

New Algorithms for Host Pathogen Systems Biology

Reporting

Project Information

SYSPATHO

Grant agreement ID: 260429

Status
Closed project

Start date
1 October 2010

End date
31 March 2015

Funded under
FP7-HEALTH

Overall budget
€ 3 882 716,41

EU contribution
€ 2 942 424,75

Coordinated by
RUPRECHT-KARLS-
UNIVERSITAET HEIDELBERG
 Germany

Final Report Summary - SYSPATHO (New Algorithms for Host Pathogen Systems Biology)

Executive Summary:

SYSPATHO focused on the development of novel and generally applicable mathematical methods and algorithms for systems biology. These methods and algorithms were applied to study the complex interactions of hepatitis C virus (HCV), a human-pathogenic virus of high medical relevance, with its host at the systems level. Using a multidisciplinary, integrative approach, SYSPATHO (a) developed methods to analyze and integrate a wide variety of data from wet lab experiments, databases and biological literature, (b) developed and applied machine learning tools to reconstruct and study intracellular interaction networks from experimental data, (c) developed new and improved existing algorithms and mathematical methods for bottom-up modelling, to fit models to data, and to analyze the dynamic behavior of models (d) generated new experimental data to gain novel insights into hepatitis C virus host interactions, and (e) used the newly developed methods and data to model and analyze HCV-host interactions at the systems level. Guided by biological data, SYSPATHO focused on the design of novel algorithms and mathematical methods for systems biology, providing generally applicable tools to

elucidate biological processes. Based on developed models and using systems analysis, SYSPATHO elucidated virus host interactions of Hepatitis C virus at an unprecedented level. As a direct spin-off, models and analysis methods developed in SYSPATHO led to the identification of sensitive candidate host cell target processes which may be applicable for the design of novel anti-viral drugs against hepatitis C. Targeting of host cell factors reduces the likelihood for the development of therapy resistance and increase the chance for broad-spectrum antivirals. Inclusion of two SME partners ensured further exploitation of results generated in SYSPATHO and their transfer into industrial and pharmaceutical applications. Thus, SYSPATHO made a significant contribution towards the economy and health care systems in Europe.

Project Context and Objectives:

The overall aim of SYSPATHO was to contribute significantly to several main technological and methodological developments in systems biology, by focusing on the improvement of existing and the development of new mathematical algorithms. In addition, by applying and rigorously testing the developed methods to model the interaction of the medically highly relevant human pathogen hepatitis C virus (HCV) with the innate immune response and the host cell, SYSPATHO contributed to the development of novel antiviral concepts.

The main methodological contributions resulting from SYSPATHO are:

- * The implementation of new methods for data preprocessing and analysis, including live cell imaging data and high-throughput screening data
- * The development and improvement of methods for text-mining and data integration
- * The development and implementation of new methods to reconstruct signal transduction networks using machine learning from high-throughput data, and the development of tools to integrate bottom-up modelling with machine learning approaches
- * The development of mathematical methods to solve highly non-linear continuous and discrete equations
- * The improvement of methods to solve the inverse problem of mathematical modelling, as well as methods to verify the reliability of parameter estimates (identifiability analysis)
- * The development of improved mathematical methods for model analysis.
- * The developed methods will be applied to a description at the systems level of hepatitis C virus - host interaction, with a particular focus on viral replication and the interplay of HCV with innate immune response.

SYSPATHO was funded under the European Call HEALTH.2010.2.1.2-3 “Developing new and improving existing mathematics algorithms for systems biology” for small or medium-scaled focused research projects, with a focus on Eastern Europe and Central Asia, in particular Russia. As such, SYSPATHO was focused on multidisciplinary research which integrates a wide variety of biological data, and develops and applies systems approaches to understand and model biological processes at all levels of organization. Our research focused on the design of algorithms for modeling complex biological systems, and algorithms developed are of general applicability for the field of systems biology. Algorithms developed within SYSPATHO were thoroughly tested on models for HCV infection. A particular aim was to increase knowledge of biological processes relevant for HCV infection, to transport this knowledge into clinical applications.

SYSPATHO established a close collaboration between EEC, EU and associated countries, in particular

SYSPATHO established a close collaboration between EECA, EU and associated countries, in particular with Russia. As such, project partners within SYSPATHO would benefit from mutual exchange of information and researchers, as well as a combination of efforts, with the aim to ultimately gain leadership in key scientific and technology areas through collaborative research. SYSPATHO ensured these aims through its research program and organizational structure, as well as specific measures implemented in SYSPATHO to tighten and extend collaborations between EU/associated countries and other projects, scientists and SMEs in the EECA area. By combining the expertise of 9 academic and 2 industrial partners, of which 3 are from two different regions in Russia and the remaining partners are located in 5 different EU / associated countries, SYSPATHO not only brought together a critical mass of scientific excellence from different European and Russian institutions, but also led to a much closer collaboration with partners in the Eastern Europe / Central Asia region. This objective was further strengthened by tight collaboration with other projects in the EE/CA region, in some of which partners in SYSPATHO were already involved; these collaborations were underlined and expanded further by the organization of an European-Russian workshop on computational systems biology, and by mutual exchange of scientists at all levels of their scientific career. Through this tight collaboration and exchange of personnel, SYSPATHO not only achieved leadership in key areas in systems biology through the combination of knowledge and efforts, but also educated young scientists and promoted their excellence through the exchange of personnel and participation in an international research effort.

Using a multidisciplinary, integrative approach, SYSPATHO

- a) developed methods to analyze and integrate a wide variety of data from wet lab experiments, databases and biological literature,
- b) developed and applied machine learning tools and graph algorithms to reconstruct and study intracellular interaction networks from experimental data,
- c) developed new and improved existing algorithms and mathematical methods for bottom-up modeling, to fit models to data, and to analyze the dynamic behavior of resulting models
- d) generated new experimental data that will provide novel insights into Hepatitis C virus host interactions, and
- e) used the newly developed methods and data to model and analyze HCV-host interactions at the systems level.

Guided by clinical data, SYSPATHO focused on the design of novel algorithms and mathematical methods for systems biology, with the aim to provide generally applicable tools to elucidate underlying biological processes. While mathematical methods and algorithms developed in SYSPATHO were applied and thoroughly tested on models of Hepatitis-C virus interactions with the host, they are of general applicability in the broader area of systems biology, and contribute also to the fields of applied mathematics, mathematical modeling, bioinformatics and computer science.

HCV infection is a major global health problem with 170 million chronically infected individuals worldwide and 3 to 4 million new infections occurring each year (Rantala et al., 2008). The data available for Europe indicate a wide variation in HCV prevalence, ranging between 0.1 to 6.0% in various countries (Esteban et al., 2008). It is assumed that about 86.000 deaths due to chronic hepatitis C occurred in the WHO European region in 2002. This is more than twice the number of deaths estimated for HIV/AIDS (Mühlberger et al., 2009). A major reason is the insidious course of diseases; HCV infection persists in up to 80% of cases and is mostly asymptomatic. However, these persons are at high risk to develop liver

to 60% of cases and is mostly asymptomatic. However, these persons are at high risk to develop liver disease most notably liver cirrhosis and hepatocellular carcinoma (HCC). Currently, there is no vaccine available. Moreover, the standard-of-care treatment consisting of pegylated interferon-alpha and ribavirin is costly, has limited efficacy, serious side-effects and poor tolerability. HCV forms the sole genus Hepacivirus in the family Flaviviridae. The single-stranded, about 9,600 nucleotides long viral genome is of positive polarity and encodes a single polyprotein approximately 3,000 amino acids in length. It is cleaved by host and viral proteases into at least 10 different products (reviewed in Moradpour et al., 2007). The N-terminal region of the polyprotein is composed of the structural protein core and the envelope glycoproteins 1 (E1) and E2. These proteins are the major building blocks of the virus particle. Two auxiliary proteins (p7 and NS2) are required for the assembly of infectious HCV particles, but it is unclear whether they are also part of the virus particle (Steinmann et al., 2007; Jones et al., 2007). The remainder of the NS proteins is sufficient for replication of the HCV genome (Lohmann et al., 1999). NS3 contains a serine-type protease in its N-terminal domain that is tightly associated with and activated by NS4A. The C-terminal NS3 domain contains NTPase/helicase activities. NS4B is a protein that induces the formation of the so-called membranous web that probably harbors the replicase complex. NS5A is an important regulator of RNA replication and virus assembly and NS5B is the RNA-dependent RNA polymerase (RdRp), the key enzyme of viral genome amplification (reviewed in Bartenschlager et al., 2006).

A hallmark of HCV is its high propensity to establish persistence. In fact, up to 80% of acute infections become chronic arguing that HCV has devised strategies to combat adaptive and innate antiviral defense. Retinoic acid inducible gene-I (RIG-I) as well as Toll-like receptor molecule 3 (TLR-3 detecting ssRNA) presumably are the key sensory molecules detecting HCV RNA and activating a signal transduction cascade that normally results in expression and secretion of interferon-beta (IFN-beta) and other cytokines. However, this signaling is blocked by the NS3/4A protease that cleaves two essential adaptor proteins (MAVS and TRIF, required for RIG-I and TLR 3 dependent signaling, respectively) and therefore, IFN-beta secretion is blocked once sufficient amounts of protease have been expressed in the infected cell (Meylan et al. 2005; Li et al., 2005). Nevertheless, treatment of cells with type 1 IFN potently suppresses HCV replication and this property is used in current treatment of chronic hepatitis C. Dahari et al (2009) have recently described a mathematical model of the dynamics of HCV RNA kinetics under type 1 IFN treatment, which describes experimental data well at the level of cell populations. Yet, although it is known that IFN-stimulation of cells induces the expression of up to 400 genes, it is unclear which of the corresponding gene products is responsible for inhibition of HCV replication (reviewed in Weber 2007). Moreover, for most of these genes we do not know the molecular mechanism by which they contribute to the antiviral state. SYSPATHO made a major contribution to the identification of those factors being involved in the inhibition of HCV replication and in the molecular mechanisms underlying this inhibition.

Based on developed models and using systems analysis, SYSPATHO elucidated virus host interactions of Hepatitis C virus at an unprecedented level. As a direct spin-off, models and analysis methods developed in SYSPATHO led to the identification of new candidate target genes for the design of anti-viral drugs against Hepatitis C. Using the systems approach pursued in SYSPATHO, host processes required by the virus were identified, targeting of which – if tolerated by the host – could make it much harder for HCV to develop resistance against the drug compound than simply to change the structure of a viral protein slightly to evade drug binding. Inclusion of two SME partners in SYSPATHO, and setting up a spin-off company ensures exploitation and transfer of results into industrial and pharmaceutical applications, even beyond the project's duration.

beyond the project's duration.

To achieve our mission, we defined six specific scientific and technological objectives:

- * To collect experimental data on Hepatitis C virus host interactions using literature and database mining approaches, as well as a experimental assays using RNAi interference, live cell imaging, over-expression experiments and yeast-2-hybrid assays.
- * To develop Bioinformatics tools for processing and analyzing the experimental data, as well as to integrate the heterogeneous data types for modeling.
- * To set up a mathematical model of Hepatitis C virus replication.
- * To develop methods to integrate host factors into the replication model, using machine learning and automatic, data-driven network reconstruction methods as well as direct modeling approaches, and use these methods to integrate host processes into the replication model
- * To develop methods for parameter estimation, model analysis and for the identification of load and choke points in models, and analyze the Hepatitis-C host model using these methods.
- * To identify potential drug targets from model analysis and possible candidate drug molecules to interfere with viral replication.

Project Results:

** Summary **

Within SYSPATHO, experimental data acquired by partners were integrated with other publicly available biological information from text mining, public databases and other sources. Furthermore, novel experimental data using screening and live cell imaging approaches were acquired to further characterize HCV-host interactions, and were used to develop mathematical models for viral replication and interactions with the innate immune system of the infected cell spanning different level of biological organization, including interactions of cells in populations, cellular level, as well as zooming into subcellular processes to study viral particle diffusion. The developed models were analyzed using our newly developed methods, to identify load and choke points of both viral replication and viral interference with cellular immune response. SYSPATHO thereby contributed to a better understanding of the pathogenicity and persistence of HCV infections, the interplay between virus and host immune defense, as well as to the identification of novel potential targets for anti viral drug design.

The methods and mathematical algorithms contributed by SYSPATHO are of paramount importance for the field of systems biology, and are of wide and general applicability to the entire field. They span a broad range of applications, encompassing the full cycle of systems biology model development and model analysis. Starting with data preprocessing, normalization and statistical data analysis as a prerequisite for all further modeling attempts, SYSPATHO contributed new methods for the analysis of high-throughput, time-resolved experimental data. This experimental data was complemented by methods to integrate data from literature using new and improved text-mining approaches, as well as data automatically parsed from databases. Based on the experimental data and integrating additional knowledge from literature and databases, SYSPATHO developed methods to automatically reconstruct underlying biological processes using machine learning and network reconstruction approaches, a process that complements traditional bottom-up modeling. SYSPATHO furthermore developed new methods to solve model equations, a key task in all systems biology projects. Improved methods and new tools were developed for parameter estimation and for verification of the reliability of parameter estimates through the identifiability analysis. Furthermore, SYSPATHO developed improved methods and novel tools to analyze the dynamic behavior

Furthermore, SYSPATHO developed improved methods and novel tools to analyze the dynamic behavior and sensitivity of the developed models. These are central to tasks such as model predictions, hypothesis generation and further experiment planning. Finally, based on results of model analysis, SYSPATHO developed new tools to screen for potential compounds that interfere at load and choke points identified in the developed models, with the aim to suggest candidate molecules that can be tested experimentally. The developed methods were rigorously tested on models of hepatitis C virus replication and its interaction with the host immune system, thus providing novel insights into this medically highly relevant viral pathogen.

** Biological discoveries **

A number of fundamental biological discoveries were made within SYSPATHO about how HCV targets the host cell, elicits an antiviral response and interacts with the cell's metabolism. An increasing number of observations now suggest that cellular metabolism and innate immunity are reciprocally interfering for an optimized adaptation to environmental perturbations, including viral infections. Viruses also manipulate host cell metabolism to ensure supply of energy and nutrients necessary for their replication and propagation.

Modulation of cellular metabolism as a factor of an effective IFN-response and a viral strategy to impair innate immunity has yet to be addressed. Using our last generation interactomic database, we identified the enzymes of the central carbon metabolism that are physically interacting with proteins of the IFN system and observed that proteins at the interface between central carbon metabolism and IFN system are highly targeted by viruses. Among these proteins glycolytic enzymes appear as potential targets for viruses to modulate metabolism and counteract innate immune-response. We therefore performed an extensive screening of interaction between HCV proteins and cellular enzymes involved in glycolysis. It appeared that NS5-A of HCV strongly binds to the hexokinase (HK), the first rate-limiting enzyme of the glycolysis whose activity strongly influences the glycolytic rate of the cell and therefore its ability to support viral replication. In addition, GCKR is a regulatory protein of HK whose presence is necessary for cells to respond to type-I IFN. We therefore focused on the characterization of the interaction between HCV-NS5A and hexokinase 2, one of the four hexokinase isoenzymes. In vitro, HCV-NS5A directly interacts with HK2 and increases its enzymatic activity. To identify the residues involved in the interaction we compared the 1H-15N-HSQC spectra of isolated domain 2 of NS5A recorded in the absence and in the presence of unlabeled HK2. NMR chemical shift perturbations identified three regions of interaction on NS5A. The major one is similar to the main binding site for Cyclophilin A, an essential cell factor for HCV replication. The isolated NS5A-D2s from JFH1 or Con1 strains were sufficient to increase the HK2 activity in-vitro. Moreover, mutations in the putative binding site of NS5A-D2s impacted both its interaction with cyclophilin A and its consequence on HK2 activity. A peptide derived from the interacting sequence of wild-type NS5A was efficient to increase HK2 activity in-vitro in a dose-dependent manner. This work thus allowed the identification of a virus-derived peptide able to modulate one of the key host-metabolism enzymes. Work is in progress to investigate the functional impact of HK2 activity perturbation by NS5-A and its derived peptide on the quality of IFN response to different stimulators.

By employing an siRNA screen we aimed to identify interferon-stimulated genes (ISGs) that suppress HCV replication after IFN-alpha or -gamma treatment. The siRNA library was composed of about 100 genes that were identified by transcriptome analysis using HCV with and without replicon and being treated with and without interferon. The assay that we employed to identify these factors was a gain of

treated with and without interferon. The assay that we employed to identify these factors was a gain of function assay where HCV replication was measured after interferon-alpha or gamma treatment and down-regulation of a given candidate gene. After performing a primary screen we selected 15 genes and validated these hits in a secondary screen. By employing stringent hit criteria seven of these genes could be confirmed to rescue HCV replication after interferon-alpha and/or -gamma treatment (Metz et al., *Hepatology*, 2012).

A hallmark of HCV infection is its propensity to establish chronic infection with late sequelae like steatosis and hepatocellular carcinoma (HCC). We were able to identify a potential novel mechanism contributing to HCV pathogenesis. Due to persistent infection HCV induces stress responses and one of these responses in combination with type 1 interferon resulted in the dynamic oscillation of so called cytoplasmic stress granules (SG). These are dense cytosolic aggregates composed of cellular proteins and cellular RNAs. To study the process of SG oscillation in real time together with the group of Rohr (UHEI) a semi-automatic approach was developed for the analysis of SG dynamics at the single cell level (Ruggieri et al, *Cell Host Microbe* 2012). We found that cells with high SG dynamics are protected against apoptosis and it seems that phases of stress release are sufficient for cell survival (Ruggieri et al., *Cell Host Microbe* 2012).

In the last years great progress has been made in identifying principle steps of the HCV replication cycle. However, there is still limited knowledge on the detailed interplay of HCV with its host cell. By using a druggable siRNA library, we conducted a comprehensive whole-virus RNA interference-based screen and identified 40 host dependency and 16 host restriction factors involved in HCV entry/replication or assembly/release. Within these factors heterogeneous nuclear ribonucleoprotein K (HNRNPK) was identified as being essential for viral particle production without affecting viral RNA replication. We also could develop a comprehensive HNRNPK-virus interaction network including also interactions of the protein with different virus families. We determined the molecular mechanisms of HNRNPK interaction with HCV and found that the cellular factor interacts specifically with viral RNA and is recruited to sites in close proximity of lipid droplets where it colocalized with core protein as well as HCV plus-strand RNA. These results suggest that HNRNPK might determine efficiency of HCV particle production by limiting the availability of viral RNA for incorporation into virions (Poenisch et al., *PLoS Pathog*, 2015).

Publications resulting from SYSPATHO:

* B. Knapp, I. Rebhan, A. Kumar, P. Matula, N.A. Kiani, M. Binder, H. Erfle, K. Rohr, R. Eils, R. Bartenschlager, L. Kaderali (2011) *BMC Bioinformatics*, 12:485

* L. Meyniel-Schicklin*, B. de Chassey*, P. Andre, V. Lotteau. *Viruses and interactomes in translation* (2012) *Mol Cell Proteomics* 11(7): M111.014738

* B. de Chassey, L. Meyniel-Schicklin, A. Aublin-Gex, P. André and V. Lotteau (2012) *New horizons for antiviral drug discovery from virus-host protein interaction networks. Curr Opin Virol* 2(5):606-13.

* B. de Chassey, A. Aublin-Gex, A. Ruggieri, L. Meyniel-Schicklin, F. Pradezynski, N. Davoust, T. Chantier, L. Tafforeau, P.-E. Mangeot, C. Ciancia, L. Perrin-Cocon, R. Bartenschlager, P. André and V. Lotteau (2013) *The interactomes of influenza virus NS1 and NS2 proteins identify new host factors and provide insights for ADAR1 playing a supportive role in virus replication. PLoS Pathog* 9(7): e1003440

* B. de Chassey, L. Meyniel-Schicklin, A. Aublin-Gex, V. Navratil, T. Chantier, P. André and V. Lotteau (2013) *Structure homology and interaction redundancy for discovering virus-host protein interactions. EMBO rep*, 14(10):938-44

* Ruggieri, E. Dezert, P. Metz, S. Hofmann, J.P. Bergoest, J. Mazur, P. Bankhead, M.S. Hiet, S. Kallio, C.

Ruggieri, E. Dazert, P. Metz, S. Nollmann, J.F. Bergeest, J. Mazur, F. Barikneau, M.S. Niet, S. Nalls, G. Alvisi, C.E. Samuel, V. Lohmann, L. Kaderali, K. Rohr, M. Frese, G. Stoecklin, R. Bartenschlager (2012) Dynamic oscillation of translation and stress granule formation mark the cellular response to virus infection. *Cell Host Microbe*. 12(1):71-85.

* P. Metz, E. Dazert, A. Ruggieri, J. Mazur, L. Kaderali, A. Kaul, U. Zeuge, M.P. Windisch, M. Trippler, V. Lohmann, M. Binder, M. Frese, R. Bartenschlager (2012) Identification of type I and type II interferon-induced effectors controlling hepatitis C virus replication. *Hepatology* 56(6):2082-93.

* Binder M, Sulaimanov N, Clausznitzer D, Schulze M, Hüber CM, Lenz SM, Schlöder JP, Trippler M, Bartenschlager R, Lohmann V, Kaderali L.(2013). Replication vesicles are load- and choke-points in the hepatitis C virus lifecycle. *PLoS Pathog*. 2013;9(8):e1003561.

* Poenisch M, Metz P, Blankenburg H, Ruggieri A, Lee JY, Rupp D, Rebhan I, Diederich K, Kaderali L, Domingues FS, Albrecht M, Lohmann V, Erfle H, Bartenschlager R. (2015) Identification of HNRNPK as regulator of hepatitis C virus particle production. *PLoS Pathog*. 11(1):e1004573.

** Modeling HCV Replication and Interactions with the Host **

As an obligate intracellular parasite, HCV replication heavily depends on the host cell. Within SYSPATHO, we established detailed mechanistic models (Binder et al., 2013; Ivanisenko et al., 2013, 2014), which adequately describe highly dynamic initial events after infection of a cell and under drug pressure. In particular, our models integrated host factors involved in the virus lifecycle and in particular also considered antiviral signaling in the cell. Using our model, we were able to simultaneously describe the dynamic initial phase of establishment of viral replication, as well as the steady state behavior. Comparing two Huh cell lines, we found that a host factor involved in the formation of a membranous replication compartment in the cell is indicated as a factor for cell permissiveness to viral replication. Furthermore, we used model analysis to predict sensitive drug targets, i.e. processes in the replication cycle which, if targeted by drugs could inhibit viral replication most.

Molecular mechanisms of the HCV drug resistance acquisition are still unclear. One may assume that the largest contribution to resistance acquisition is made by the preexisting mutant HCV quasispecies, which have the selection advantages in the drug presence. In particular, it was previously demonstrated that the resistance observed during in vitro replicon studies results from the outgrowth of the preexisting NS5B mutations rather than from the generation of mutations after an onset of antiviral suppression. Our modeling results showed, that in the whole pool of mutant RNAs the preexisting NS3 protease mutants prevail over mutants generated during the wild type viral RNA replication at high drug concentration. However, at low inhibitor concentrations an impact of the mutants newly emerged during the medical treatment into the whole pool of resistant mutants is comparable with that of the preexisting mutants. We also use our model to study the efficacy of potential inhibitors targeted to a host cellular factor involved in the formation of the active viral replicase. In principle these inhibitors can either stop the integral cellular factor expression or reduce its activity in the viral replication complex. Our analysis revealed that the NS3 protease inhibitors are less effective in reduction of the viral RNA amount when combined with the potential inhibitor of the cellular factor expression than with the inhibitor of the cellular factor activity. Since the replicon system is the major model for design and evaluation of anti-HCV drugs, our results offer guidelines for selection of clinical drug concentrations and combinations of antivirals.

Furthermore, we integrated a description of an intracellular antiviral immune response in our model as a

Furthermore, we integrated a description of an intracellular antiviral immune response in our model as a crucial host process to modulate infection and shape infection outcome. We included various interactions between viral replication and immune response, capturing the viral repression by the immune response, as well as viral evasion and persistence at the cellular level. The probability of the immune response to be activated depends on the one hand on the level of viral RNA, and on the other hand on viral negative feedback. The immune response affects various replication processes of the virus, namely, degradation, translation and import in a protective replication compartment. The viral feedback affects the activation rate of the immune response. Fitting this model to additional time course data for co- or pretreated we could estimate a number of parameters for the antiviral response.

By studying populations of cells, we significantly contributed to approaches for multi-scale modeling by coupling different layers of biological organization relevant for viral infection. We developed a stochastic model for HCV spread and the dynamics of the immune response in a two-dimensional cell population. Here, production of the antiviral cytokine interferon represents a measure for the immune response, which can be stimulated by viral infection of the cell or IFN molecules secreted by neighboring cells. The dynamics of the cell population depends on a number of factors. For instance, cells can be pre-treated with IFN to initiate IFN production in a proportion of cells. Transitions between cell states depend on each cell's own state as well as the number of viral particles and IFN molecules from neighboring cells at the location of the cell. Hence, this model is flexible to simulate *in silico* different experimental conditions and predict population outcomes.

At the subcellular level, we analyzed individual motion trajectories of HCV particles quantified from microscopy images and individual trajectories determined by the tracking approach developed within SYSPATHO using microscopy images from the consortium (Godinez et al, 2009; Chenouard et al., 2014; see below). A motion model was inferred for each of them. To discriminate between different motion types we applied the mean-square displacement (MSD) analysis (Monnier et al., 2012), as well as the analysis of diffusion based on the equation of the Jeffreys type (Rukolaine & Samsonov, 2013). We showed that the motion of replication complexes from the first dataset is well explained by diffusion based on the equation of the Jeffreys type, while the motion of particles for the second dataset is well described as directed motion with diffusion. We showed that distributions of segment lengths of viral trajectories in different experiments are qualitatively consistent with distribution of microtubule lengths. We demonstrate that the population dynamics of viral particles can be reconstructed from data on both HCV lifetime and trajectories initiation.

Publications resulting from SYSPATHO:

* Marco Binder, Nurgazy Sulaimanov, Diana Clausznitzer, Manuel Schulze, Christian M. Hüber, Simon M. Lenz, Johannes P. Schlöder, Martin Trippler, Ralf Bartenschlager, Volker Lohmann, Lars Kaderali (2013) Replication Vesicles are Load- and Choke-Points in the Hepatitis C Virus Lifecycle. *PLoS Pathogens* 8(9): e1003561

* N. V. Ivanisenko , E. L. Mishchenko , I. R. Akberdin , P. S. Demenkov , V. A. Likhoshvai , K. N. Kozlov , D. I. Todorov , M. G. Samsonova , A. M. Samsonov , N. A. Kolchanov , V. A. Ivanisenko (2013) Replication of the subgenomic hepatitis C virus replicon in the presence of the NS3 protease inhibitors: a stochastic model. *Biophysics* (58)5: 592-606

* Nikita V. Ivanisenko , Elena L. Mishchenko , Ilya R. Akberdin , Pavel S. Demenkov , Vitaly A. Likhoshvai , Konstantin N. Kozlov , Dmitry I. Todorov , Vitaly V. Gurev , Maria G. Samsonova , Alexander M.

Konstantin N. Rozlov , Dmitry I. Rodionov , Vitaly V. Gurovsky , Maria G. Samsonova , Alexander M.

Samsonov , Diana Clausznitzer , Lars Kaderali , Nikolay A. Kolchanov , Vladimir A. Ivanisenko (2014) A New Stochastic Model for Subgenomic Hepatitis C Virus Replication Considers Drug Resistant Mutants . PLoS One 9(3): e91502

* Amberkar S, Kiani NA, Bartenschlager R, Alvisi G, Kaderali L (2013) High-throughput RNA interference screens integrative analysis: Towards a comprehensive understanding of the virus-host interplay. World J Virol 2(2):18-31

* S.A. Rukolaine, A.M. Samsonov. Inner particle immobility in a dense crowd: mass transfer, described by the equation of the Jeffreystype. Phys. Rev E., 2013

** Network Reconstruction from High-Throughput Data **

The problem of reconstructing signal transduction and regulatory pathways directly from data using machine learning has been studied extensively by many researchers in the last decade, for a recent review see Kaderali & Radde (2008). Methods proposed include Boolean models, correlation based methods and Relevance Networks approaches, Bayesian Networks, S-Systems, Differential Equation Models, and

others. Most approaches are based on microarray gene expression data, and concentrate on the reconstruction of genetic regulatory networks from either static or time series data. Some methods such as Bayesian networks allow the integration of biological prior information. In spite of these efforts, the problem is still a very challenging one, as evidenced by the recent DREAM challenge (Stolovitzky et al, 2009).

Fewer approaches exist for the inference of signal transduction networks from RNAi perturbation data, a relatively young research field enabled through technical advances in experimental highthroughput screening assays (Moffat and Sabatini, 2006). RNAi is a very powerful tool to screen for genes involved in cellular functions of interest, however, the spatial and temporal placement of proteins within pathways based on RNAi screens, and even more so the reconstruction of unknown pathways from RNAi data, are very difficult problems. Markowitz et al. (2007) recently proposed Nested Effect Models (NEMs) for this purpose. These models reconstruct signal transduction pathways from the nested structure of phenotype observations. However, while well-suited for cause-effect data such as RNAi data, NEMs require high-dimensional readouts such as microarray experiments for each knockout to be able to reconstruct the underlying pathway. We previously developed an alternative approach that computes probability distributions over alternative network topologies and model parameters consistent with the experimental data, based on probabilistic Boolean threshold functions. This approach allows the reconstruction of networks from RNAi data with low-dimensional readouts. However, due to the nonlinear model used, and high computational demands, this method is applicable only to small networks. Furthermore, neither method is able to handle time series data from live cell imaging experiments.

Within SYSPATHO we developed novel methods for network reconstruction from perturbation data, which due to the linear programming formulation used and by integrating different data sources allow for the derivation of larger network involving several dozen components from RNAi perturbation data. Specifically, linear programming approaches have been developed, which can be solved efficiently, even for large-scale problems. First, one method is based on heuristics and integer linear optimization models which make use of previous knowledge of protein-protein interactions. The method reconstructs the signaling network from the given protein protein interaction (PPI) network satisfying RNAi data by making minimum changes on the given network. The methods SiNeC (Signaling Network Construction) and S-SiNeC

changes on the given network. The methods SINEC (Signaling Network Construction) and S-SINEC (Scalable Signaling Network Construction) by Hashemikhabir et al. (2012) are heuristic graph algorithms which provide near optimal results and can scale well for reconstructing networks up to hundreds of components. We validated the proposed methods on synthetic and real data sets. Secondly, a second method scales up the linear programming model by using a divide-and-conquer approach. The signaling network is divided into smaller components by different heuristics and solutions for small components are merged to get the final signaling network. The method is tested again on real and synthetic data and the results showed that the proposed method is better in terms of accuracy and speed over existing techniques. A third method developed is based on the idea of an information flow going through the network, which is modulated by perturbations of individual nodes and their propagating down-stream effects, such as in RNAi experiments (Knapp & Kaderali, 2013). A constrained linear optimization algorithm, which takes into account real-valued activating or repressing effects of parent nodes on child nodes, was developed. This method was applied to small to medium sized networks using simulated data, as well as on an experimental data set.

Furthermore, we developed novel efficient clustering approaches which includes biological knowledge into the network reconstruction process. Based on Yeast-2-Hybrid data as well as including annotation data from databases such as gene ontology, we have analyzed the structure of the underlying protein-protein interaction graphs. Currently known clustering algorithms often do not perform well under such conditions and small clusters are often neglected. The algorithm we developed uses this knowledge on the structure of protein-protein interaction networks by reducing the search space to nodes that attain similar or close to similar degrees. Most known graph clustering algorithms use a distance measure based on the number of hops in the path between two nodes in the graph. This distance measure is not appropriate in the case of graphs representing biological interactions since these graphs contain a significant rate of false positive and false negative edges. Moreover, the computation of a dissimilarity measure based on the hop distance in a graph can be time and memory consuming in our case of large and complex networks. We suggested a new approach to finding clique-like clusters in PPI networks. Our algorithm, being very fast, simple, and robust, allows finding clusters of different size. The underlying approach consists in reducing the problem to looking for clusters only among the nodes of similar degrees, along with applying the Farthest-Point-First clustering algorithm of Gonzalez with the Jaccard distance function. The optimal number of clusters is determined by using the elbow criterion.

We developed a method to detect the significance of the links (interactions) between the nodes in complex biological networks by using a Gaussian graphical model (GGM). Furthermore we apply the threshold gradient descent algorithm (TGD) and cross-validation methods in junction with GGM as an alternative of the classical GGM based on the inference for biochemical systems. We also found that applying normalization methods can improve results of GGM further.

To integrate various data sets for HCV, we developed a modular enrichment method for HCV-related genes on host protein-protein interaction (PPI) networks built from public repositories (Amberkar & Kaderali, 2015). Using a robust Markov Chain Length graph clustering algorithm, we identify functional modules in the network and map RNAi hits and the full set of screened genes on these functional modules. This approach allows us to identify modules showing significant overlap of RNAi hits.

The integration of mathematical models describing the interoperable gene networks for different biological

The integration of mathematical models describing the interoperable gene networks for different biological processes is very important task in gene network theory. Modeling of interactions between biological processes requires knowledge about shared genes. Mistakes in the integration of gene networks due to incomplete information on the interaction between genes lead to the creation of false integrated mathematical model. A new method for integration of mathematical models describing dynamic properties of interconnected gene networks in absence of data about shared genes is proposed. The method is based on the Control Theory. Construction of the control function was performed using RNAi screening data from the consortium. The criteria of applicability of the method and the requirements for the experimental data on RNAi knockouts, depending on the structural and functional organization (cycles in the regulatory circuits, Betweenness, Centrality, etc) of the analyzed gene networks were suggested. Computer algorithms for determining of steady states and cycles in gene networks were set up and tested with simulated and experimental data.

Generally, by permitting the integration of different data sources (literature, yeast-2-hybrid, live cell imaging data, etc.), our methods allow more reliable predictions, and general software tools for this purpose are available or being developed and made available to the scientific community within the next year.

Publications resulting from SYSPATHO:

- * M. Böck, S. Ogishima, H. Tanaka, S. Kramer, L. Kaderali (2012) Hub-Centered Gene Network Reconstruction using Automatic Relevance Determination. PLoS ONE 7(5): e35077
- * O. Eren Ozsoy and T. Can (2013) Divide and Conquer Approach for Construction of Large-Scale Signaling Networks from PPI and RNAi Data Using Linear Programming. IEEE/ACM Transactions on Computational Biology and Bioinformatics. Institute of Electrical and Electronics Engineers Inc. 1 – 1
- * V. Purutçuoglu, T. Erdem, G.W. Weber, Inference of the JAK-STAT gene network via graphical models. Proceeding of the 23rd International Conference on Systems Research, Informatics and Cybernetics, Baden, Germany, 46-50, 2011.
- * V. Purutçuoglu. Stochastic modelling and parameter estimation of the HCV network. Proceeding of the 16th INFORMS Applied Probability Conference. 6-8 June 2011, Stockholm, Sweden, pages: 81-82.
- * Hashemikhabir S, Ayaz ES, Kavurucu Y, Can T, Kahveci T (2012) Large-Scale Signaling Network Reconstruction. IEEE/ACM Transactions on Computational Biology and Bioinformatics. Institute of Electrical and Electronics Engineers Inc. 1696-1708
- * Narsis A Kiani , Lars Kaderali (2014) Dynamic probabilistic threshold networks to infer signaling pathways from time-course perturbation data. BMC Bioinformatics 15(1): 250
- * Amberkar SS, Kaderali L (2015) An integrative approach for a network based meta-analysis of viral RNAi screens. Algorithms Mol Biol 10:6. eCollection 2015.
- * Ö. Defterli, V. Purutçuoglu, G.W. Weber, Advanced mathematical and statistical tools in the dynamic modeling and simulation of gene-environment regulatory networks.preprint: 1-22. Chapter in Modelling, Optimization and BioEconomy - MOBE, Eds. Alberto Pinto and David Zilberman, Springer. 2013.
- * E. Ayyildiz, Gaussian graphical approaches in estimation of biological systems, MSc Thesis, Department of Statistics, Middle East Technical University, May 2013.
- * V. Purutçuoglu, E. Ayyildiz, E. Wit, Inference of the complex system via Gaussian graphical models (submitted).
- * E. Ayyildiz, V. Purutçuoglu. Inference of the biological systems via L1-penalized lasso regression. Proceeding of the 29th European Meeting of Statisticians, Budapest, Hungary, 2013, page: 38

- * O. Eren Ozsoy and T. Can. A divide and conquer approach for construction of large-scale signaling networks from PPI and RNAi data using linear programming. IEEE/ACM Transactions on Computational Biology and Bioinformatics, in press, July 2013. doi:10.1109/TCBB.2013.80
- * B. H. Akman, T. Can, and A. E. Erson-Bensan, Estrogen-induced upregulation and 3'-UTR shortening of CDC6. Nucl. Acids Res., September 12, 2012. doi:10.1093/nar/gks855
- * Mehmet Emin Gönen, Counting and listing a special class of directed graphs, Master of Science Thesis, Bahcesehir University, The Graduate School of Natural and Applied Sciences, Applied Mathematics, Istanbul, Turkey, August 2013.
- * B. Knapp, L. Kaderali (2013). Reconstruction of Cellular Signal Transduction Networks using Perturbation Assays and Linear Programming. PLoS ONE 8(7): e69220.

Software tools resulting from SYSPATHO:

- * Bioconductor package "lpNet" for the reconstruction of cellular signal transduction networks using perturbation assays and linear programming. B. Knapp, J. Mazur and L. Kaderali (2013).
- * A software package implementing newly developed methods for network inference will be released within the next year, making the tools developed available to the scientific community.

** High-Throughput Data Analysis and Image Processing **

Establishing models of the HCV life cycle required automated analysis of the acquired microscopy image data. To this end image analysis methods were developed for quantifying model parameters and for extracting important spatial-temporal biological information. In particular, for analyzing large scale high-throughput images from RNAi screens fully automatic, accurate, and efficient methods are required (e.g. Gudla et al. 2008, Harder et al. 2008, Li et al. 2007, Paran et al. 2007, Carpenter et al. 2006, Lindblad et al. 2004, Perlman et al. 2004). For a detailed quantification of the image data it was crucial to exploit the information on a single cell level in comparison to using average values of cell populations. Central tasks in high-throughput applications are the segmentation of cell nuclei, the segmentation of the cytoplasm, the quantification of protein markers, the extraction of cellular image features, and the classification of phenotypes. Although a number of approaches for cell segmentation and classification had been proposed in the literature, the analysis of different cell types under different perturbations (e.g. RNAi knockdown) in our particular application typically required extensions and adaptations of these approaches, respectively the development of new approaches. An additional challenge for image analysis at a single cell /particle level was the clustering of objects. Moreover, it was important to exploit the temporal information in microscopy image sequences based on tracking approaches (e.g. Padfield et al. 2009, Harder et al. 2009, Wang et al. 2007, Chen et al. 2006).

In this project, we developed and extended a computational approach for cell nuclei segmentation based on level set deformable models. The approach combines different convex energy functionals and enables coping with inhomogeneities of the image intensities as well as with clustered cells. Based on the 2D approach an extension for 3D image data was developed. The approach was applied to 2D and 3D cell microscopy image data and thoroughly tested. The performance was quantified based on manual annotation and compared with previous approaches. It turned out the approach yields superior results compared to previous approaches (Bergeest & Rohr, Medical Image Analysis 2012, ISBI 2014). In addition, extracted parameters of an approach for quantifying the dynamic host stress response induced

addition, extracted parameters of an approach for quantifying the dynamic host stress response induced by HCV from time-resolved microscopy images (Ruggieri et al., *Cell Host & Microbe*, 2012) were used for mathematical modeling. With this image analysis approach different compartments are distinguished (e.g. cell nuclei, stress granules).

In addition, we quantified HCV model parameters within different subcellular compartments to yield a more detailed and comprehensive characterization of the imaged phenotypes as compared to earlier work. We developed a computational tracking approach for spatial-temporal analysis of HCV replication dynamics using live cell image data (Chenouard et al., *Nature Methods*, 2014). Challenges of the image data were a high density and heterogeneous size of the image structures, features not uncommon for general imaging of biological molecules. Our approach is based on a Bayesian paradigm and spatial-temporal filtering. The quantified parameters of HCV replicons were further used for mathematical modelling and motion classification using a new transportation model of Jeffrey's type. Data analysis and image processing are core technologies bridging the experimental work with the theoretical and modelling work packages in SYSPATHO.

Publications resulting from SYSPATHO:

- * Ruggieri A, Dazert E, Metz P, Hofmann S, Bergeest JP, Mazur J, Bankhead P, Hiet MS, Kallis S, Alvisi G, Samuel CE, Lohmann V, Kaderali L, Rohr K, Frese M, Stoecklin G, Bartenschlager R (2012) Dynamic oscillation of translation and stress granule formation mark the cellular response to virus infection. *Cell Host Microbe* 12(1): 71-85
- * S. Reiss, I. Rebhan, P. Backes, I. Romero-Brey, H. Erfle, P. Matula, L. Kaderali, M. Pönisch, H. Blankenburg, M.-S. Hiet, T. Longerich, S. Diehl, F. Ramirez, T. Balla, K. Rohr, A. Kaul, S. Bühler, R. Pepperkok, T. Lengauer, M. Albrecht, R. Eils, P. Schirmacher, V. Lohmann, R. Bartenschlager (2011) Recruitment and activation of a lipid kinase by NS5A of the hepatitis C virus is essential for integrity of the membranous replication compartment. *Cell Host Microbe* 9(1): 32-45
- * Bettina Knapp, Ilka Rebhan, Anil Kumar, Petr Matula, Narsis A Kiani, Marco Binder, Holger Erfle, Karl Rohr, Roland Eils, Ralf Bartenschlager, Lars Kaderali (2011) Normalizing for individual cell population context in the analysis of high-content cellular screens. *BMC Bioinformatics* 12: 485ff
- * J.P. Bergeest, K. Rohr (2012) Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals. *Medical Image Analysis* 16 (7), 1436-1444
- * J.P. Bergeest, K. Rohr. Segmentation of Cell Nuclei in 3D Microscopy Images Based on Level Set Deformable Models and Convex Minimization. Proc. IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'14), Beijing, China, 28 April - 2 May, 2014
- * N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I.F.Sbalzarini Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A.R. Cohen, W.J.Godinez K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K.E.G. Magnusson, J. Jaldén, H.M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S.L. Shorte, J. Willemsse, K. Celler, G.P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. Ortiz de Solórzano, J.-C. Olivo-Marin, E. Meijering (2014) Objective comparison of particle tracking methods. *Nature Methods* 11(3): 281-289.

Software packages resulting from SYSPATHO:

- * R script for Normalizing for Individual Cell Population Context in the Analysis of High-Content Cellular Screens. B. Knapp, I. Rebhan, A. Kumar, P. Matula, N.A. Kiani, M.Binder H. Erfle, K. Rohr, R. Eils, R. Bartenschlager, L. Kaderali (2011) *BMC Bioinformatics* 12:485

** Text Mining**

The biomedical literature is a huge repository of poorly structured data in systems biology. Extraction of facts concerning molecular-genetic interactions from literature and providing access to them can be divided on two directions: manual analysis of literature and automated information extraction with text mining techniques. Manual curation is more accurate but time-consuming. Increasing growth of the literature demands the usage of the automated text mining techniques for molecular-genetic networks reconstruction on proteome and genome levels (Winnenburg et al, 2008). A number of computer tools for extracting information about molecular genetic interactions had been developed (see review of Krallinger et al, 2008). One of widely used approaches implies calculation of statistically meaningful values of object name co-occurrence in the texts. A well known computer tool based on this approach is PUBGENE (Jenssen et al. 2001). It allows the user to find co-occurrences of biological objects in PubMed abstracts. This method shows high recall, but quite low precision. MedScan from the PathwayStudio (Nikitin et al., 2003) is an example of a full parsing-based system. Text analysis algorithm based on formal grammar implemented in this system shows high accuracy but is very time-consuming.

In the frame of SYSPATHO methods that are both fast and accurate were developed and used for reconstruction of large-scale HCV-Host molecular-genetic networks on the basis of information from PubMed abstracts. The Hepatitis C Associome (HCA) relational database containing integrated literature network with Y2H data describing virus-host interactions was developed (<http://www-bionet.sccc.ru/psd/andhcv/>). The HCA database contains virus-host, host-host and virusvirus molecular-genetic interactions including:

- 4 309 629 interactions of 24 types (expression, physical interactions and others);
- 221 430 PubMed protein-protein interactions;
- 37 319 Databases protein-protein interactions (including yeast two-hybrid system data);
- 859 human proteins associated with 11 HCV proteins;
- 1 815 1st-level interactions;
- 11 167 proteins involved in 164 606 2nd-level interactions.

We showed that proteins associated with HCV are more connected to each other than random proteins, indicating that HCV proteins regulate host biological pathways. The pathways in which host proteins associated with HCV are localized mostly in single GO cell components. Fig. 3.1 shows an example of network reconstruction of interactions between hepatitis C proteins and human biological processes using the HCA database. The main methodical advantages of the used approaches are original, and corresponding RF patents are held by PBsoft. A key achievement of SYSPATHO is the development and application of proprietary text mining software of PBSoft to mine the scientific literature for HCV-host interactions.

Publications resulting from SYSPATHO:

* Sommer B, Tiys ES, Kormeier B, Hippe K, Janowski SJ, Ivanisenko TV, Bragin AO, Arrigo P, Demenkov PS, Kochetov AV, Ivanisenko VA, Kolchanov NA, Hofestädt R (2010) Visualization and analysis of a cardio vascular disease- and MUPP1-related biological network combining text mining and data warehouse approaches. J Integr Bioinform 7(1): 148

* Turonne N, Tiys E, Ivanisenko V, Yudin N, Ignatieva E, Valour D, Desroches SA, Hue I (2012) Finding

Turekova N, Hys E, Ivanisenko V, Tuzhik N, Ignatieva E, Valouf D, Degreffe SA, Hue T (2012) Finding biomarkers in non-model species: literature mining of transcription factors involved in bovine embryo development. *BioData Mining* 1(5): 12

* Demenkov P.S. Ivanisenko T.V. Kolchanov N.A. Ivanisenko V.A. (2011) ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol* 11(3):149-61

* Sommer B, Kormeier B, Demenkov PS, Arrigo P, Hippe K, Ates Ö, Kochetov AV, Ivanisenko VA, Kolchanov NA, Hofestädt R (2013) Subcellular localization charts: a new visual methodology for the semi-automatic localization of protein-related data sets. *Journal of Bioinformatics and Computational Biology* 11(1): 1340005

Proprietary tools resulting from SYSPATHO:

* relational database containing integrated information both from literature and experiments on virus-host interactions (PBSoft)

** Computational Drug Design **

The traditional approaches for testing small molecules inhibitors in vivo to animals are costly, time consuming, and have a low throughput. Accurate prediction of the biological action of chemical substances on living systems, identification of possible toxic alerts, and compound prioritization for animal testing are the primary goals of computational drug design. In the literature although individual successes have been reported, in general a strong need remains in developing widely accessible and reliable computational drug design modeling techniques and specific end-point predictors (Oprea TI et. al. 2007).

Through the combination of industry-recognized expertise, state of the art software, large-scale computing infrastructure, and advanced in silico capabilities in molecular design and simulation, SYSPATHO provided effective paths to drug innovation. We developed a robust and reliable QSAR (Quantitative Structure – Activity Relationship) models for the prediction of HCV inhibitory activity of small molecules. To this end, we developed a predictive workflow including the following main steps: (i) input data, (ii) calculate descriptors, (iii) select descriptors, (iv) develop model, (v) validate model, (vi) define domain of applicability. We have used KNIME platform to integrate these steps and developed our in house nodes called "Enalos KNIME nodes" (<http://www.novamechanics.com/knime.php>). "Enalos KNIME nodes" by Novamechanics were developed to perform the following tasks: (a) Define the Applicability Domain (APD) based on the Euclidean distances, (b) Define the Applicability Domain based on the Leverages, (c) Calculate Molecular Descriptors using Mold2 software and (d) Calculate Quality of Fit and Predictive Ability of a continuous QSAR Model. Based on this workflow and a large publicly available database of known HCV inhibitors, we developed highly predictive and robust QSAR models that can be used to predict the HCV inhibition. We developed several models by applying different combinations of calculated descriptors and modeling techniques and for the given dataset we have concluded in a reliable predictive classification model.

Furthermore, we performed a ligand virtual screening study. Based on the models from our QSAR model a virtual screening procedure was initiated for the identification of promising small molecules as possible HCV inhibitors. We successfully developed a computational Virtual Screening workflow combining Molecular Docking, 3D CoMFA/CoMSIA QSAR and Similarity Search that was applied to PubChem and

molecular docking, 3D-QSAR/COMSIA-QSAR and similarity search that was applied to PubChem and ChEMBL database with the aim to identify small molecules that could act as inhibitors of HCV replication. In the proposed approach we have first developed a robust, validated and predictive 3D-QSAR model that was subsequently used to virtually screen compounds from PubChem and ChEMBL database. We narrowed the chemical space search by focusing only on compounds that contain specific scaffold. Similarity search was then used to retrieve the compounds that are similar to known active inhibitors. In this way, we identified the most promising compounds from a pool of new analogues and to prioritize a list of compounds for screening. The approach revealed several promising chemistry driven compounds with potential high activity. The workflow can also be used to screen other databases or virtual combinations to identify derivatives with desired activity. The results have been published in two scientific papers (Vrontaki et al., 2015; Vrontaki et al., in press).

Publications resulting from SYSPATHO:

* G. Melagraki & A. Afantitis (2011) Ligand and structure based virtual screening strategies for hit-finding and optimization of hepatitis C (HCV) inhibitors. *Curr. Med Chem* 18(17): 2612-2619

* Eleni Vrontaki, Georgia Melagraki, Thomas Mavromoustakos, Antreas Afantitis (2015) Exploiting

ChEMBL database to identify indole analogs as HCV replication inhibitors. *Methods* 71:4-13

* E. Vrontaki, G. Melagraki*, T. Mavromoustakos, A. Afantitis. Searching for Anthranilic ?cid=?ased Thumb Pocket 2 HCV NS5B Polymerase Inhibitors through a Combination of Molecular Docking, 3D-QSAR and Virtual Screening, in press.

** Solving nonlinear ODEs and PDEs **

It is evident that linear mathematical models for the dynamics of biological systems, while being the most simple, are typically inconsistent with the highly non-linear behavior of the underlying biological system. Moreover, in most models, any spatial inhomogeneities of species distribution are not considered, and the models consist of linear ordinary equations describing the underlying kinetics. Any spatial processes, such as diffusion, dissipation or localization in specific cellular compartments are usually not taken into account. There is practically no general method to solve highly non-linear continual and discrete equations, and neither theorem to prove the existence of solutions. For this reason, any new method for the solution of non-linear equations is of great interest for mathematicians and for tentative applications in systems biology. The theory of standard reaction-diffusion models is well developed, however, not for coupled equations and not for complex nonlinearities.

Irreversibility is one of the most specific features of biological processes, including virus infection and replication. The irreversible thermodynamics (IT) introduced by Onsager and Prigogine in the 70-ies is nowadays the classical one, and it was generalized later as an extended one (EIT) by Jou, Casas-Vazquez and Lebon (Jou et al., 1988; Casas-Vázquez & Jou, 1994; Jou et al, 2006; Lebon et al., 2008) to govern various processes, running far from their local stability. Its main idea is in the dependence of the entropy of a biological system on the fluxes as independent variables (Lebon et al., 2008; Sobolev, 1994; Uchaikin & Saenko, 2001). These physical theories provided several new mathematical models to biological dynamical processes, which were reconsidered within SYSPATHO.

The methods of mathematical physics to solve the corresponding nonlinear partial differential equation (p.d.e.) are scanty and mostly are based on reduction of initial problem to the p.d.e. of lower order. In

(p.d.e.) are scanty and mostly are based on reduction of initial problem to the p.d.e. of lower order. In addition, the analytical methods of ordinary differential equations (o.d.e.) theory can be used to study the system behavior in a phase space. An independent analysis of the model based on discretization requires, as a rule, knowledge of a partial analytical solution as a check point for algorithms and numerical simulation methods used. Consequently, it was used in SYSPATHO for the necessary tuning and refinement of the Optimal Steepest Descent Algorithm (OSDA), which runs the cost functional to a minimal value. Improvements were based on formal mathematical analysis of the problem in a new formulation.

Publications resulting from SYSPATHO:

* S. A. Rukolaine and A. M. Samsonov (2011) Diffusion vs telegraph equation: what is the better approximation of the delayed uncoupled continuous-time random walks? Problems in mathematical physics and applied mathematics. The Ioffe Institute St. Petersburg. 114-137

* S.A. Rukolaine, A.M. Samsonov (2012) The delayed uncoupled continuous-time random walks do not provide a model for the telegraph equation. Phys Rev E 85: 021150

** Inverse Problem of Mathematical Modeling **

The design of efficient algorithms and systems to solve the inverse problem of mathematical modeling continues to be a challenge due to large volume and heterogeneity of biomedical data, as well as high computational complexity of biomedical applications. It is well established that an efficient optimization method should not only be fast and scalable across modern high performance architectures, but also reliable and robust. In systems biology the most commonly used global optimization algorithm is parallel Simulated Annealing (SA) (Chu et al., 1999). This method uses the considerable quantities of CPU time, but is capable to find the global extremum and runs efficiently in parallel. The wide range of methods called Genetic Algorithms (GA) has been developed and successfully applied to biological problems (Spirov & Kazansky, 2002). Modern Evolutionary algorithms such as Evolution Strategies (ESs) or Differential Evolution (DE) can outperform other methods at the estimation of parameters of some biological models (Fomekong Nanfack et al. 2007, Kozlov & Samsonov, 2009). The common challenge in the efficient implementation of global optimization methods is that they depend on problem-specific assumptions and thus are not easily adapted to other problems. For example, in SA the final result and computational time depend on the so called cooling schedule, the success of the GA optimization is closely connected with the selected mutation, recombination and selection rules and the evolutionary algorithms heavily rely on the algorithmic parameters which define the model of evolution.

In the past decade numerous software packages, which implemented algorithms to solve the inverse problem of mathematical modeling, had been developed. These systems were usually designed with a certain hardware architecture in mind. This is especially true for the time consuming applications typical for systems biology. SYSPATHO contributed significant advancements to the field by improving existing optimization algorithms. We improved the Differential Evolution Entirely Parallel (DEEP) method developed by us earlier and developed the prototype of Combined Optimization Technique by implementing a new selection rule for Differential Evolution that allows us to use several different objective functions in offspring evaluation. We compared the convergence of the original and improved methods by calculating the number of serial iterations $Q(F)$ necessary to attain the particular value of the quality functional F in a parallel implementation. Furthermore, we developed an adaptive Combined Optimization Technique (COT): DEEP obtains the rough approximation of the parameter set that is refined by Optimal Steepest

(OSDA), DEEP obtains the rough approximation of the parameter set that is refined by Optimal Steepest Descent Algorithm (OSDA) (Kozlov&Samsonov, 2003). OSDA is a local search method based on optimal control theory. OSDA uses the variation of the expanded functional that combines all constraints. The necessary condition of the minimum of the first order is used to derive the numerical algorithm. OSDA computes all derivatives using their analytical representation. This reduces the computational cost for each iteration significantly.

Publications resulting from SYSPATHO:

* Kozlov K, Surkova S, Myasnikova E, Reinitz J, Samsonova M (2012) Modeling of Gap Gene Expression in Drosophila Kruppel Mutants. PLoS Comput Biol 8(8): e1002635

* N. V. Ivanisenko , E. L. Mishchenko , I. R. Akberdin , P. S. Demenkov , V. A. Likhoshvai , K. N. Kozlov , D. I. Todorov , M. G. Samsonova , A. M. Samsonov , N. A. Kolchanov , V. A. Ivanisenko (2013) Replication of the subgenomic hepatitis C virus replicon in the presence of the NS3 protease inhibitors: a stochastic model. Biophysics (58)5: 592-606

* Nikita V. Ivanisenko , Elena L. Mishchenko , Ilya R. Akberdin , Pavel S. Demenkov , Vitaly A. Likhoshvai , Konstantin N. Kozlov , Dmitry I. Todorov , Vitaly V. Gursky , Maria G. Samsonova , Alexander M.

Samsonov , Diana Clausznitzer , Lars Kaderali , Nikolay A. Kolchanov , Vladimir A. Ivanisenko (2014) A New Stochastic Model for Subgenomic Hepatitis C Virus Replication Considers Drug Resistant Mutants . PLoS One 9(3): e91502

Software packages resulting from SYSPATHO:

* We have developed a package for finding any real and integer unknown parameters for data-driven models of biological processes using one or even several objective functions called DEEP. DEEP is an open source and free software distributed under the terms of GPL licence version 3. The sources are available at <http://deepmethod.sourceforge.net/> and binary packages for Fedora GNU/Linux are provided for RPM package manager at <https://build.opensuse.org/project/repositories/home:mackoel:compbio>. The package was already used in a number of applications: In collaboration with CHLA (Prof. T. Tatarinova) K. Kozlov et al, BMC Genomics 16 (2015), p. S9; K. Kozlov, V. Gursky, I. Kulakovskiy, and M. Samsonova, BMC Genomics 15 (2014), p. S6.; in collaboration with ICG RAS (Prof. V.A. Ivanisenko) N.V. Ivanisenko et al, PLoS ONE 9 (2014), p. e91502; in collaboration with ICG RAS (Prof. N. Kolchanov) M. Nuriddin et al, Russian Journal of Genetics: Applied Research 17 (2013), pp. 686–704. The software is useful especially in the area of systems biology and bioinformatics.

** Identifiability of Model Parameters **

The number of model parameters that are estimated by fitting to experimental data is typically large. For the analysis of the estimation results it is necessary to know how reliable the obtained estimates are. For example, in practice, insufficient or noisy data, as well as strong parameter correlation or even their functional relation, may prevent the unambiguous determination of parameter values. In addition, many models used in systems biology exhibit parameter “sloppiness” (Gutenkunst et al., 2007). This means that there may exist model parameters, estimations of which can vary by orders of magnitude without significantly influencing the quality of the fit. The detection of non-identifiable and sloppy parameters is the subject of identifiability analysis.

Two approaches are generally used to handle non-identifiability: first, the model structure itself is investigated with respect to non-identifiabilities. This approach is referred to as a priori or structural identifiability analysis, as the model structure is examined before simulating and fitting procedures. Within the second approach, a posteriori or practical identifiability study (Ashyraliyev, 2008), non-identifiabilities are detected by fitting to data and investigating parameter estimates. The a posteriori identifiability analysis is used to verify the reliability of parameter estimates, their correlations and to determine the non-identifiable and sloppy parameters in order to provide the reliable biological interpretation of modeling results. These uncertainties in the model parameters may become particularly problematic when models are used explicitly to extract biological information from estimated parameter values or for prediction of dynamical behavior of the model at different parameter values. For example, if several parameter values are fixed, and if these parameters are correlated with those estimated by fitting, the correct prediction of the system behavior may become infeasible.

Within SYSPATHO, we implemented a sequential method: first, we apply the method based on confidence intervals (Bates et al., 1988, Ashyraliyev et al., 2008). Confidence intervals are constructed for parameter estimates in the vicinity of model solution. Second, a collinearity analysis method is applied to detect correlations between parameters (Brun, 2001). The method was thoroughly tested on a gene circuit model of *Drosophila* segment determination (Jaeger et al., 2004) and a model of transcriptional control of the *Drosophila* even-skipped gene (Janssens et al., 2006).

Publications resulting from SYSPATHO:

* Kozlov K, Surkova S, Myasnikova E, Reinitz J, Samsonova M (2012) Modeling of Gap Gene Expression in *Drosophila* Kruppel Mutants. *PLoS Comput Biol* 8(8): e1002635

* N. V. Ivanisenko , E. L. Mishchenko , I. R. Akberdin , P. S. Demenkov , V. A. Likhoshvai , K. N. Kozlov , D. I. Todorov , M. G. Samsonova , A. M. Samsonov , N. A. Kolchanov , V. A. Ivanisenko (2013) Replication of the subgenomic hepatitis C virus replicon in the presence of the NS3 protease inhibitors: a stochastic model. *Biophysics* (58)5: 592-606

* Nikita V. Ivanisenko , Elena L. Mishchenko , Ilya R. Akberdin , Pavel S. Demenkov , Vitaly A. Likhoshvai , Konstantin N. Kozlov , Dmitry I. Todorov , Vitaly V. Gursky , Maria G. Samsonova , Alexander M. Samsonov , Diana Clausznitzer , Lars Kaderali , Nikolay A. Kolchanov , Vladimir A. Ivanisenko (2014) A New Stochastic Model for Subgenomic Hepatitis C Virus Replication Considers Drug Resistant Mutants . *PLoS One* 9(3): e91502

* E. Myasnikova (2013) Identifiability analysis and predictive power of the gene circuit model, *Proc. Int. Moscow Conf. Comput. Mol. Biol. MCCMB - 13*, 25-28 August, 2013, Moscow.

** Model analysis **

Models in the field of systems biology are usually formulated in terms of high-dimensional systems of ordinary differential equations (dynamical systems), which depend on a vector of parameters. Methods of the dynamical systems theory are widely used for analysis of qualitative and quantitative behavior in these models (Hirsh et al., 2004). The methods aim to characterize with as much details as computationally possible a phase portrait of the dynamical system based on placement in the phase space of stationary and non-stationary attractors, invariant manifolds and basins of attraction of all attracting sets. Methods of the dynamical systems theory are important to study robustness of the system, which is defined as the

the dynamical systems theory are important to study robustness of the system, which is defined as the property of reliably performing a biological function under variable conditions. The biological robustness finds its mathematical formulation in terms of sensitivity of the phase portrait of the system to changes in parameter values.

There had been numerous applications of this approach to biological robustness in many biological systems, from bacteria chemotaxis to development in *Drosophila* (Tyson et al., 2001, MacArthur et al., 2009, Manu et al., 2009). Previously existing algorithms and software for calculating these objects are mostly suited for comparatively low-dimensional dynamical systems. On the other hand, being essentially multi-dimensional, systems biology oriented dynamical systems have various specific features, such as scalability, which make it possible to extend the existing algorithms to these systems. The main innovations within SYSPATHO are a detailed description of the phase portrait and its dependence on parameter values, as well as the prediction of various qualitative scenarios for the system dynamics for various parameter regimes or biological conditions.

We developed computationally efficient methods and implemented them in our in-house software Basin to calculate attractors and their basins of attraction for multi-dimensional dynamical systems for signaling and genetic networks. The program calculates attractors and attraction basins in models of genetic networks by direct evaluation of the model equations from a randomly sampled set of initial conditions. New methods for this calculation were investigated, including reverse-in-time calculations of attraction basins and local search methods for refinement of the attraction basin boundaries in the phase space. The reverse-in-time method calculates attraction basins by evaluating the model equations with reversed time and starting from a local vicinity of a given stationary attractor. The local search methods for the attraction basin boundaries allow reconstructing the boundary starting from a given point on it by means of specific sampling of initial conditions only in a local vicinity of this and newly found points of the boundary. These methods are computationally more feasible than methods based on the direct sampling of initial conditions and are applied for calculations in low-dimensional projections of the phase space. An extended version of the program Basin can be applied to a wider class of systems biology models (nonlinear ordinary differential equations with polynomial nonlinearities which are particularly used for modeling mass-action kinetics in signaling and genetic networks).

Publications resulting from SYSPATHO:

- * Kozlov K, Surkova S, Myasnikova E, Reinitz J, Samsonova M (2012) Modeling of Gap Gene Expression in *Drosophila* Kruppel Mutants. *PLoS Comput Biol* 8(8): e1002635
- * N. V. Ivanisenko , E. L. Mishchenko , I. R. Akberdin , P. S. Demenkov , V. A. Likhoshvai , K. N. Kozlov , D. I. Todorov , M. G. Samsonova , A. M. Samsonov , N. A. Kolchanov , V. A. Ivanisenko (2013) Replication of the subgenomic hepatitis C virus replicon in the presence of the NS3 protease inhibitors: a stochastic model. *Biophysics* (58)5: 592-606
- * Nikita V. Ivanisenko , Elena L. Mishchenko , Ilya R. Akberdin , Pavel S. Demenkov , Vitaly A. Likhoshvai , Konstantin N. Kozlov , Dmitry I. Todorov , Vitaly V. Gursky , Maria G. Samsonova , Alexander M. Samsonov , Diana Clausznitzer , Lars Kaderali , Nikolay A. Kolchanov , Vladimir A. Ivanisenko (2014) A New Stochastic Model for Subgenomic Hepatitis C Virus Replication Considers Drug Resistant Mutants . *PLoS One* 9(3): e91502

Software packages resulting from SYSPATHO:

software packages resulting from SYSPATHO.

* Methods are implemented for the calculation and visualization of attractors and attractor basins in our in-house software Basin, which is freely available upon request from SPBSPU.

Potential Impact:

SYSPATHO has had an impact on the following areas:

- * Development of novel methods for systems biology, with general applicability in the field
- * A closer collaboration between European, associated and EECA countries, in particular with Russia
- * Advancement of our understanding of hepatitis C infection and interactions of HCV with the immune response of the host
- * Leadership in key scientific areas in systems biology and HCV research through collaboration

** European Dimension and Impact on the collaboration between European and EECA countries **

A key objective achieved through SYSPATHO was the establishment of closer collaboration between EU, associated countries and countries in the Eastern Europe/ Central Asia (EECA) region, in particular Russia. All scientific work packages in SYSPATHO involved both Russian and EU groups, and the

successful completion of almost all tasks required close collaboration and a combination of efforts between European and Russian partners. This led to tight exchange of knowledge between Russian and EU scientists, and significantly contributed to close scientific ties. The inclusion of two academic groups from different provinces in Russia in SYSPATHO was further complemented by a Russian SME partner, who linked the research objectives of SYSPATHO with industrial exploitation in the EECA region, and as such played a major role in the consortium.

To further strengthen collaboration between EU, associated country and Russian groups, parts of the SYSPATHO budget were set aside for the exchange of scientists at all levels of their academic career, starting with PhD students, but extending also to the exchange of Principal Investigators and Professors. The partners mutually exchanged scientists for joint work on SYSPATHO on a number of occasions, with extended stays of personnel at the partner institution. By the exchange of personnel, not only SYSPATHO itself benefited from the resulting knowledge transfer, but also involved young scientists were educated at an international level and got to know different scientific systems. Thus, they have been qualified further for an academic career in their own home countries or abroad. In addition to tight collaboration between the partners and partner institutions, SYSPATHO combined European and Russian research efforts in the field of mathematical modeling and systems biology further through collaboration with other research projects funded by the Russian Federal Agency of Science and Innovations (FASI). Partners involved in SYSPATHO also participated in FASI-funded collaborative projects in the EECA region, with related scientific objectives. For example, the Russian Federal Agency of Science and Innovations supported a collaborative research project headed by the department of participant 4 on the “Development of Web-oriented expert system for identification of interrelated proteins identified using the post-genomics techniques” and a second project on “Development of software-informational resource on the eukaryotic gene expression control”. Related scientific questions on method development for Systems Biology and Bioinformatics arose in these projects, and exchange of methods and knowledge between SYSPATHO and these projects led to benefits for both sides.

As an additional means to strengthen collaboration with scientists and other projects in the EECA region, to increase the visibility of SYSPATHO in the region, and to attract interest of scientists and industry alike

to increase the visibility of SYSPATHO in the region, and to attract interest of scientists and industry alike, SYSPATHO organized a joint Russian-European workshop on Computational Systems Biology in Russia in September 2012 in St. Petersburg. With close to 100 participating researcher from relevant fields such as systems biology, virology, mathematics and computer science, this workshop served as a platform to install and expand an EECA-European research network on systems biology, in order to combine and capitalize on the highly developed skills and knowledge in both EU/AC and EECA countries.

** Impact on the competitiveness of the participants **

Due to the interdisciplinary nature of systems biology research, involving molecular and cellular biology, mathematical modeling, computer science and optimization, dynamical systems theory, and other fields, successful research with a visible impact on the field requires collaboration and knowledge exchange between participants with different backgrounds. Furthermore, due to the extensive genomic and molecular data and technological expertise required, for example high throughput screening, large-scale live cell imaging, and large-scale sequencing projects, single academic groups cannot compete internationally anymore, and large collaborative projects at the national and European level are essential to maintain competitiveness at the international level. Through the cooperation between the partner institutions in SYSPATHO, the project contributed to the establishment of a research network on the systems biology of HCV-Host interactions in Europe and Russia, which made it possible to bundle the expertise from different member states and thus gain scientific and technological leadership in this important research field. This network with its unique resources and expertise made it more attractive also for third parties to join, and attracted researches from abroad to work at partner institutes that are well connected within this kind of network. Furthermore, both Russian and EU/AC projects and personnel benefited from mutual exchange of information and researchers, and a combination of efforts within and beyond SYSPATHO.

By integrating partners at the European level, SYSPATHO brought together trans-national expertise on virus host interactions, systems biology and method development. Building on strong experimental groups in Germany and France, integrating mathematical, systems biology and bioinformatics groups from Russia, Israel, Turkey and Germany, and blending this with the expertise of two industrial partners from Russia and Cyprus, SYSPATHO brought together a critical mass of scientific researches that enabled SYSPATHO participants to gain leadership in a key scientific area, which is not only the foundation for further transnational collaboration in the established research network, but due to research results obtained in the project also brings about important health-related implications for Hepatitis C research. Importantly, based on developed models and biological insight into HCV replication, several partners from SYSPATHO were successful to form SysVirDrug, one of 7 successful consortia in the first call of ERA-Net for Systems Biology Applications. SysVirDrug is taking approaches from SYSPATHO further in order to translate systems virology data into broad-spectrum antiviral drugs, which can be used to efficiently treat a panel of different viruses of medical importance.

The two SME partners involved in SYSPATHO benefitted strongly from the participation in this international consortium by gaining access to technologies and knowledge developed within the consortium, and by establishing collaboration and scientific contacts with partners at the European level. This led to the opening of new markets for the involved SMEs, and broadly expanded their visibility both geographically as well as in the scientific community. Additionally, a spin-off company, ENVO pharma, has

geographically as well as in the scientific community. Additionally, a spin-off company, ENVO pharma, has been created under participation of INSERM, which will exploit results from SYSPATHO.

**** Societal and economic impact, and impact on Systems Biology research ****

SYSPATHO addressed two key issues which have an important societal and economic impact at the European level: Firstly, the improvement of existing and development of novel mathematical methods for systems biology and their dissemination in the scientific community has significant influence on basic research in systems biology in the European Union and beyond. The integration of different data types in SYSPATHO and the associated data analysis and modeling required novel computational tools that integrate different data types and span several levels of biological systems, considering both intracellular and inter-cellular effects. Due to the complexity of the underlying biological processes, for example affecting virus-host interactions, which can only be understood at the systems level, neither forward modeling using differential equations nor data driven approaches alone can succeed. The combination of forward and inverse modeling approaches as pursued in SYSPATHO and their rigorous testing on an important, health-related biological system has laid the foundation for the application of these methods also in other projects on the systems biology of human diseases.

Secondly, the development of new mathematical models of hepatitis C virus (HCV) infection have been used for the identification of new potential drug candidates, as well as treatment regimes, using dynamical systems analysis. Hepatitis C virus (HCV) infection is a major global health problem with 170 million chronically infected individuals worldwide and 3 to 4 million new infections occurring each year (Rantala et al., 2008). The data available for Europe indicate a wide variation in HCV prevalence, ranging between 0.1 to 6.0% in various countries (Esteban et al., 2008). It is assumed that about 86.000 deaths due to chronic hepatitis C occurred in the WHO European region in 2002. This is more than twice the number of deaths estimated for HIV/AIDS (Mühlberger et al., 2009). A major reason is the insidious course of diseases; HCV infection persists in up to 80% of cases and is mostly asymptomatic. However, these persons are at high risk to develop liver disease later, most notably liver cirrhosis and hepatocellular carcinoma (HCC). Currently, there is no vaccine available. Moreover, the standard-of-care treatment consisting of pegylated interferon-alpha and ribavirin is costly, has limited efficacy, serious side-effects and poor tolerability. Despite the recent approval of new interferon-free treatments using direct acting antivirals, these drugs are still not available to many HCV patients as they are expensive. Hence, there is still a need for low-cost alternatives, in particular for the countries most affected by HCV. To this aim, models developed in SYSPATHO will continued to be used in further projects analyzing treatment regimes and drug targets. In particular, SysVirDrug is aimed at translating systems biological results from SYSPATHO into drug development for broad acting antivirals.

Using the methods and models of HCV-host interactions developed in SYSPATHO, novel insights into the pathogenicity of HCV were obtained. SYSPATHO developed new models of intracellular HCV replication, antiviral innate immune response of the cell upon recognition of viral infection and its interaction with viral replication, and IFN induction and viral spread at the level of cell populations. Jointly, these models elucidated how HCV establishes persistent infection, and mathematical model analysis generated new hypotheses on load and choke points in this process. In particular, the formation of viral replication vesicles was identified to critically regulate persistent cellular replication and enable a constant low-level replication required for chronicity of infection. Several targets were checked for possible drug candidates

replication required for chronicity of infection. Several targets were checked for possible drug candidates that could inhibit viral replication directly and compounds were identified. Hence, SYSPATHO contributed to the competitive abilities of the European economy, and formed the basis for understanding fundamental mechanisms that contribute to chronification of Hepatitis C infection with obvious health care related implications and economic potential.

** Contribution to policy development and societal objectives **

SYSPATHO provided tools for translational research, as developed in the systems biology research field, and thus contributed to structuring EU efforts and better understand disease mechanisms, drug discovery and translational research, including

- * Integration of data and knowledge driven approaches to modeling complex biological systems
- * Integration of a wide variety of different data types, including RNAi screening, protein interaction data, time resolved live cell imaging data, literature mining, information retrieved from databases and other data sources
- * Pre-processing, dimension reduction and normalization of the acquired data, to make it accessible for modeling
- * Efficient and optimal parameter estimation of models from experimental data, including analysis of structural identifiability
- * Efficient solution of models, and provide the knowledge of a partial analytical solution as a check point for algorithms and numerical simulation methods used and
- * Efficient model analysis regarding robustness and dynamic behavior.

Developed methods were thoroughly tested and applied on an important health related problem, to understand and model biological processes underlying Hepatitis C infection. By developing mathematical models of HCV-host interactions at the systems level, focusing on the struggle between HCV and cellular antiviral responses to infection, SYSPATHO generated novel insights into the processes leading to chronification of Hepatitis C. Using model analysis, key steps and central players in this process were studied, and load and choke points identified, which can be translated into current and future therapies. With about 80.000 to 90.000 deaths attributed to HCV infection in the European region, about twice the number than is due to AIDS, research results obtained within SYSPATHO may have a significant impact for health and well-being of European citizens.

** Dissemination activities **

SYSPATHO extended excellence and disseminated knowledge within and outside the project. Dissemination of data, mathematical methods and algorithms and systems biology models generated by SYSPATHO to the scientific community was given highest priority within the consortium. Furthermore, dissemination of knowledge to the public enhanced recognition of the consortium, as well as the overall state of knowledge in the field.

A special focus of SYSPATHO was in strengthening collaboration with other research groups and industry in the Eastern Europe / Central Asia (EECA) region. For this purpose, SYSPATHO organized a joint Russian-European workshop on Computational Systems Biology in September 2012 in St. Petersburg, Russia. With close to 100 participating researcher from relevant fields such as systems biology, virology

Russia. With close to 100 participating researchers from relevant fields such as systems biology, virology, mathematics and computer science, this workshop served as a platform to install and expand an EECA-European research network on systems biology, in order to combine and capitalize on the highly developed skills and knowledge in both EU/AC and EECA countries. A further successful workshop on Systems biology was organized with participation of SYSPATHO in Brazil in 2014. The members of the consortium actively communicated results from SYSPATHO in more than 50 talks at European and international workshops and conferences, giving the project a lot of visibility in the scientific community. Peer-reviewed publications resulting from PATHOSYS were published mostly in open access journals where possible, to ensure wide and free public access to research results of the project.

The general scientific community, as well as science-interested public has been made aware of the work within SYSPATHO through public lectures held in Dresden in 2012, as well as through an interview at the Russian scientific popular web portal <http://academcity.org/content/my-snabzhaem-organizm-oruzhiem-protiv-virusov>. Furthermore, an article has been published on modeling the HCV replication cycle in a journal distributed mainly to subscribed experimental laboratories (Laborwelt 6:13-15, 2011) to familiarize this community with systems biology approaches and to exemplify potential advances in the generation of biological knowledge through mathematical modeling. The project website (www.syspatho.eu) was kept up to date with published findings and news from the project throughout the project duration. In addition, a project flyer was printed in English and Russian language, and has been distributed 2012-2015.

Internal dissemination of knowledge within SYSPATHO was achieved by increasing electronic communication and by creating and continuously updating common knowledge bases. The organizational structure of SYSPATHO and joint work between different partners on crucial tasks within SYSPATHO further strengthened collaboration and knowledge sharing between the partners. This was supported by exchange of scientists between the partners at all levels of their academic career, with a particular focus on exchange of PhD students, but also research stays of principal investigators at partner institutions. Furthermore, at least annual project meetings were organized in order to effectively disseminate the project's results among partners.

**** Use and exploitation of foreground ****

SYSPATHO developed a suite of methods, algorithms and software packages which can be used widely for basic and translational systems biology research. Hence, a significant outcome from the project is the advancement of knowledge, both on the algorithmic side, as well as applied to investigating HCV infection as a medically highly relevant disease. A number of scientific discoveries were made in the course of the project which advanced our understanding of HCV biology and mechanisms for chronicity, which will have a major impact on HCV research, and therefore clinical development in general.

SYSPATHO integrated two small and medium enterprises (SME) as industrial partners, PBSOft LLC in Novosibirsk (Russia) and NovaMechanics Ltd. in Cyprus. PBSOft LLC is a research-based IT company dedicated to develop innovative software tools directed toward the resolving of biomedical and biotechnological tasks. PBSOft is a winner of the competition "START 07" in the field of information technologies, software tools, telecommunication systems, organized by state fund for development of start-up companies in scientific-technical field, with the project "Development of software for computer proteomics". The company is expert in text mining and automated knowledge extraction from factual and textual databases, and holds patents on Associative network database (ANDCell) and reconstruction

textual databases, and novel patents on Associative Network Database (ANDCell) and reconstruction, visualization and analysis of associative networks (ANDVisio). PBsoft deals with the development of a new bioinformatics. NovaMechanics is a consulting and contract research biopharmaceutical company committed to the computer aided design of small-molecule medicines for the treatment of a variety of diseases such as AIDS, hepatitis C, cancer etc. NovaMechanics' techniques can reduce the costs for candidate development by providing activity screening at a very early stage of product development and eliminating inactive compounds early and thus reducing the number of expensive (and sometimes animal consuming) efficacy experiments.

Results from SYSPATHO not only include new algorithms of general importance and applicability for systems biology, but also major improvements in the text-mining technique, which is commercialized through PBSoft. Our improved text-mining methods were used to reconstruct large-scale HCV-Host molecular-genetic networks on the basis of information from PubMed abstracts. The Hepatitis C Associome (HCA) relational database containing integrated literature network with Y2H data describing virus-host interactions was developed and is of a commercial value. These networks represent new informational resources which are of value for solving a number of actual tasks in pharmacology and medicine. PBSoft uses these networks integrating a huge amount of various biological facts and relations as an expert system for reconstruction of gene networks and also for the prediction of potential regulatory pathways, potential drugs and drug targets, etc. PBSoft provides this expertise for research institutes and companies dealing with the investigation of HCV and the development of new efficient drugs. Furthermore, PBSoft uses the technology of text mining and associative networks reconstruction, which was significantly developed in the frame of SYSPATHO, for analysis of other biological data for pharmacological and biotechnological companies. Hence, SYSPATHO increased the competitiveness of PBSoft significantly due to the development of these new bioinformatics resources.

NovaMechanics developed computational drug design techniques further within SYSPATHO, and applied them to screen for potential new-generation anti-HCV drugs. The need for high-quality drug candidates has never been higher, and the pressure to discover those drug candidates is already high, and growing. Hence, SYSPATHO represented an important opportunity for NovaMechanics as technological progress in this field is based mainly on human knowledge and computational systems and does not require extensive financial investments. Within SYSPATHO, Novamechanics Ltd developed further its suite of methods for computational drug design, including small molecule virtual screening. SYSPATHO gave NovaMechanics' work and research activities a wider audience and therefore increased its market share by the attraction of new partners for collaboration. Hence, opportunities have emerged and will continue to emerge for further collaborations and research programs, providing benefits for NovaMechanics, as well as exploitation of results from SYSPATHO targeted at medically relevant problems in Europe and world-wide.

The interaction of SYSPATHO's academic research groups with NovaMechanics enabled the interaction with an industrial research environment. Importantly, this has already led to the successful application for a new European collaborative project within the frame of ERANet, "SysVirDrug", under participation of a number of partners from SYSPATHO. This project is using systems biology approaches including bioinformatics and mathematical modeling to develop novel broad-acting antivirals.

In January 2014, the company Enyo Pharma was created under participation of INSERM as a spin-off based on knowledge generated from SYSPATHO. Enyo Pharma's business model relies on its ability to

based on knowledge generated from STS/AIHO. Enyo Pharma's business model relies on its ability to (1) Identify cellular protein as therapeutic targets for a variety of human-infecting viruses, (2) Validate such cellular therapeutic targets, (3) Initiate clinical trials with repositioned drugs, (4) Develop therapeutic virus-derived peptides and peptidomimetics. With its internal lead programs, Enyo Pharma expects to conduct a phase II clinical trial for the cure of chronic hepatitis B in 2016 and to validate original molecules for the treatment of seasonal and severe flu, Polyomaviruses and HIV reactivation.

Last update: 17 December 2015

Record number: 174228