medical bioinformatics against the hepatitis C virus

Novel approaches and software in systems medicine to combat viral infection

by Mario Albrecht, Hagen Blankenburg, Nadezhda T. Doncheva and Sven-Eric Schelhorn

More than a quarter of a billion people worldwide carry the hepatitis C virus, although most of them are unaware of it. This viral infection manifests itself through chronic inflammation of and considerable damage to the liver. Drug therapies applied until now are only effective in one out of two patients and have numerous side effects. Therefore, scientists intensively seek new target molecules and drugs for better therapeutic approaches. Medical bioinformatics supports this research with new methodological approaches of systems medicine and suitable software for analysing and visualising the large volumes of data generated in the lab.

A widespread, but hidden infectious disease

Hepatitis C is an inflammatory liver disease, which occurs worldwide and is triggered by infection with the hepatitis C virus (HCV) discovered first in 1989. Initially, the disease is inconspicuous in the human body, but, in the course of years, it causes the progressive destruction of the liver, which, if untreated, often leads to liver cancer and ultimately to the patient's death. The virus is transmitted between people primarily by blood and blood products, though the path of infection cannot be traced in about 30% of those infected.

HCV, like the human immunodeficiency virus (HIV) and many other viruses, is highly adaptable and changes its genome sequence continuously. This is why no vaccine against HCV exists yet and why antiviral drugs become ineffective rapidly. A further difficulty is that, for various reasons, the drugs approved until recently show an insufficient effect in half of all patients infected with HCV in Europe.

Therefore, our bioinformatics research at the Max Planck Institute for Informatics in Saarbrücken supports the worldwide search for better drugs to fight HCV. Innovative software helps to analyse viral genome variations and their impact on virus function and drug action. We also develop novel methods for the analysis of experimental data, which will be useful for discovering new target molecules of anti-HCV drugs.

Analysis of viral sequence changes

A starting point for bioinformatics support of HCV therapy is the computational analysis of the viral genome sequences found in the patient. This includes determining the sequence changes by means of which the virus responds to the latest antiviral drugs

Figure 1: The web service geno2pheno[hcv]

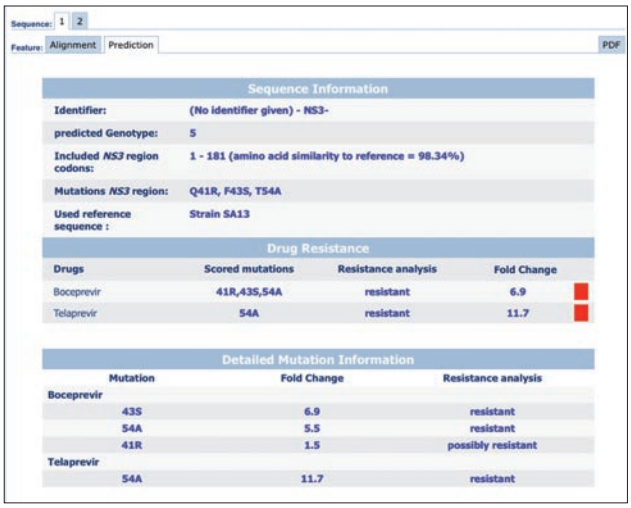
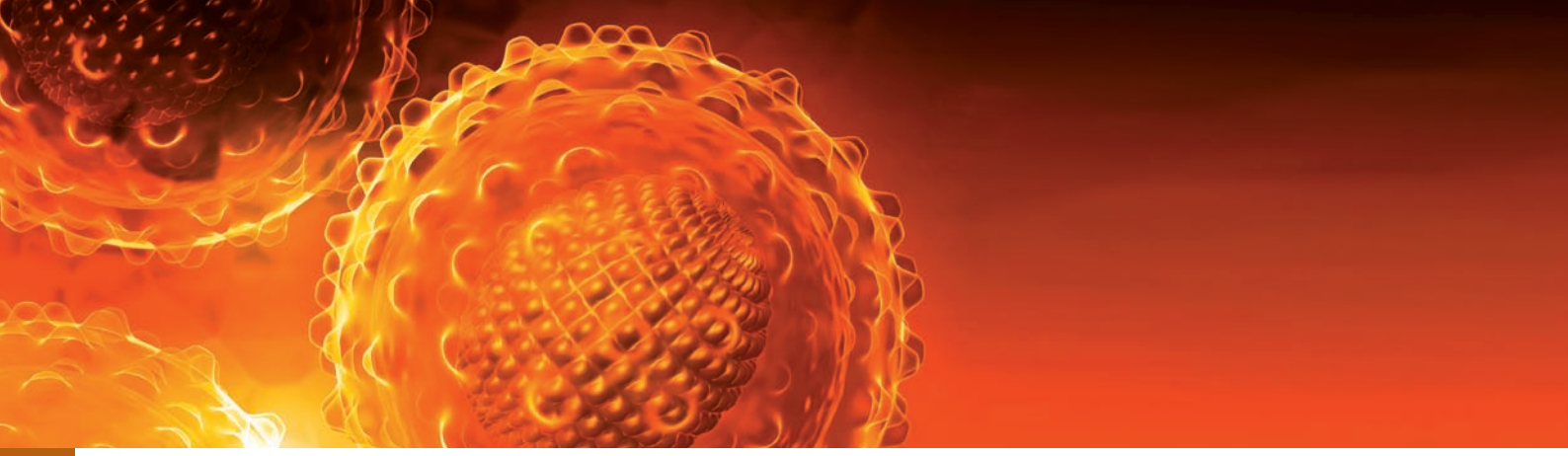


Image of the web service geno2pheno[hcv] for planning the therapy of HCV-infected patients. The identification of viral sequence changes requires different analysis steps of the virus genome, which are bundled in this web service (Image: Sven-Eric Schelhorn).



Picture: © 4designersart – Fotolia.com

and prevents them from binding directly to viral molecules and interrupting its reproductive cycle. To this end, in cooperation with the University Hospital Frankfurt and other hospitals, the viral sequences are determined at different time points of the drug therapy.

Since the latest generation of sequencing technology generates very large amounts of fragmented sequencing data, millions of short sequence fragments first have to be assigned to the right location in the viral genome. This challenging task can only be accomplished by using powerful computers and by developing and applying appropriate computational methods. Subsequently, statistical methods are used in order to identify those changes in the viral genome that make the hepatitis C virus resistant to antiviral drugs.

The resistance information provides doctors with important clues on the course of the viral infection and assists pharmaceutical researchers in finding patient-specific therapies and more effective drugs. For planning these personalised therapies, our research group develops the web service geno2pheno[hcv] (Fig. 1). The web service enables doctors to examine, quickly and free of charge, viral sequences in patients for sequence changes that confer drug resistance. This helps combating the virus more effectively and with fewer side effects for the patient.

Analysis of viral protein structures

Variations in the genome of the hepatitis C virus frequently cause changes of the spatial structures of viral protein molecules. Such three-dimensional structures and the drugs that bind to them have already been determined experimentally with single-atom resolution using X-ray crystallography. Thus, the development of new therapies can be supported by the accurate analysis of the HCV protein structures and the drugs that interact with them. For this purpose, structural biologists and pharmacologists often use specialized software programs for the spatial visualisation of viral protein structures that may consist of up to several thousand atoms. This provides important insights into the molecular mechanisms of the viral protein structure and its function, which might be influenced by drug binding and sequence changes. For example, the latest antiviral drugs block

Figure 2: Visualisation and analysis of the protein structure of the HCV protease NS3-4A using RINalyzer

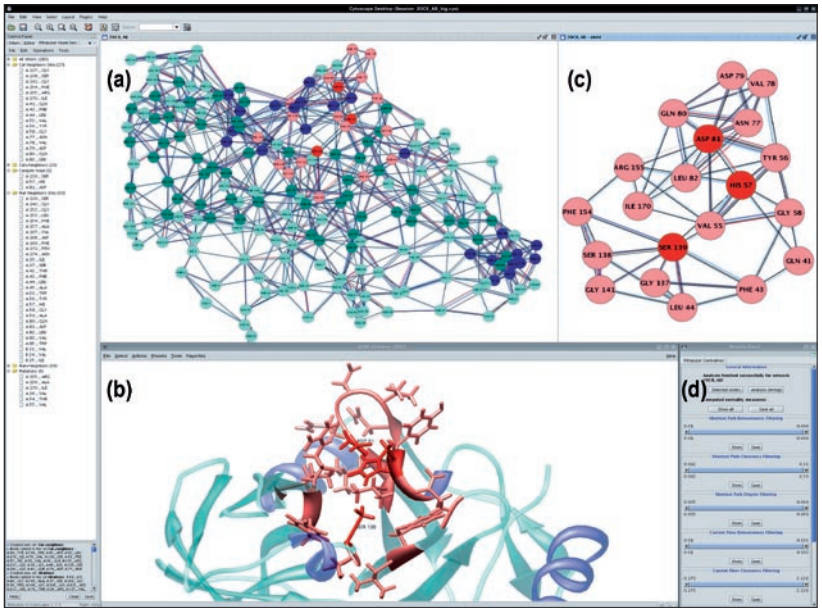


Image: Nadezhda T. Doncheva

(a) Structure network of the HCV protease in 2D;
(b) Protein structure of the protease in 3D;
(c) Close-up showing functionally relevant interactions in the active site of the HCV protease;
(d) Slide control for the visual analysis of the structure network.

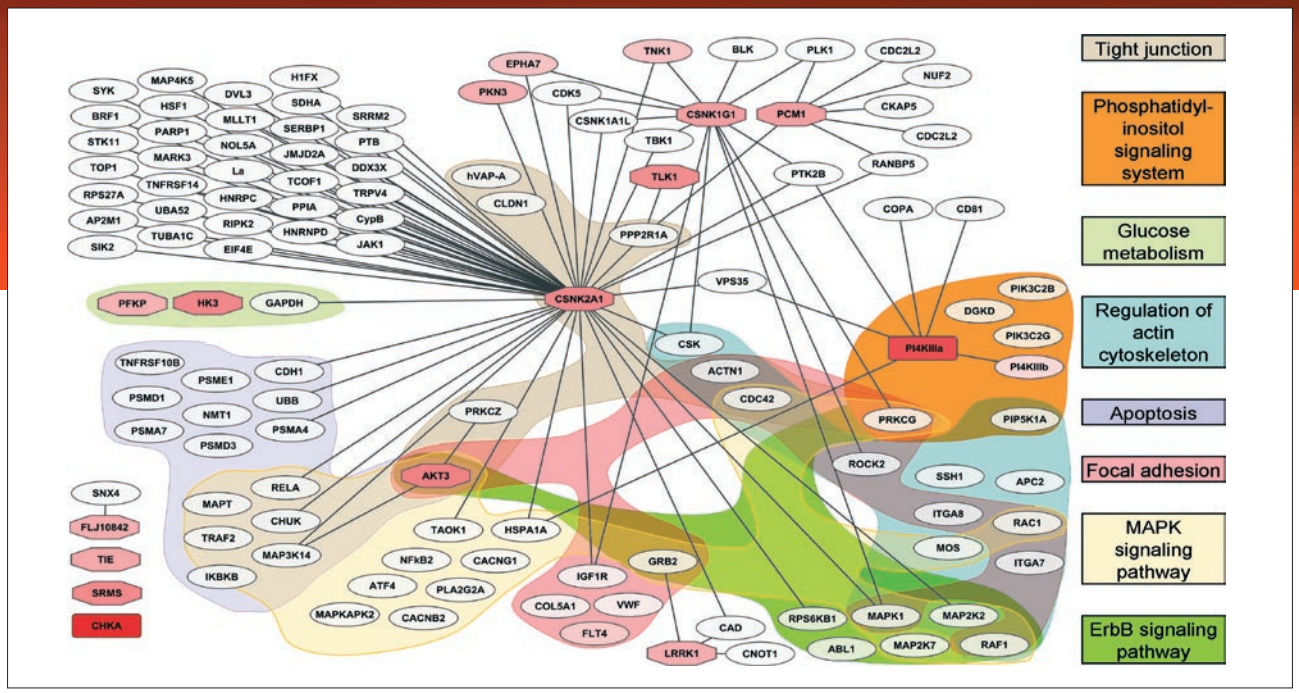


Figure 3: Schematic drawing of selected human host factors of HCV and their interactions in different molecular processes of the human cell

The individual host factors are depicted as ovals, octagons, or rectangles, depending on whether they have been previously known host factors, newly found in the lab during the present study, or already known from earlier experiments, respectively. The deeper the red shade of a host factor, the greater its importance in the experiment for the viral life cycle in the human cell. Interactions between host factors are shown as connecting lines, with some signalling pathways in which certain host factors are involved are highlighted in colour (Image: Hagen Blankenburg).

the function of the viral protease NS3-4A, which is vital for HCV reproduction, whereupon the virus responds by changing its sequence and structure (Welsch *et al.*, 2008).

To further simplify the structural analysis of proteins, in view of the large number of atoms to be taken into account, our research group develops novel integrative visualisation software (Doncheva *et al.*, 2011, 2012). It visualises the atoms of protein structures and their interactions as network in only two dimensions in addition to three dimensions as before (Fig. 2). This simplified network representation complements the established three-dimensional structure analysis, particularly when studying numerous complex interactions between atoms in the protein structure. Because of the great number of interactions, especially across large molecular distances, they can no longer be displayed clearly in three dimensions on a two-dimensional computer screen. This makes our new approach especially suitable for identifying all sequence variations that are relevant to drug effectiveness as well as for visualising and detailed understanding their molecular effects on protein structure and function.

Analysis of human host factors

Drugs that do not target viral molecules directly, but act indirectly by targeting molecular factors in the human host that are essential for the virus, offer an alternative approach to combating HCV. Viruses need these human host factors in order to enter liver cells, to replicate in them, ultimately to leave them and to infect further cells. Better knowledge of the many host factors therefore enables a more comprehensive understanding of the different stages of the viral life cycle in human cells. The most important factors are then potential human target molecules of antiviral drugs, because HCV cannot inhibit the effectiveness of this kind of drug therapy by changing its viral genome.

To this end, our research group cooperates with virologists at the University Hospital Heidelberg in their efforts to discover new human host factors for HCV. The lab experiments performed for this purpose generate extensive measurements that require special bioinformatics methods for data analysis and interpretation. A main focus of our work is to enrich these experimental findings with additional function information from other molecular bio-

logy data sources in order to interpret the individual lab results in the context of a more global, cellular network of interacting molecules (Fig. 3). Amongst other things, this facilitates the identification of relevant host factors and allowed us to explain the molecular mechanisms in which HCV uses a human protein, the lipid kinase PI4KIIIα, for viral replication (Reiss *et al.*, 2011). Thus, impairing the protein function of this critical host factor is now a possible target of future drugs to block the viral life cycle.

In summary, medical bioinformatics contributes, through integrative data analysis, towards uncovering host factors and their molecular interactions as potential targets of new drugs in human cells. Computational methods speed up the elucidation of disease causes at the molecular level and enable the faster development of drugs.

The research project in brief:

Dr. Mario Albrecht has been research group leader for Molecular Networks in Medical Bioinformatics at the Max Planck Institute for Informatics and in the Multimodal Computing and Interaction (MMCI) cluster of excellence in Saarbrücken. He has also been a member of the clinical research group "Mechanisms of resistance development and optimisation of antiviral strategies in hepatitis C virus infection using integrative models of biomathematics and bioinformatics" (KFO 129) funded by the German Research Foundation (DFG). Recently, he has become Professor of Bioinformatics at the Institute of Biometrics and Medical Informatics at the University Medicine Greifswald.

References:

geno2pheno[hcv], http://hcv.bioinf.mpi-inf.mpg.de

RINalyzer, http://www.rinalyzer.de

Doncheva, N.T., Klein, K., Domingues, F.S., Albrecht, M. (2011). Analyzing and visualizing residue networks of protein structures. Trends Biochem. Sci., 36(4):179-182.

Doncheva, N.T., Assenov, Y., Domingues, F.S., Albrecht, M. (2012) Topological analysis and interactive visualization of biological networks and protein structures. Nat. Protoc., in press.

Reiss, S., Rebhan, I., Backes, P., Romero-Brey, I., Erfle, H., Matula, P., Kaderali, L., Poenisch, M., Blankenburg, H., Hiet, M.S., Longerich, T., Diehl, S., Ramírez, F., Balla, T., Rohr, K., Kaul, A., Bühler, S., Pepperkok, R., Lengauer, T., Albrecht, M., Eils, R., Schirmacher, P., Lohmann, V., Bartenschlager, R. (2011). Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. Cell Host Microbe, 9(1):32-45.

Welsch, C., Domingues, F.S., Susser, S., Antes, I., Hartmann, C., Mayr, G., Schlicker, A., Sarrazin, C., Albrecht, M., Zeuzem, S., Lengauer, T. (2008). Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of HCV. Genome Biol., 9(1):R16.1-18.

Contact:

Dr. Mario Albrecht
Research Group Leader
Max Planck Institute for Informatics
Saarbrücken
mario.albrecht@mpi-inf.mpg.de

Hagen Blankenburg
Max Planck Institute for Informatics
Saarbrücken
hagen.blankenburg@mpi-inf.mpg.de

Nadezhda T. Doncheva
Max Planck Institute for Informatics
Saarbrücken
nadezhda.doncheva@mpi-inf.mpg.de

Sven-Eric Schelhorn
Max Planck Institute for Informatics
Saarbrücken
sven.schelhorn@mpi-inf.mpg.de

http://www.medbioinf.de