_i(x_i). \quad (7)$$

This operation can be extended to define all higher order terms. If one assumes  $f(x)$  to be square integrable then by squaring and integrating (4), the following can be obtained

$$\int f^2(x) dx - f_0^2 = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} \int f_{i_1, \dots, i_s}^2 dx_{x_1, \dots, x_s} \quad (8)$$

If  $x$  is a random point uniformly distributed in the hypercube, then  $f(x)$  and  $f_{i_1, \dots, i_s}(x_1, \dots, x_s)$  will be random variables. From the equation (8), one defines the constants,

$$D = \int f^2(x) dx - f_0^2 \quad (9)$$

and

$$D_{i_1, \dots, i_s} = \int f_{i_1, \dots, i_s}^2 dx_{x_1, \dots, x_s} \quad (10)$$

where  $D$  is called a total variance, and  $D_{i_1, \dots, i_s}$  are called variances.

## Fisher Information Matrix

The Fisher Information Matrix plays a central role in the estimation, identification and information theory and provides a summary of the information in the data relative to the quantities of interest [140]. In the following, we will define a Fisher information matrix from Navarro *et al.* [101]. Let  $X = (X_1, \dots, X_n)$  be a random sample and let  $f(X|\theta)$  denote a probability density function for some model of the data with the parameter vector  $\theta = (\theta_1, \dots, \theta_k)$ . The Fisher information matrix of sample size  $n$  then is given by the  $k \times k$  matrix whose elements are given by the covariance between the first partial derivatives of the log-likelihood,

$$I_n(\theta)_{ij} = Cov \left[ \frac{\partial \ln f(X|\theta)}{\partial \theta_i}, \frac{\partial \ln f(X|\theta)}{\partial \theta_j} \right] \quad (11)$$

Another definition of Fisher information matrix is given based on the expected values of the second partial derivatives,

$$I_n(\theta)_{ij} = -E \left[ \frac{\partial^2 \ln f(X|\theta)}{\partial \theta_i \partial \theta_j} \right] \quad (12)$$

If no expectation is taken, one obtains a data-dependent quantity called an observed Fisher information. Here, we can give a small example: for a normal distribution with the parameters  $\mu$  and  $\sigma^2$ , the Fisher information matrix is given by

$$I_n(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \quad (13)$$

## Cramer-Rao lower bound

The Cramer-Rao lower bound provides a lower bound on the variance of any unbiased parameter estimation method and is an important tool in the assessment of a parameter estimation method [56]. The Cramer-Rao lower bound is computed as

the inverse of Fisher Information Matrix. Now we review Cramer-Rao lower bound from [101]. Let  $T(X)$  denote any statistic and let  $\psi(\theta)$  be the expectation of  $T(X)$  given by  $\psi(\theta) = E[T(X)]$ . Then for all  $\theta$ , the following is obtained,

$$\text{Var}(T(X)) \geq \frac{\left(\frac{d\psi(\theta)}{d\theta}\right)^2}{I_n(\theta)} \quad (14)$$

which is called Cramer-Rao inequality, whereas the right hand side of this inequality is called Cramer-Rao lower bound. If  $T(X)$  is an unbiased estimator for  $\theta$ , then (14) is given by,

$$\text{Var}(T(X)) \geq \frac{1}{I_n(\theta)} \quad (15)$$

When one increases  $I_n(\theta)$ , the variance becomes smaller which gives more precise information about the location of unknown parameter. This inequality can be generalized to the multi-parameter case,  $\theta = (\theta_1, \dots, \theta_k)$ ,

$$\text{Var}(T_{mult}(X)) \geq \gamma(\theta)^T I_n(\theta)^{-1} \gamma(\theta) \quad (16)$$

where  $T_{mult}(X)$  is a statistic for multi-parameter case and  $\gamma(\theta)$  is a  $k \times 1$  vector with elements  $\gamma(\theta_i) = \partial g(\theta) / \partial \theta_i$ .

# Appendix B

In the following, we will give a short overview about the biological background. This overview summarizes most of biological terminologies that have been appeared throughout the thesis.

- **HCV genotypes:** There are 6 major genotypes and they differ in their nucleotide sequence by 30 – 35%, and within an HCV genotype, 11 common subtypes (1a-c, 2a-c, 3a, 3b, 4a, 5a, and 6a) can be defined that differ in their nucleotide sequence by 20 – 25% [32].
- **NS5B [32] :** NS5B RNA-dependent RNA polymerase belongs to a class of membrane proteins. It is a key enzyme responsible for both plus and minus-strand RNA synthesis. HCV replication proceeds by the synthesis of a complementary minus-strand RNA using the plus-strand RNA as a template and the subsequent synthesis of genomic plus-strand RNA from this minus-strand RNA.
- **NS4B [32]:** NS4B is a relatively poorly characterized protein. One of its functions is to induce the formation of the replication vesicles which serves as a scaffold for the HCV replication complex.
- **NS5A [32]:** NS5A is a phosphoprotein that can be found in basally phosphorylated and hyperphosphorylated forms. Phosphorylation of NS5A is a conserved feature among hepaciviruses and pestiviruses and is found in flavivirus NS5B proteins, arguing that it has an important role in the HCV lifecycle. Some studies suggested that the phosphorylated state of NS5A modulates the efficiency of HCV RNA replication.
- **NS2-3 protease:** The NS2-3 protease is known also as an autoprotease and it is not important for RNA replication *in vitro* but is essential for the complete

replication cycle *in vitro* and *in vivo* [32]. NS2-3 protease cleaves the site between NS2 and NS3 and NS2 is a short lived protein that loses its protease activity after self-cleavage from NS3 [108].

- **NS3-NS4A protease [108]:** NS3 is a multifunctional viral protein containing a serine protease and NS4A is a cofactor of NS3 protease activity. This protease is essential for the HCV lifecycle and catalyzes HCV polyprotein cleavage at NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B junctions. It is one of the most popular viral targets for anti-HCV therapeutics. It has been shown that this protease blockades the intracellular double-stranded RNA sensor protein (RIG-I) pathway.
- **p7[32] :** P7 is a small polypeptide that is often incompletely cleaved from E2. It is not required for RNA replication *in vitro* but is essential for productive infection in *in vivo*.
- **E1 and E2 Envelope Glycoproteins:** The two envelope glycoproteins, E1 and E2 are the essential components of the HCV virion envelope and necessary for viral entry and fusion [108].
- **Core protein:** The first structural protein encoded by the HCV is the core protein which forms the viral nucleocapsid [32].
- **Low density lipoprotein receptor [108]:** Low density lipoprotein receptor is an endocytic receptor that transports lipoproteins into the cells through receptor-mediated endocytosis. Virus-like particles complexed with LDLs have been reported to enter the cells via LDL receptor.
- **SR-BI [108]:** A scavenger receptor B type I is a glycoprotein with a large extracellular loop anchored to the plasma membrane. It is considered as a fatty acylated protein located in the lipid raft domain and is highly expressed in the hepatocytes. It has been proposed as the another candidate receptor for HCV.
- **IRES [75]:** “An internal ribosome entry site is a nucleotide sequence that allows for a translation initiation in the middle of a messenger RNA (mRNA) sequence as a part of the greater process of protein synthesis. Usually, in eukaryotes, the translation can be initiated only at the 5' end of the mRNA molecule, since 5' cap recognition is required for the assembly of the initiation complex. IRES are often used by viruses as a means to ensure that the viral

translation is active during periods of time when the host translation is inhibited. These mechanisms of host translation inhibition are varied, and can be initiated by both the virus and the host, depending on the type of virus in question.”

- **CD81:** The cell surface protein which is encoded by CD81 gene and this protein mediates the signal transduction events that play a role in the regulation of cell development, activation, growth and motility [118]. It has been shown that HCV binds to this protein through its envelope glycoprotein E2 [108].
- **Endocytosis:** Endocytosis is a mechanism for cells to remove ligands, nutrients, and plasma membrane (PM) proteins, and lipids from the cell surface, bringing them into the cell interior [105]. HCV enters the cell through endocytosis [108].
- **Viral nucleocapsid [76]:** A capsid plus the enclosed nucleic acid is called a nucleocapsid. Capsid is a protein coat which encloses the nucleic acid of a virion and composed of multiple copies of one protein or a few different proteins, each of which is encoded by a single viral gene.
- **5' cap [76]:** 5' cap is a guanine nucleotide which has been added 5' end of mRNA shortly after the start of transcription. The cap protects an mRNA from an enzymatic degradation and assists in its export to the cytoplasm. The cap is also bound by a protein factor required to begin the translation in the cytoplasm .
- **Ribosomes [76]:** An enormously complex molecular machine composed of both the RNA and the protein that carry out the translation. During the translation, the ribosomes assemble and link together the amino acids in the precise order dictated by the mRNA sequence according to the nearly universal genetic code.
- **RNA polymerase:** A large enzyme that catalyzes the linkage of nucleotides into a single-stranded ribonucleic acid (RNA) [76].
- **Mutation:** Changes or mistakes which occasionally occur spontaneously during DNA or RNA replication causing changes in the sequence of nucleotides [76].
- **Gene expression:** The overall process of selectively reading and using genetic information [76].

- **Fatty acids:** Fatty acids are an important source of many cells and are stored in the form of triacylglycerols within an adipose tissue [76].
- **DNA [76]:** Deoxyribonucleic acid contains all the information required to build the cells and the tissues of an organism. The information stored in DNA is arranged in the hereditary units, known as genes, that control the identifiable traits of an organism. In the process of transcription, the information stored in DNA is copied into a ribonucleic acid (RNA).
- **mRNA [76]:** Messenger RNA carries the genetic information and the instructions from DNA that specify the correct order of amino acids during a protein synthesis. In HCV, its genome acts as an mRNA.
- **Viruses [76]:** Viruses are the parasites of the cellular genetic system. They cannot reproduce by themselves and must commandeer a host cell's machinery to synthesize viral proteins and some cases to replicate the viral genome. The entire infectious virus particle is called a virion which consists of the nucleic acid and an outer shell of protein.
- **Cell line:** A culture of cells with an indefinite lifespan which is considered immortal [76].
- **Protease [76]:** Protease is an enzyme that cleaves the polypeptide bonds in a protein. Through this mechanism, the proteases degrade a variety of proteins and peptides.
- **Nuclease [76]:** Nuclease is an enzyme that cleaves the bonds between the nucleotide subunits of nucleic acids. They degrade RNA and DNA into their mononucleotide building blocks.
- **Co-precipitation (Immunoprecipitation):** Immunoprecipitation (IP) is a method that uses the antigen-antibody reaction principle to identify a protein that reacts specifically with an antibody from mixture of proteins so that its quantity or physical characteristics can be examined [3].
- **Immunostaining [49]:** Immunostaining is a technique widely used for a co-localization of multiple peptide antigens. Immunostaining can be used for a variety of applications based on investigating the presence or the absence of a protein, its tissue distribution, its sub-cellular localization, or changes in a protein expression or degradation.

- **ER membranes**[76] : The endoplasmic reticulum is an organelle of cells that synthesizes lipids and detoxifies certain hydrophobic compounds. It also functions in the synthesis, processing and sorting of secreted proteins, lysosomal proteins and certain membranes.
- **Golgi apparatus** [76]: Golgi processes and sorts the secreted proteins, the lysosomal proteins and the membrane proteins synthesized on the rough ER before they reach to the final destination.
- **Subgenomic replicons**: Subgenomic replicon is a subset of full length replicon that is capable of autonomous replication in the Huh7 cells.
- **miRNAs** Micro RNA is a short RNA with 21 and 22 nucleotides long that hybridizes to the 3' non-translated regions of specific target mRNAs. It represses the translation of these mRNAs by yet unknown mechanism [76].
- **RNA interference** [76]: RNA interference (RNAi) is a post-transcriptional process triggered by the introduction of double-stranded RNA (dsRNA) which leads to a gene silencing in a sequence-specific manner. The studies with the extracts of *Drosophila* embryos showed that a long double-stranded RNA that mediates interference is initially processed into a double-stranded intermediate referred to as *short interfering RNA (siRNA)*.
- **Cyclosporin A (CsA)** Cyclosporin A (CsA) is a cyclic peptide of fungal origin and a potent immunosuppressant. It is used to prevent a rejection of kidney and liver transplants. It appears to act on the immune system by inhibiting the initial steps of T-lymphocyte activation [51]. It has been observed that CsA inhibits the HCV replication in vitro [32].
- **Cyclophilin B** [60]: Cyclophilin B is a secreted protein which binds to the immunosuppressive drug cyclosporin A. Cyclophilin B possesses a cellular peptidyl-prolyl ciss-trans isomerase activity and interacts with the C-terminal region of NS5B to directly stimulate its RNA binding activity, and thereby contributes to the efficient replication of HCV RNA.
- **PI4KIII** [124]: PI4KIII is an enzyme that catalyzes the synthesis of phosphatidylinositol 4-phosphate which is prevalent in the membrane of the Golgi apparatus. It was reported that an enzymatic activity is critical for HCV replication.

- **p68:** p68 also known as DEAD box protein, is a putative RNA helicase. They are implicated in a number of cellular processes involving an alteration of RNA secondary structure, such as a translation initiation, a nuclear and mitochondrial splicing, and a ribosome and spliceosome assembly [118]. It has been reported that p68-NS5B interaction may serve to mediate HCV replication [45].
- **Nucleolin:** Nucleolin is a eukaryotic nucleolar phosphoprotein involved in the synthesis and maturation of ribosomes [118].
- **hnRNP A1 [118]:** The heterogeneous nuclear ribonucleoprotein A1 (hnRNPs) are the RNA binding proteins and they form a complex with the heterogeneous nuclear RNA (hnRNA). These proteins are associated with pre-mRNAs in the nucleus and appear to influence a pre-mRNA processing and other aspects of the mRNA metabolism and transport.
- **Septin 6:** Septin 6 is a member of the septin family of GTPases and the members of this family are required for cytokinesis [118].
- **Cis-acting elements:** Cis-acting elements are regions of a non-coding DNA or RNA that bind to the transcription factors that act to repress or activate the gene expression [55].
- **PTB:** Polypyrimidine tract binding (PTB) protein is an ubiquitous cellular factor binding to RNA or single-stranded DNA, and is highly conserved during evolution [26]. It was reported that PTB may also be involved in the viral replication and has been observed to modulate an HCV IRES activity binding to the several sites within the viral genome [64].
- **hVAP-A [118]:** Human vesicle associated membrane protein A (hVAP-A) is present in the plasma membrane and the intracellular vesicles. It may also be associated with the cytoskeleton. This protein may function in a vesicle trafficking, a membrane fusion, a protein complex assembly and a cell motility.
- **Lipid rafts:** Lipid rafts are the specialized membrane microdomains that serve as organizing centers for an assembly of signaling molecules, an influence membrane fluidity and a trafficking of membrane proteins, and regulate different cellular processes such as a neurotransmission and a receptor trafficking [66].

- **Recombinant protein [76]:** A protein that is encoded from a recombinant DNA or RNA which is simply any DNA or RNA molecule composed of the sequences derived from different sources. The difference between the recombinant DNA and the DNA is only in the sequence of the nucleotides within that identical overall structure.



# Appendix C

## Publications

The work in this thesis has been partially published in the following conferences and journals:

1. Sulaimanov N, Knapp B., Mazur J., Kaderali L. *Modeling HCV Virus-Host Interactions*. Poster presentation at the Symposium on Viral Kinetic Modeling, Frankfurt, September 19-20, 2008.
2. Sulaimanov N., Binder M., Lohmann V., Bartenschlager R., Kaderali L. *Modeling the Dynamics of Hepatitis C Virus Intracellular Replication*. Poster presentation at the German Symposium on Systems Biology, Heidelberg, May 12-15, 2009.
3. Sulaimanov N., Binder M., Lohmann V., Bartenschlager R., Kaderali L. *Systems Biology of Hepatitis C Virus Replication*. Poster presentation at the Conference on Systems Biology of Mammalian Cells, Freiburg, June 3-5, 2010.
4. Sulaimanov N., Binder M., Lohmann V., Bartenschlager R., Kaderali L. *Modeling the Intracellular Dynamics of HCV*. Poster presentation at the International Conference on Systems Biology of Human Disease, Boston, June 16-18, 2010.
5. Sulaimanov N., Binder M., Lohmann V., Bartenschlager R., Kaderali L. *Systems Biology of Hepatitis C Virus Replication*. Poster presentation at the 1st Cellular Networks Conference, Heidelberg, September 23-26, 2010.
6. Sulaimanov N., Binder M., Lohmann V., Bartenschlager R., Kaderali L. *Systems Biology of Hepatitis C Virus Replication*. Poster presentation at the 11th International Conference on Systems Biology, Edinburgh, October 11-14, 2010.

7. Sulaimanov N., Binder M., Lenz S., Schlöder J., Lohmann V., Bartenschlager R., Kaderali L. *Modeling Virus-Host Interactions*. Poster presentation at the International Conference on Systems Biology of Human Disease, Boston, June 22-24, 2011.
8. Clausnitzer D., Sulaimanov N., Binder M., Lohmann V., Bartenschlager R., Kaderali L (2011). *Systembiologie der Hepatitis C-Virus-Wirts-Interaktionen*. *Laborwelt* 6:13-15.
9. Sulaimanov N., Binder M., Lenz S., Schlöder J., Lohmann V., Bartenschlager R., Kaderali L. *Replication Vesicles are Load- and Choke Points in the Hepatitis C Virus Lifecycle*. In preparation for the submission to *Molecular Systems Biology*.

# References

- [1] <http://www.who.int>. 6
- [2] <http://huh7.com/>. 7
- [3] <http://www.protocol-online.org>. 124
- [4] <http://scienceworld.wolfram.com>. 117
- [5] Wasley A and Alter MJ. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin Liver Dis*, 20(1):1–16, 2000. 5
- [6] Nezam H Afdhal. The natural history of Hepatitis C. *Semin Liver Dis*, 24 Suppl 2:3–8, 2004. 5, 6
- [7] V. Agnello, G. Abel, M. Elfahal, G. B. Knight, and Q. X. Zhang. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci U S A*, 96(22):12766–71, 1999. 9
- [8] H. Aizaki, K. J. Lee, V. M. Sung, H. Ishiko, and M. M. Lai. Characterization of the hepatitis C virus RNA replication complex associated with lipid rafts. *Virology*, 324(2):450–61, 2004. 10
- [9] Bree B Aldridge, John M Burke, Douglas A Lauffenburger, and Peter K Sorger. Physicochemical modelling of cell signalling pathways. *Nature Cell Biology*, 8(11):1195–1203, 2006. 39, 61
- [10] Anonymous. Management of hepatitis C. *NIH Consens Statement*, 15:1–41, 1997. 6
- [11] Anonymous. Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium. *J Viral Hepat*, 6:35–47, 1999. 5

- [12] Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, and Perelson AS. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- therapy. *Science*, 282(5386):103–107, 1998. 116
- [13] E. Baake, M. Baake, H.G. Bock, and K.M. Briggs. Fitting ordinary differential equations to chaotic data. *Phys. Rev. A*, 45:5524–5529, 1992. 44
- [14] E. Baake and J.P. Schlöder. Modelling the fast fluorescence rise of photosynthesis. *Bull. Math. Biol.*, 6(54):999–1021, 1992. 44
- [15] R. Bartenschlager, F. L. Cosset, and V. Lohmann. Hepatitis C virus replication cycle. *J Hepatol*, 53(3):583–5, 2010. 12
- [16] Ralf Bartenschlager, Michael Frese, and Thomas Pietschmann. Novel insights into hepatitis C virus replication and persistence. volume 63 of *Advances in Virus Research*, pages 71 – 180. Academic Press, 2004. 7
- [17] Ralf Bartenschlager and Volker Lohmann. Replication of hepatitis C virus. *Journal of General Virology*, 81(7):1631–1648, 2000. 7, 26
- [18] S. Bellentani and C. Tiribelli. The spectrum of liver disease in the general population: lesson from the Dionysos study. *J Hepatol*, 35(4):531–7, 2001. 6
- [19] M. Binder, D. Quinkert, O. Bochkarova, R. Klein, N. Kezmic, R. Bartenschlager, and V. Lohmann. Identification of determinants involved in initiation of hepatitis C virus RNA synthesis by using intergenotypic replicase chimeras. *J Virol*, 81(10):5270–83, 2007. 61, 64, 66
- [20] Marco Binder. *Characterization of host cell determinants for hepatitis C virus replication in cell culture*. PhD thesis, Heidelberg, Univ., Diss., 2007, 2007. Zsfassung in dt. Sprache. 67
- [21] A. M. Di Bisceglie. Hepatitis C. *Lancet*, 351:351–355, 1998. 5
- [22] H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics. In K.H. Ebert, P. Deuffhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, volume 18 of *Springer Series in Chemical Physics*, pages 102–125. Springer, Heidelberg, 1981. 44, 47
- [23] H.G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen*, volume 183 of *Bonner Mathematische Schriften*. Universität Bonn, Bonn, 1987. 44, 47

- [24] H.G. Bock, E. Kostina, and J.P. Schlöder. Numerical methods for parameter estimation in nonlinear differential algebraic equations. *GAMM Mitteilungen*, 30/2:376–408, 2007. 44, 45, 46, 47
- [25] H.G. Bock, E.A. Kostina, and J.P. Schlöder. On the role of natural level functions to achieve global convergence for damped Newton methods. In M.J.D. Powell and S. Scholtes, editors, *System Modelling and Optimization. Methods, Theory and Applications*. Kluwer, 2000. 46
- [26] Franck Brunel, Mario M. Zakin, Henri Buc, and Malcolm Buckle. The polypyrimidine tract binding (PTB) protein interacts with single-stranded DNA in a sequence-specific manner. *Nucleic Acids Research*, 24(9):1608–1615, 1996. 126
- [27] M. E. Burlone and A. Budkowska. Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. *J Gen Virol*, 90(Pt 5):1055–70, 2009. 9
- [28] Dan C Collins and Roni Avissar. An evaluation with the Fourier Amplitude Sensitivity Test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling. 1994. 86
- [29] Jordi Cortadella and Wolfgang Reisig, editors. *Applications and Theory of Petri Nets 2004, 25th International Conference, ICATPN 2004, Bologna, Italy, June 21-25, 2004, Proceedings*, volume 3099 of *Lecture Notes in Computer Science*. Springer, 2004. 61
- [30] R. I. Cukier, C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59(8):3873–3878, 1973. 86
- [31] R. I. Cukier, H. B. Levine, and K. E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26(1):1 – 42, 1978. 86
- [32] Moradpour D, Penin F, and Rice CM. Replication of hepatitis C virus. *Nat Rev Microbiol*, 5:453–463, 2007. 7, 8, 9, 10, 11, 13, 67, 115, 121, 122, 125
- [33] Harel Dahari, Bruno Sainz, Alan S Perelson, and Susan L Uprichard. Modeling subgenomic hepatitis C virus RNA kinetics during treatment with alpha interferon. *Journal of Virology*, 83(13):6383–6390, 2009. 116

- [34] Tomoko Date, Takanobu Kato, Michiko Miyamoto, Zijiang Zhao, Kotaro Yasui, Masashi Mizokami, and Takaji Wakita. Genotype 2a hepatitis C virus subgenomic replicon can replicate in HepG2 and IMY-N9 cells. *Journal of Biological Chemistry*, 279(21):22371–22376, 2004. 7
- [35] H. Wedemeyer et al. E. Jaekel, M. Cornberg. Treatment of acute hepatitis C with interferon alfa-2b. *N Engl J Med*, 345:1452–1457, 2001. 6
- [36] D. Egger, B. Wolk, R. Gosert, L. Bianchi, H. E. Blum, D. Moradpour, and K. Bienz. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol*, 76(12):5974–84, 2002. 10
- [37] Manfred Eigen, Christof K. Biebricher, Michael Gebinoga, and William C. Gardiner. The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry*, 30(46):11005–11018, 1991. 1
- [38] M. J. Evans, C. M. Rice, and S. P. Goff. Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proc Natl Acad Sci U S A*, 101(35):13038–43, 2004. 11
- [39] M. J. Evans, T. von Hahn, D. M. Tscherner, A. J. Syder, M. Panis, B. Wolk, T. Hatzioannou, J. A. McKeating, P. D. Bieniasz, and C. M. Rice. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature*, 446(7137):801–5, 2007. 9
- [40] Centers for Disease Control and Prevention (CDC). Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related disease. *MMWR Morb Mortal Wkly Rep*, 47:1–40, 1998. 5
- [41] P. Friebe, V. Lohmann, N. Krieger, and R. Bartenschlager. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J Virol*, 75(24):12047–57, 2001. 8
- [42] M. W. Fried, M. L. Shiffman, K. R. Reddy, C. Smith, G. Marinos, F. L. Goncales, Jr, D. Haussinger, M. Diago, G. Carosi, D. Dhumeaux, A. Craxi, A. Lin, J. Hoffman, and J. Yu. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med*, 347(13):975–82, 2002. 7

- [43] L. Gao, H. Aizaki, J. W. He, and M. M. Lai. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J Virol*, 78(7):3480–8, 2004. 11
- [44] K. R. Godfrey and J. J. DiStefano. *Identifiability of Model Parameters in Identifiability of Parametric Models*. Chapter 1. Pergamon Press, Oxford, 1987. 98
- [45] P. Y. Goh, Y. J. Tan, S. P. Lim, Y. H. Tan, S. G. Lim, F. Fuller-Pace, and W. Hong. Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. *J Virol*, 78(10):5288–98, 2004. 126
- [46] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, 1 edition, January 1989. 47, 48, 49, 110
- [47] Ju-Tao Guo, Vadim V. Bichko, and Christoph Seeger. Effect of Alpha Interferon on the Hepatitis C Virus Replicon. *J. Virol.*, 75(18):8516–8523, 2001. 55
- [48] Dahari H, Ribeiro RM, Rice CM, and Perelson AS. Mathematical modeling of subgenomic hepatitis C virus replication in Huh-7 cells. *J. Virol.*, 81(2):750–760, 2007. 1, 16, 18, 20, 21, 22, 25, 33, 35, 76, 103, 104, 108, 109, 114
- [49] G.W. Hacker and R.R. Tubbs. *Molecular morphology in human tissues: techniques and applications*. Advances in pathology, microscopy & molecular morphology. CRC Press, 2005. 124
- [50] Jacques Hadamard. *Lectures on Cauchy’s problem in linear partial differential equations / Jacques S. Hadamard*. New Haven,, 1923. 40
- [51] R E Handschumacher, M W Harding, J Rice, R J Drugge, and D W Speicher. Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science*, 226(4674):544–547, 1984. 125
- [52] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, 23(19):2612–2618, 2007. 98, 99, 103

- [53] J H Holland. *Adaptation in Natural and Artificial Systems*, volume Ann Arbor. University of Michigan Press, 1975. 48
- [54] H. Huang, F. Sun, D. M. Owen, W. Li, Y. Chen, M. Gale, Jr, and J. Ye. Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins. *Proc Natl Acad Sci U S A*, 104(14):5848–53, 2007. 13
- [55] Cristian I. and Castillo-Davis. The evolution of noncoding DNA: how much junk, how much func? *Trends in Genetics*, 21(10):533 – 536, 2005. 126
- [56] Raimund J. and Ober. The Fisher information matrix for linear systems. *Systems and amp; Control Letters*, 47(3):221 – 226, 2002. 119
- [57] J. A. Jacquez and T. Perry. Parameter estimation: local identifiability of parameters. *Am J Physiol*, 258(4 Pt 1):E727–36, 1990. 101, 102
- [58] John A Jacquez and Peter Greif. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences*, 77(1-2):201–227, 1985. 102
- [59] C. L. Jopling, M. Yi, A. M. Lancaster, S. M. Lemon, and P. Sarnow. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*, 309(5740):1577–81, 2005. 11
- [60] Watashi K, Ishii N, Hijikata M, Inoue D, Murata T, Miyanari Y, and Shimotohno K. Cyclophilin B is a functional regulator of hepatitis C virus RNA polymerase. *Mol. Cell*, 19:111–122, 2005. 11, 125
- [61] J. Kallrath, H.G. Bock, and J.P. Schlöder. Least squares parameter estimation in chaotic differential equations. *Celestial Mechanics and Dynamical Astronomy*, 56:353–371, 1993. 44
- [62] Takanobu Kato, Tomoko Date, Michiko Miyamoto, Akihiro Furusaka, Katsutoshi Tokushige, Masashi Mizokami, and Takaji Wakita. Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology*, 125(6):1808–1817, 2003. 7
- [63] Artur Kaul, Sarah Stauffer, Carola Berger, Thomas Pertel, Jennifer Schmitt, Stephanie Kallis, Margarita Zayas Lopez, Volker Lohmann, Jeremy Luban, and Ralf Bartenschlager. Essential role of cyclophilin A for hepatitis C virus

- replication and virus production and possible link to polyprotein cleavage kinetics. *PLoS Pathog*, 5(8):e1000546, 08 2009. 11
- [64] C. S. Kim, S. K. Seol, O. K. Song, J. H. Park, and S. K. Jang. An RNA-binding protein, hnRNP A1, and a scaffold protein, septin 6, facilitate hepatitis C virus replication. *J Virol*, 81(8):3852–65, 2007. 126
- [65] A. Kirsch. *An introduction to the mathematical theory of inverse problems*. Applied mathematical sciences. Springer, 1996. 40
- [66] Zeljka Korade and Anne K Kenworthy. Lipid rafts, cholesterol, and the brain. *Neuropharmacology*, 55(8):1265–73, 2008. 126
- [67] G. Koutsoudakis, E. Herrmann, S. Kallis, R. Bartenschlager, and T. Pietschmann. The level of CD81 cell surface expression is a key determinant for productive entry of hepatitis C virus into host cells. *J Virol*, 81(2):588–98, 2007. 9
- [68] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, and N. Shah. Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering and System Safety*, 94(7):1135 – 1148, 2009. Special Issue on Sensitivity Analysis. 86
- [69] G M Leclerc, F R Boockfor, W J Faught, and L S Frawley. Development of a destabilized firefly luciferase enzyme for measurement of gene expression. *Biotechniques*, 29(3):590–591, 594–596, 598 passim, 2000. 36
- [70] Y. Lecourtier and others. *Volterra and Generating Power Series Approaches to Identifiability in Identifiability of Parametric Models*. Chapter 5. Pergamon Press, Oxford, 1987. 99
- [71] Simon M. Lenz, H. Georg Bock, Johannes P. Schlöder, Ekaterina A. Kostina, Gottlob Gienger, and Gerald Ziegler. Multiple Shooting Method for initial satellite orbit determination. *Journal of Guidance Control and Dynamics*, 33:1334–1346, 2010. 44, 46
- [72] Yun-Sook Lim and Soon B. Hwang. Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type III  $\alpha$  and regulates viral propagation. *Journal of Biological Chemistry*, 286(13):11290–11298, 2011. 12
- [73] L. Ljung. *Model validation and model error modeling*. The Aström Symposium on Control. Studentlitteratur, Lund, Sweden, 1999a. 100

- [74] L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2):265 – 276, 1994. 99, 101
- [75] Books Llc. *Internal Ribosome Entry Site: Hepatitis C Virus Internal Ribosome Entry Site, Picornavirus Internal Ribosome Entry Site*. General Books LLC, 2010. 122
- [76] Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, Anthony Bretscher, Hidde Ploegh, and Paul Matsudaira. *Molecular Cell Biology (Lodish, Molecular Cell Biology)*. W. H. Freeman, 6th edition, June 2007. 9, 25, 36, 123, 124, 125, 127
- [77] Harvey F. Lodish and Marion Jacobsen. Regulation of hemoglobin synthesis. *Journal of Biological Chemistry*, 247(11):3622–3629, 1972. 33, 35
- [78] Volker Lohmann, Sandra Hoffmann, Ulrike Herian, Francois Penin, and Ralf Bartenschlager. Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *Journal of Virology*, 77(5):3007–3019, 2003. 61
- [79] J. M. Lyle, E. Bullitt, K. Bienz, and K. Kirkegaard. Visualization and functional analysis of RNA-dependent RNA polymerase lattices. *Science*, 296(5576):2218–22, 2002. 10, 73
- [80] Charlton M. Hepatitis C infection in liver transplantation. *Am J Transplant*, 1(3):197–203, 2001. 5
- [81] K. Krawczynski et al. M. J. Alter, H. S. Margolis. The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. *N Engl J Med*, 327:1899–1905, 1992. 6
- [82] Han Ma, Vincent Leveque, Anniek De Witte, Weixing Li, Than Hendricks, Saundra M. Clausen, Nick Cammack, and Klaus Klumpp. Inhibition of native hepatitis C virus replicase by nucleotide and non-nucleoside inhibitors. *Virology*, 332(1):8 – 15, 2005. 35
- [83] M. P. Manns, J. G. McHutchison, S. C. Gordon, V. K. Rustgi, M. Shiffman, R. Reindollar, Z. D. Goodman, K. Koury, M. Ling, and J. K. Albrecht. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet*, 358(9286):958–65, 2001. 7

- [84] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J Theor Biol*, 254(1):178–96, 2008. 81, 86, 89
- [85] M Martell, J I Esteban, J Quer, J Genesca, A Weiner, R Esteban, J Guardia, and J Gomez. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *Journal of Virology*, 66(5):3225–3229, 1992. 26
- [86] MATLAB. *Global Optimization Toolbox, version 7.10.0 (R2010a)*. The Math-Works Inc., Natick, Massachusetts, 2010. 47, 110
- [87] Harley H. McAdams and Adam Arkin. Stochastic mechanisms in gene-expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997. 115
- [88] G W McCaughan, P H McGuinness, G A Bishop, D M Painter, A S Lien, R Tulloch, B R Wylie, and G T Archer. Clinical assessment and incidence of Hepatitis C RNA in 50 consecutive RIBA-positive volunteer blood donors. *The Medical journal of Australia*, 157(4):231–233, 1992. 6
- [89] Alison K McLean, Fabio Luciani, and Mark M Tanaka. Trade-offs in resource allocation in the intracellular life-cycle of hepatitis C virus. *Journal of Theoretical Biology*, 267(4):565–572, 2010. 113, 114
- [90] J. C. Meunier, R. S. Russell, R. E. Engle, K. N. Faulk, R. H. Purcell, and S. U. Emerson. Apolipoprotein c1 association with hepatitis C virus. *J Virol*, 82(19):9647–56, 2008. 13
- [91] Hongyu Miao, Xiaohua Xia, Alan S Perelson, and Hulin Wu. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, 53(1):3–39, 2011. 97, 98, 99, 101
- [92] Elena L. Mishchenko, Kirill D. Bezmaternykh, Vitali A. Likhoshvai, Alexander V. Ratushny, Tamara M. Khlebodarova, Natalia Yu. Sournina, Vladimir A. Ivanisenko, and Nikolay A. Kolchanov. Mathematical model for suppression of subgenomic hepatitis C virus RNA replication in cell culture. *J. Bioinformatics and Computational Biology*, 5(2b):593–609, 2007. 113, 114
- [93] M. Mitchell. *An introduction to genetic algorithms*. Complex adaptive systems. MIT Press, 1998. 51

- [94] Y. Miyanari, K. Atsuzawa, N. Usuda, K. Watashi, T. Hishiki, M. Zayas, R. Bartenschlager, T. Wakita, M. Hijikata, and K. Shimotohno. The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol*, 9(9):1089–97, 2007. 13
- [95] Yusuke Miyanari, Makoto Hijikata, Masashi Yamaji, Masahiro Hosaka, Hitoshi Takahashi, and Kunitada Shimotohno. Hepatitis C Virus Non-structural Proteins in the Probable Membranous Compartment Function in Viral Genome Replication. *Journal of Biological Chemistry*, 278(50):50301–50308, 2003. 35
- [96] Alter MJ. The epidemiology of acute and chronic hepatitis C. *Clin Liver Dis*, 1(3):559–68, 1997. 5
- [97] Alter MJ, Margolis HS, Krawczynski K, Judson FN, Mares A, Alexander WJ, Hu PY, Miller JK, Gerber MA, Sampliner RE, and et al. The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. *N Engl J Med*, 327(27):1899–1905, 1992. 5
- [98] D. Moradpour, F. Penin, and C. M. Rice. Replication of hepatitis C virus. *Nat Rev Microbiol*, 5(6):453–63, 2007. 8
- [99] Kohji Moriishi and Yoshiharu Matsuura. Host factors involved in the replication of hepatitis C virus. *Reviews in Medical Virology*, 17(5):343–354, 2007. 67
- [100] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):pp. 161–174, 1991. 86
- [101] Jay I. Myung and Daniel J. Navarro. *Information Matrix*. John Wiley and Sons, Ltd, 2005. 119, 120
- [102] Dixit NM, Layden-Almer JE, Layden TJ, and Perelson AS. Modeling how ribavirin improves interferon response rates in hepatitis C virus infection. *Nature*, 432:922–924, 2004. 116
- [103] Marek Obitko. <http://www.obitko.com/tutorials/genetic-algorithms>, 1998. 51

- [104] National Institutes of Health Consensus Development Conference Statement. Management of Hepatitis C 2002 (June 10-12, 2002). *Gastroenterology*, 123(6):2082–99, 2002. 5, 6
- [105] Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 27(1):29–34, 1999. 123
- [106] Jong-Won Oh, Takayoshi Ito, and Michael M. C. Lai. A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J. Virol.*, 73(9):7694–7702, 1999. 35
- [107] Richard D. Palmiter. Ovalbumin Messenger Ribonucleic Acid Translation. *Journal of Biological Chemistry*, 248(6):2095–2106, 1973. 33, 35
- [108] Jean-Michel Pawlotsky. HCV Genome and Life Cycle. *Organization*, pages 5–47, 2006. 122, 123
- [109] I. M. Pedersen, G. Cheng, S. Wieland, S. Volinia, C. M. Croce, F. V. Chisari, and M. David. Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature*, 449(7164):919–22, 2007. 11
- [110] Lawrence Perko. Differential equations and dynamical systems. *Bulletin of Mathematical Biology*, 54(6):1086–1088, 1991. 13, 15
- [111] F. Pessione, F. Degos, P. Marcellin, V. Duchatelle, C. Njapoum, M. Martinot-Peignoux, C. Degott, D. Valla, S. Erlinger, and B. Rueff. Effect of alcohol consumption on serum hepatitis C virus RNA and histological lesions in chronic hepatitis C. *Hepatology*, 27(6):1717–22, 1998. 6
- [112] M. G. Peters and N. A. Terrault. Alcohol use and hepatitis C. *Hepatology*, 36(5 Suppl 1):S220–5, 2002. 6
- [113] U. Pfeifer, R. Thomssen, K. Legler, U. Bottcher, W. Gerlich, E. Weinmann, and O. Klinge. Experimental non-A, non-B hepatitis: four types of cytoplasmic alteration in hepatocytes of infected chimpanzees. *Virchows Arch B Cell Pathol Incl Mol Pathol*, 33(3):233–43, 1980. 10
- [114] T. Pietschmann, V. Lohmann, G. Rutter, K. Kurpanek, and R. Bartenschlager. Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. *J Virol*, 75(3):1252–64, 2001. 35

- [115] P. Pileri, Y. Uematsu, S. Campagnoli, G. Galli, F. Falugi, R. Petracca, A. J. Weiner, M. Houghton, D. Rosa, G. Grandi, and S. Abrignani. Binding of hepatitis C virus to CD81. *Science*, 282(5390):938–41, 1998. 9
- [116] H. Pohjanpalo. System identifiability based on the power series expansion of the solution. *Mathematical Biosciences*, 41(1-2):21 – 33, 1978. 99
- [117] Huntington Potter and Richard Heller. *Transfection by Electroporation*. John Wiley & Sons, Inc., 2001. 108
- [118] K Pruitt, T Tatusova, and J Ostell. The Reference Sequence (RefSeq) Database. Website. <http://www.ncbi.nlm.nih.gov/books/NBK21091/>. 123, 126
- [119] T. Quaiser and M. Mönnigmann. Systematic identifiability testing for unambiguous mechanistic modeling - application to JAK-STAT, MAP kinase, and NF-kappaB signaling pathway models. *BMC Systems Biology*, 3(50), 2009. 101
- [120] Doris Quinkert, Ralf Bartenschlager, and Volker Lohmann. Quantitative Analysis of the Hepatitis C Virus Replication Complex. *J. Virol.*, 79(21):13594–13605, 2005. 10, 16, 18, 19, 35, 55, 57, 67, 69, 73, 76, 108, 109, 110, 112
- [121] Gosert R, Egger D, Lohmann V, Bartenschlager R, Blum HE, Bienz K, and Moradpour D. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J. Virol.*, 77(9):5487–5492, 2003. 10
- [122] G. Randall, M. Panis, J. D. Cooper, T. L. Tellinghuisen, K. E. Sukhodolets, S. Pfeffer, M. Landthaler, P. Landgraf, S. Kan, B. D. Lindenbach, M. Chien, D. B. Weir, J. J. Russo, J. Ju, M. J. Brownstein, R. Sheridan, C. Sander, M. Zavolan, T. Tuschl, and C. M. Rice. Cellular cofactors affecting hepatitis C virus infection and replication. *Proc Natl Acad Sci U S A*, 104(31):12884–9, 2007. 11
- [123] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37:81–91, 1945. 100
- [124] Simon Reiss, Ilka Rebhan, Perdita Backes, Ines Romero-Brey, Holger Erfle, Petr Matula, Lars Kaderali, Marion Poenisch, Hagen Blankenburg, Marie-Sophie Hiet, Thomas Longerich, Sarah Diehl, Fidel Ramirez, Tamas Balla,

- Karl Rohr, Artur Kaul, Sandra Buhler, Rainer Pepperkok, Thomas Lengauer, Mario Albrecht, Roland Eils, Peter Schirmacher, Volker Lohmann, and Ralf Bartenschlager. Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell Host and amp; Microbe*, 9(1):32 – 45, 2011. 11, 125
- [125] M. Rodriguez-Fernandez and J. R. Banga. SensSB: a software toolbox for the development and sensitivity analysis of systems biology models. *Bioinformatics*, 26(13):1675–6, 2010. 104
- [126] Maria Rodriguez-Fernandez, Jose A Egea, and Julio R Banga. Novel meta-heuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*, 7(1):483, 2006. 99
- [127] Maria Rodriguez-Fernandez, Pedro Mendes, and Julio R Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Bio Systems*, 83(2-3):248–265, 2006. 99
- [128] H. Sakamoto, K. Okamoto, M. Aoki, H. Kato, A. Katsume, A. Ohta, T. Tsukuda, N. Shimma, Y. Aoki, M. Arisawa, M. Kohara, and M. Sudoh. Host sphingolipid biosynthesis as a target for hepatitis C virus therapy. *Nat Chem Biol*, 1(6):333–7, 2005. 10
- [129] A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity Analysis*. Wiley series in probability and statistics. Wiley, 2000. 81
- [130] A. Saltelli, S. Tarantola, and K. P.-S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56, February 1999. 86
- [131] A. Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, 1 edition, April 2004. 86
- [132] Andrea Saltelli and Ricardo Bolado. An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics and amp; Data Analysis*, 26(4):445 – 460, 1998. 86
- [133] G. Saracco, A. Olivero, A. Ciancio, S. Carenzi, and M. Rizzetto. Therapy of chronic hepatitis C: a critical review. *Curr Drug Targets Infect Disord*, 3(1):25–32, 2003. 7

- [134] E. Scarselli, H. Ansuini, R. Cerino, R. M. Roccasecca, S. Acali, G. Filocamo, C. Traboni, A. Nicosia, R. Cortese, and A. Vitelli. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J*, 21(19):5017–25, 2002. 9
- [135] J H Schaibly and K E Shuler. Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 2. applications. *Journal Of Chemical Physics*, 59(8):3879–3888, 1973. 86
- [136] M. Schwartz, J. Chen, M. Janda, M. Sullivan, J. den Boon, and P. Ahlquist. A positive-strand RNA virus replication complex parallels form and function of retrovirus capsids. *Mol Cell*, 9(3):505–14, 2002. 10, 73
- [137] S. T. Shi, K. J. Lee, H. Aizaki, S. B. Hwang, and M. M. Lai. Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2. *J Virol*, 77(7):4160–8, 2003. 10
- [138] Chen SL and Morgan TR. Natural history of hepatitis C virus (HCV) infection. *Int J Med Sci*, 3:47–52, 2006. 5, 6, 7
- [139] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271 – 280, 2001. 86, 118
- [140] J.C. Spall. Cramer-Rao bounds and Monte Carlo calculation of the Fisher information matrix in difficult problems. In *American Control Conference, 2004. Proceedings of the 2004*, volume 4, pages 3140 –3145, 30-July 2 2004. 119
- [141] R Srivastava, L You, J Summers, and J Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *Journal of Theoretical Biology*, 218(3):309–321, 2002. 115
- [142] Tetsuro Suzuki, Koji Ishii, Hideki Aizaki, and Takaji Wakita. Hepatitis C viral life cycle. *Advanced Drug Delivery Reviews*, 59(12):1200 – 1212, 2007. 10
- [143] Z. Szallasi, J. Stelling, and V. Periwal. *System modeling in cell biology: from concepts to nuts and bolts*. Bradford Book. MIT Press, 2006. 98, 99
- [144] Poynard T, Ratziu V, Benhamou Y, Opolon P, Cacoub P, and Bedossa P. Natural history of HCV infection. *Baillieres Best Pract Res Clin Gastroenterol*, 14(2):211–28, 2000. 5

- [145] P Opolon et al. T. Poynard, P Bedossa. Natural history of liver fibrosis progression in patients with chronic hepatitis C. *Lancet*, 349:825–832, 1997. 5, 6
- [146] Hengli Tang and Henry Grise. Cellular and molecular biology of HCV infection and hepatitis. *Clinical Science*, 117(2):49–65, 2009. 11, 12
- [147] Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*, volume 1. SIAM, 2005. 39
- [148] Timothy L. Tellinghuisen, Matthew J. Evans, Thomas von Hahn, Shihyun You, and Charles M. Rice. Studying hepatitis C virus: Making the best of a bad virus. *J. Virol.*, 81(17):8853–8867, 2007. 7, 9, 67
- [149] D. L. Thomas, J. Astemborski, R. M. Rai, F. A. Anania, M. Schaeffer, N. Galai, K. Nolt, K. E. Nelson, S. A. Strathdee, L. Johnson, O. Laeyendecker, J. Boitnott, L. E. Wilson, and D. Vlahov. The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA*, 284(4):450–6, 2000. 7
- [150] John F. Thompson, Lisa S. Hayes, and David B. Lloyd. Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene*, 103(2):171–177, 1991. 36
- [151] B. J. Thomson and R. G. Finch. Hepatitis C Virus infection. *Clinical Microbiology and Infection*, 11(2):86–94, 2005. 6, 7
- [152] M. J. Tong, N. S. el Farra, A. R. Reikes, and R. L. Co. Clinical outcomes after transfusion-associated hepatitis C. *N Engl J Med*, 332(22):1463–6, 1995. 5, 6
- [153] T. Umehara, M. Sudoh, F. Yasui, C. Matsuda, Y. Hayashi, K. Chayama, and M. Kohara. Serine palmitoyltransferase inhibitor suppresses HCV replication in a mouse model. *Biochem Biophys Res Commun*, 346(1):67–73, 2006. 10
- [154] Lohmann V, Koerner F, Koch JO, Herian U, Theilmann L, and Bartenschlager R. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science*, 285(5424):110–113, 1999. 7, 35, 55
- [155] S. Vajda, K. R. Godfrey, and H. Rabitz. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math Biosci*, 93(2):217–48, 1989. 99

- [156] A. Varma, M. Morbidelli, and H. Wu. *Parametric sensitivity in chemical systems*. Cambridge series in chemical engineering. Cambridge University Press, 1999. 81, 82
- [157] Takaji Wakita, Thomas Pietschmann, Takanobu Kato, Tomoko Date, Michiko Miyamoto, Zijiang Zhao, Krishna Murthy, Anja Habermann, Hans-Georg Kräusslich, Masashi Mizokami, and et al. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Medicine*, 11(7):791–6, 2005. 8
- [158] Chunfu Wang, Jill Pflugheber, Jr. Sumpter, Rhea, Donald L. Sodora, Daniel Hui, Ganes C. Sen, and Jr. Gale, Michael. Alpha interferon induces distinct translational control programs to suppress hepatitis C virus RNA replication. *J. Virol.*, 77(7):3898–3912, 2003. 33, 35
- [159] Thomas Weise. *Global Optimization Algorithms. Theory and Application.*, 2008. 48, 49, 50, 51
- [160] Benno Wölk, Benjamin Büchele, Darius Moradpour, and Charles M. Rice. A dynamic view of hepatitis C virus replication complexes. *Journal of Virology*, 82(21):10519–10531, November 1, 2008. 109
- [161] X Xia and C H Moog. Identifiability of nonlinear systems with application to HIV/AIDS models. *IEEE Transactions on Automatic Control*, 48(2):330–336, 2003. 98
- [162] Feng Yang, Jason M. Robotham, Heather B. Nelson, Andre Irsigler, Rachael Kenworthy, and Hengli Tang. Cyclophilin A is an essential cofactor for hepatitis C virus infection and the principal mediator of Cyclosporine resistance in vitro. *Journal of Virology*, 82(11):5269–5278, June 1, 2008. 11
- [163] Mitsutoshi Yoneyama and Takashi Fujita. Function of RIG-I-like receptors in antiviral innate immunity. *Journal of Biological Chemistry*, 282(21):15315–15318, 2007. 116