insilico biotechnology

7 October 2011 – Stuttgart

Current Status and Future Perspectives of HPC at Insilico Biotechnology

Klaus Mauch

Outline

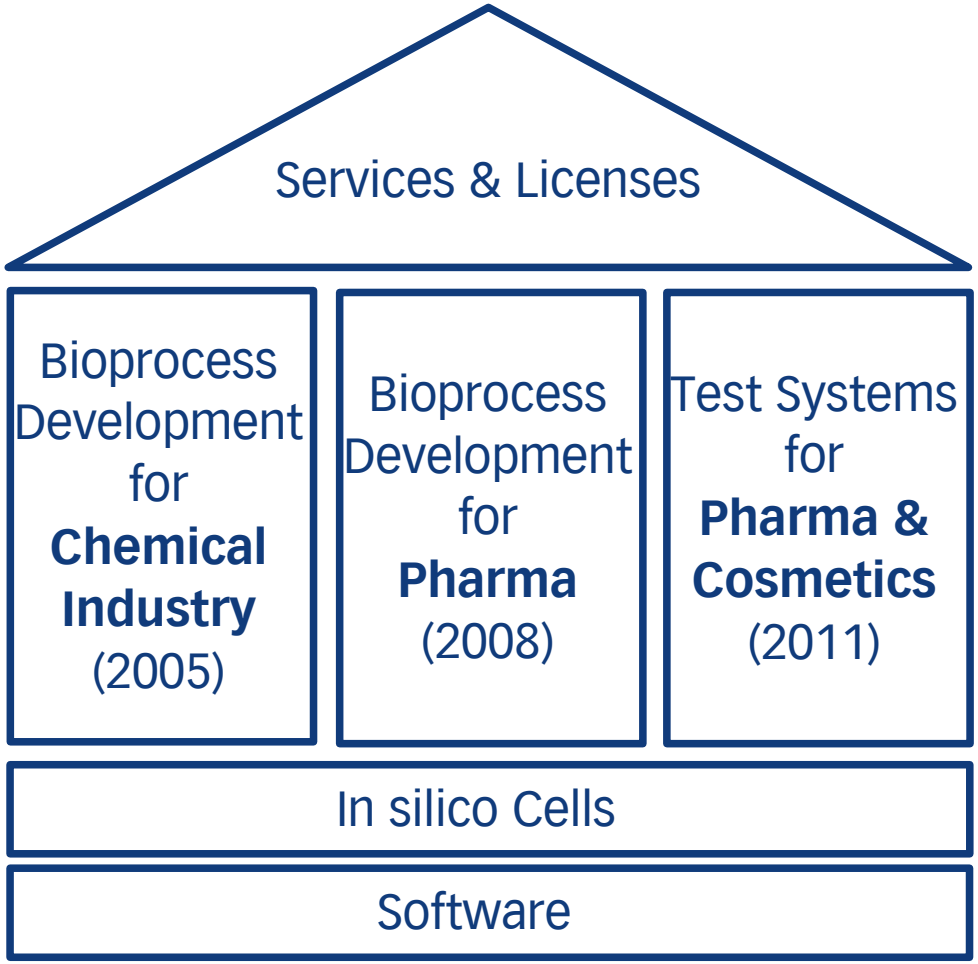
- **In silico Biotechnology** and the Bio-Economy at a Glance
- **Current Status of HPC** at Insilico: Verification of Dynamic Network Models from Time-Series Data
- **Future Perspectives of HPC** at Insilico: Simulation of Genome Based Whole-Body Models

Company Overview – 1

- Founded in 2001, headquartered in Stuttgart, Germany
- Offering services since 2005
- Inter-disciplinary team comprised of biologists, chemists, computer scientists, physicist, and bioprocess engineers
- Unique technology platform and expertise
- Solution provider for more than 20 international companies and academic research institutes

Insilico designs and optimises cellular processes for the Life Science Industries by building computational models of living cells.

Company Overview – 2



Insilico reconstructs, simulates and predicts the performance of complex cellular systems.

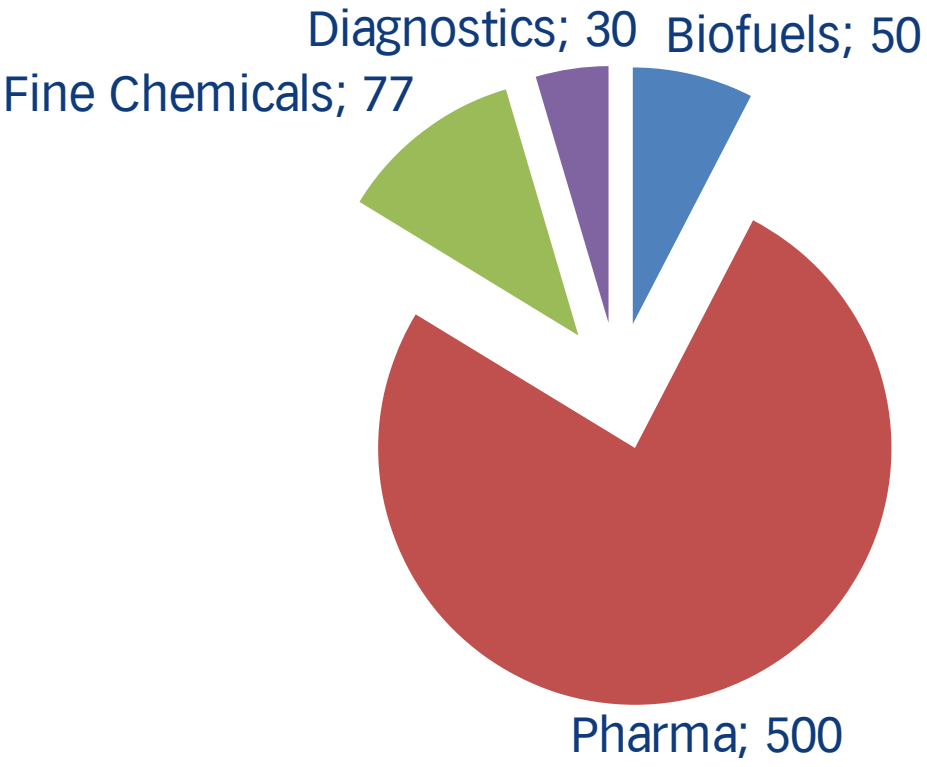
Insilico positions itself as »High-Tech« service provider for Chemical and Pharmaceutical Industries.

Insilico Biotechnology – Stuttgart Office @STEP*

New Address in 2011:
Meitnerstraße 8, 70563 Stuttgart

*Stuttgart Engineering Park

Bio-Economy – Revenues (Billion US$)*



Growth rates of the global Bio-Economy are well above 10%.

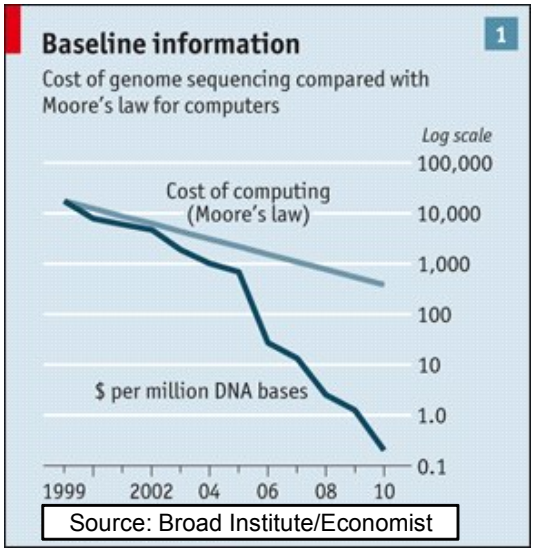
*Numbers depend on market definition; plant biotechnology excluded

Costs for Drug Development

- In 2009, costs for developing a new drug could have been as high as **1.7 billion US$*** compared to 0.8 billion in 2003
- The main reason for rising costs are – in particular – **increasing costs for clinical trials** and the failure of many drugs in late clinical trials
- 6 out of 7 new drugs offer little, if any, clinical advantages over existing drugs*
- The forthcoming complete **ban on animal testing in Europe for cosmetic products** and the lack of validated alternative assessment methods for **long term systemic toxicity testing**, have emphasized on an urgent need to develop better prediction models in this area

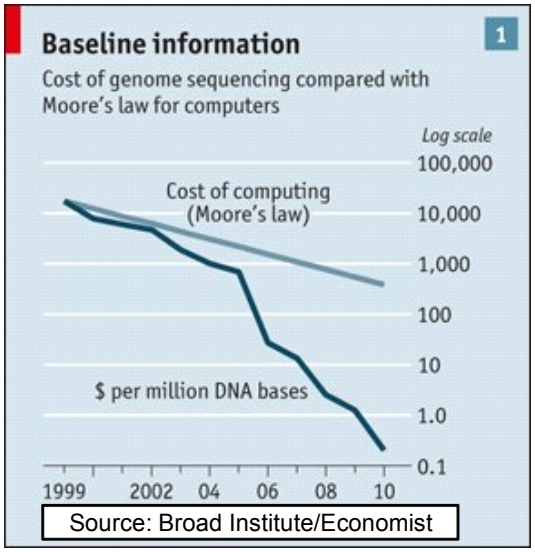
*CMAJ. 2009 February 3; 180(3): 279–280.

Cost of Sequencing vs.Cost of Computing – 1



- Decrease of sequencing costs by factor **70,000** in 12 years (1998 – 2010)
- Pacific Biosciences (US): Mapping the human genome in 15 minutes for less than **$1,000** in 2013
- Decrease of computing costs by factor **4,000** in 12 years (1998 – 2010)

Cost of Sequencing vs.Cost of Computing – 2



- In biotechnology, **data processing and analysis** will become key for generating **added value and competitiveness**
- Increase in speed and reduction in cost generates **new business opportunities** for industrial service providers

Bio-Economy Needs

– Development of Industrial Fermentation Processes

- Need for enhanced efficiency
- Need for new, sustainable bio-processes
- Need for risk assessment
- Need for decreasing process development time
- Need for Know-how/IP

Knowledge-Based Bio-Economy (KBBE) can be defined as "transforming life sciences knowledge into new, sustainable, eco-efficient and competitive products"*.

*2007 EU Cologne Conference: »En Route to the Knowledge-Based Bio-Economy.«

Bio-Economy Needs

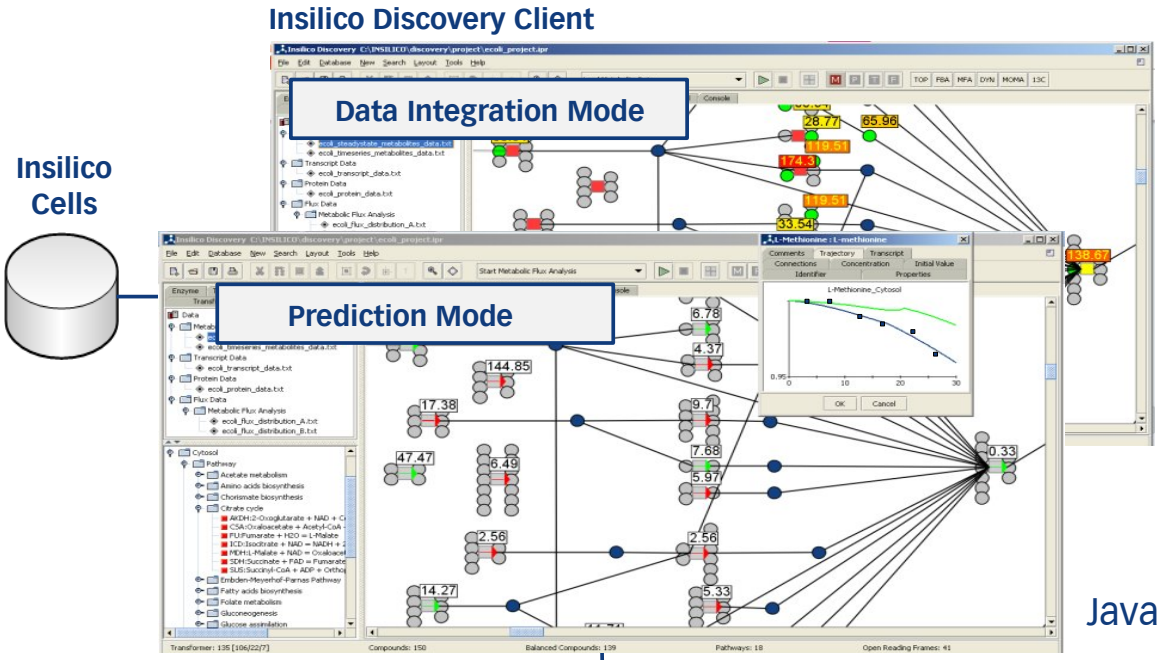
– Drug Development/Testing

insilico biotechnology

- Need for novel methods for realizing
  - **personalized diagnostics**
  - **personalized medicine**
  - **personalized nutrition**
- Need for **alternative toxicity tests**
- Need for integral use of modelling approaches and data integration along the value chain

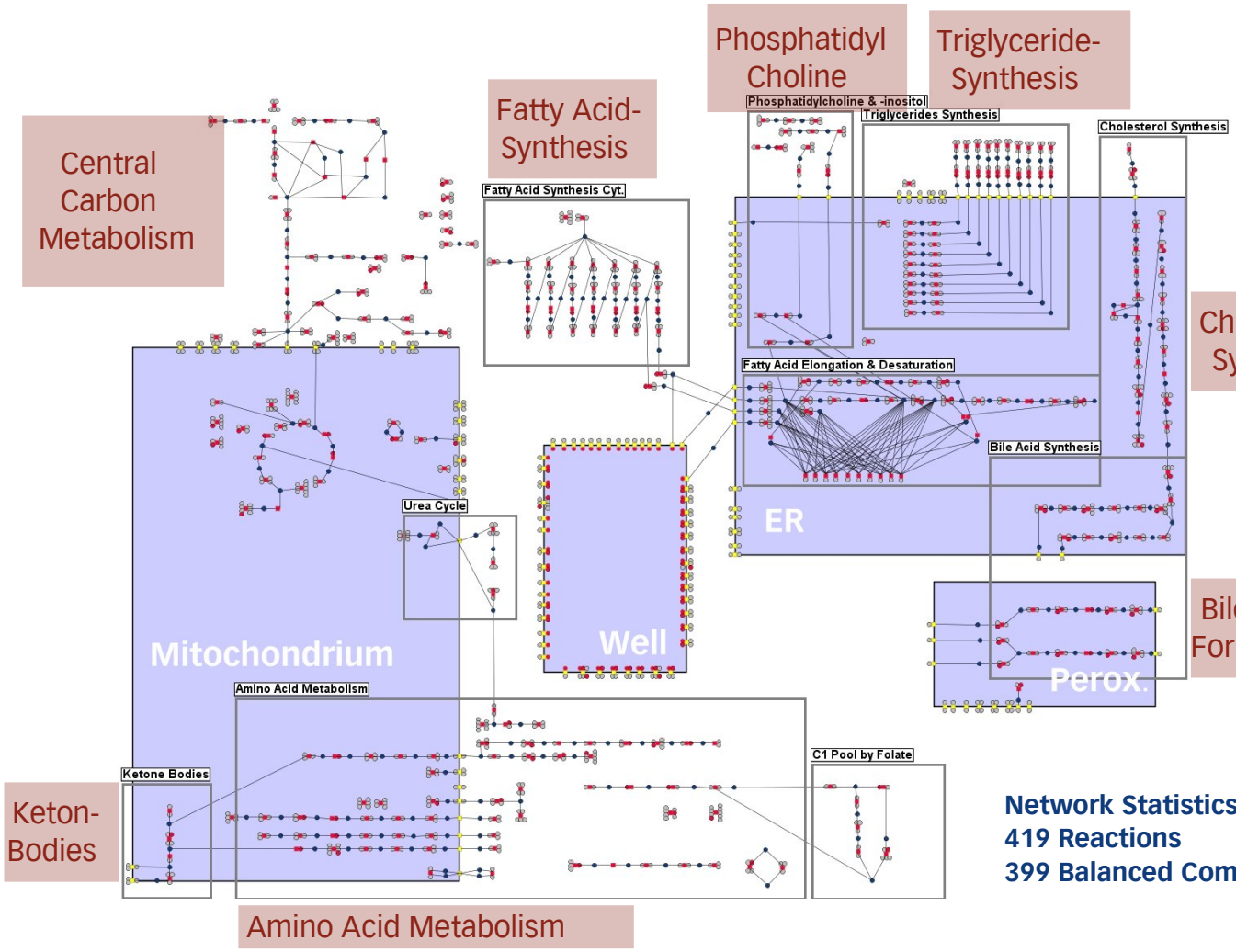
The Knowledge-Based Bio-Economy needs to transform Life-Sciences from a Discovery Science into a quantitative and predictive Engineering Science.

Insilico's Modelling and Simulation Platform



Insilico's modeling tool allows for a graphically oriented reconstruction and simulation of large-scale, genome-based reaction networks.

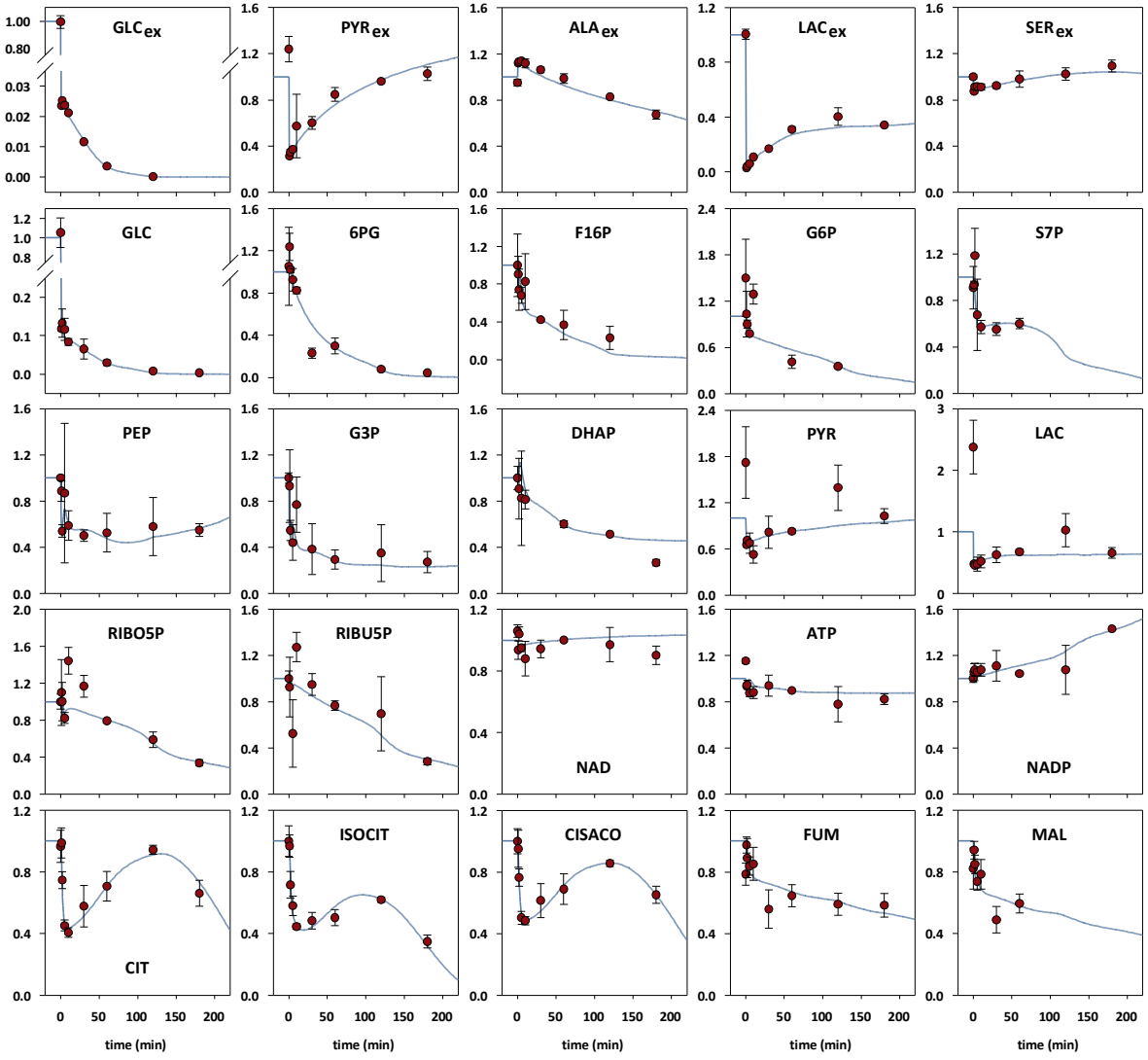
Insilico Liver – Metabolic Network



Network Statistics:
419 Reactions
399 Balanced Compounds (ODEs)

Network-Verification

– Cell Based Assays for Dose-Response Predictions



Glucose depletion at t=0 minutes triggered significant changes in metabolite levels. These changes were used for parametrising reaction kinetics through an evolutionary strategy with covariance matrix adaptation.

In silico dynamics were obtained after 6 million simulation runs (function evaluations).

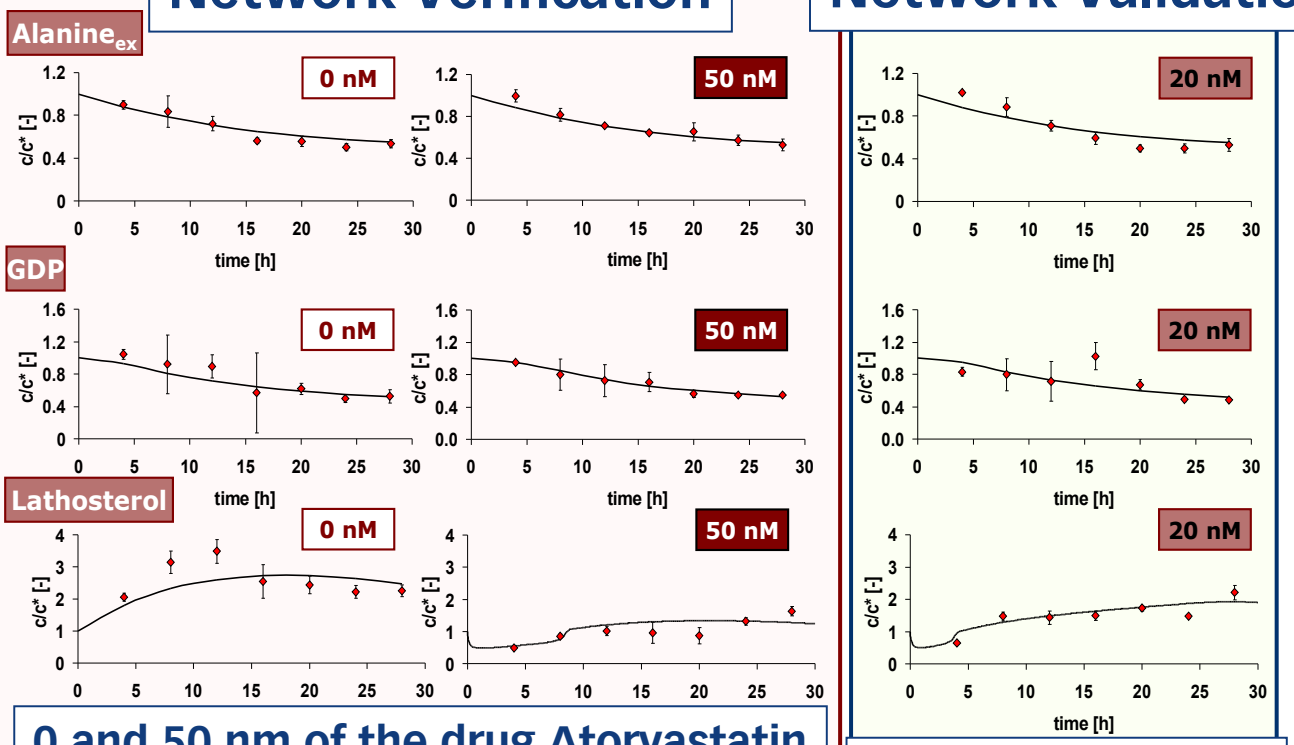
Maier et al. (2010) BMC Syst Biol 4:54.

Network Validation: Predicting the Effect of 20 nM Atorvastatin on Cholesterol Formation

insilico biotechnology

Network Verification

Network Validation



Right column:
Predicted (line)
and **measured** (dots)
metabolite concentrations

Hepatocytes, grown adherently in 6-well plates

0 and 50 nm of the drug Atorvastatin were used for network verification

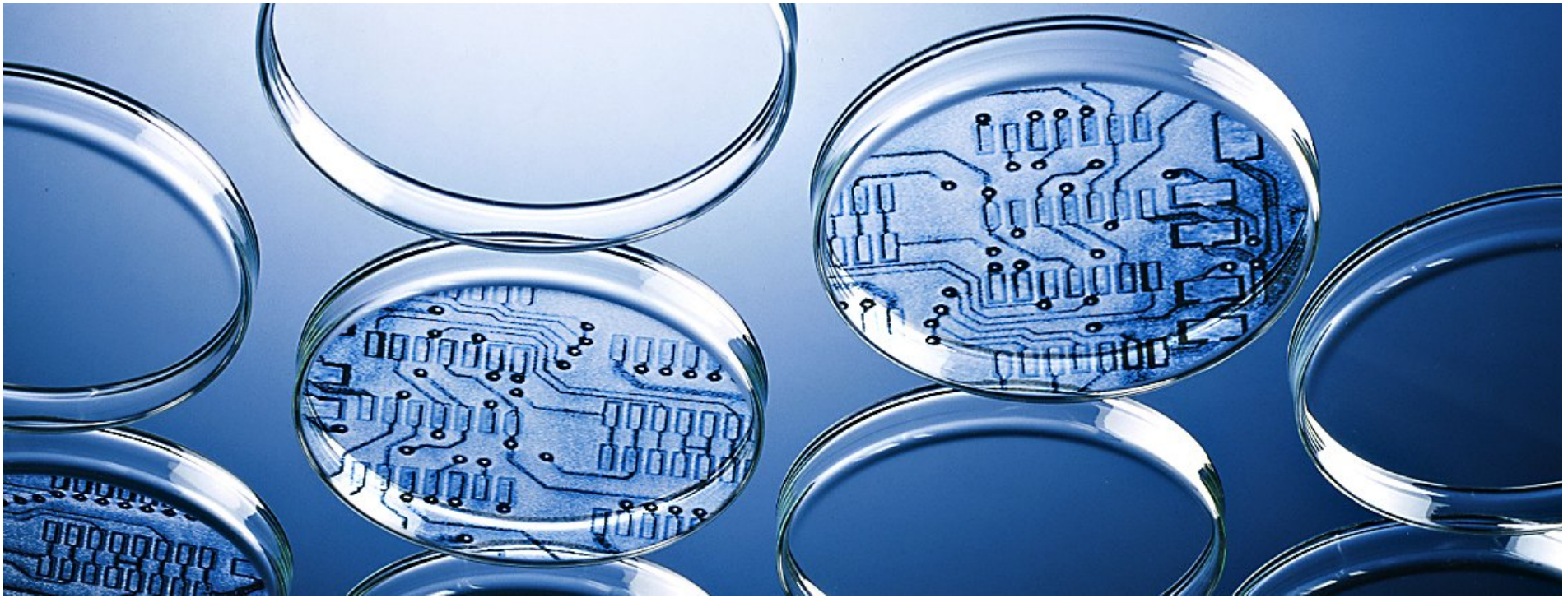
20 nm Atorvastatin

In Silico

In Vitro

HPC for Network Verification (Parameter Identification)

insilico biotechnology



High Performance Network Verification

Aims

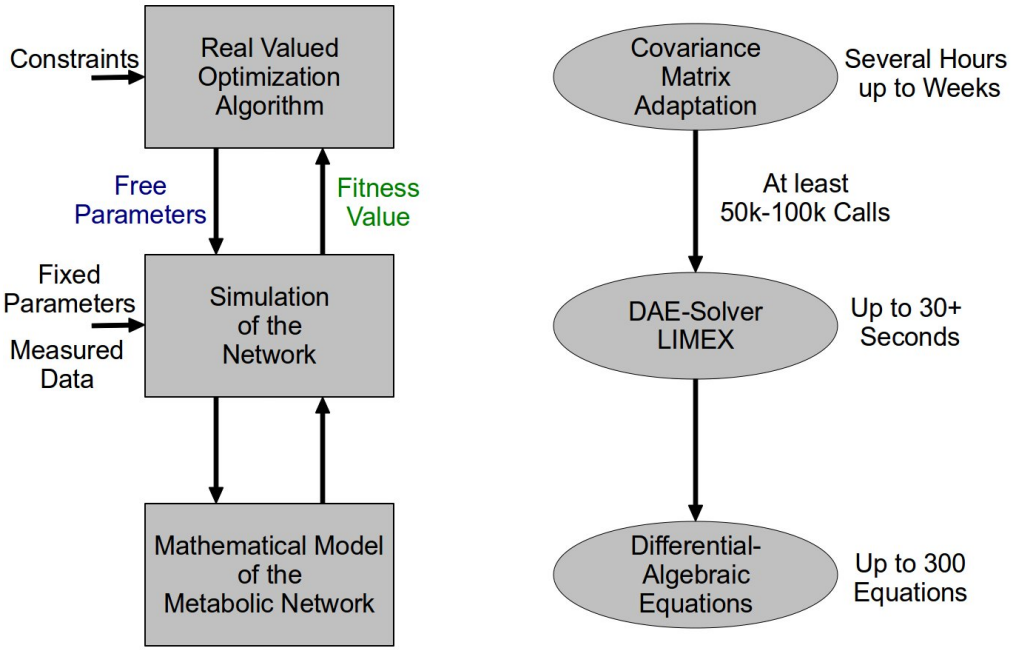
- Downscaling optimization time (from days/months to hours)

Solution

- Modularization of the optimization problem by network decomposition
- Parallelization of the pCMALib algorithm[1)] (synchronous cooperative island model) and the DAE solver LIMEX[2)] by using Open MPI
- Running system on Shanghai (Insilico) and Nehalem (HLRS) nodes
- Documentation/visualization of optimization results

1) C. L. Müller, B. Baumgartner, G. Ofenbeck, B. Schrader et al. 2009 2) Deuflhard, Hairer & Zugck 1987

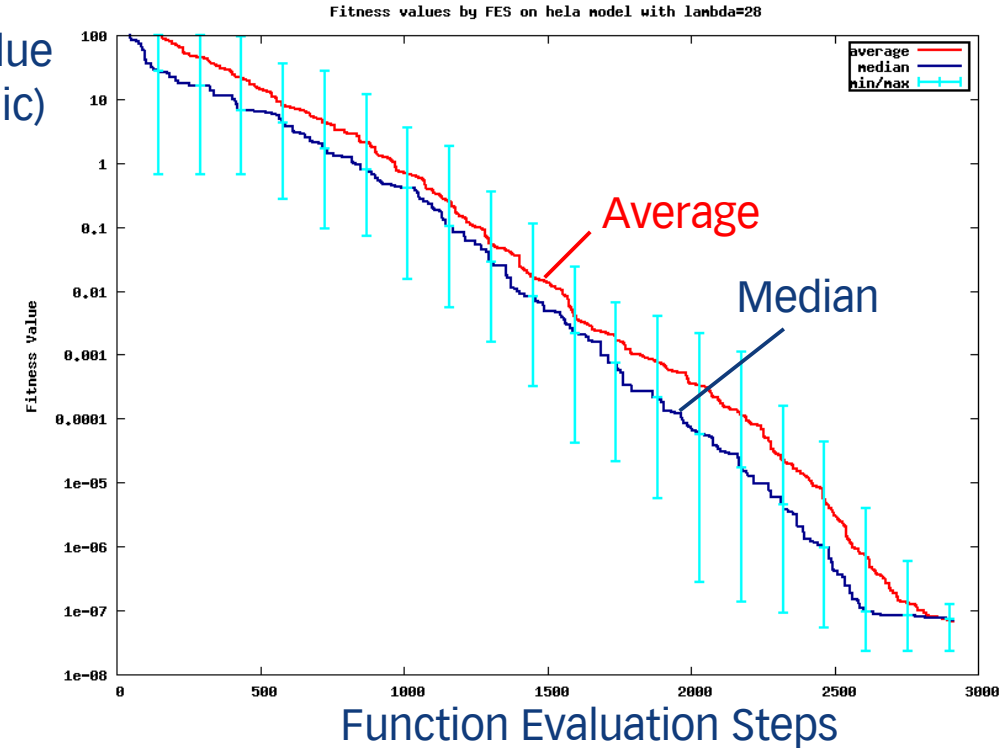
Network Verification – Simplified Workflow



The network simulation time (30+ seconds) rather than the optimization step (< 0.3 seconds) is the critical time step. For estimating 50-100 parameters, about 100.000 simulation runs are needed.

Convergence of Fitness Value
– HeLa Cell-Line

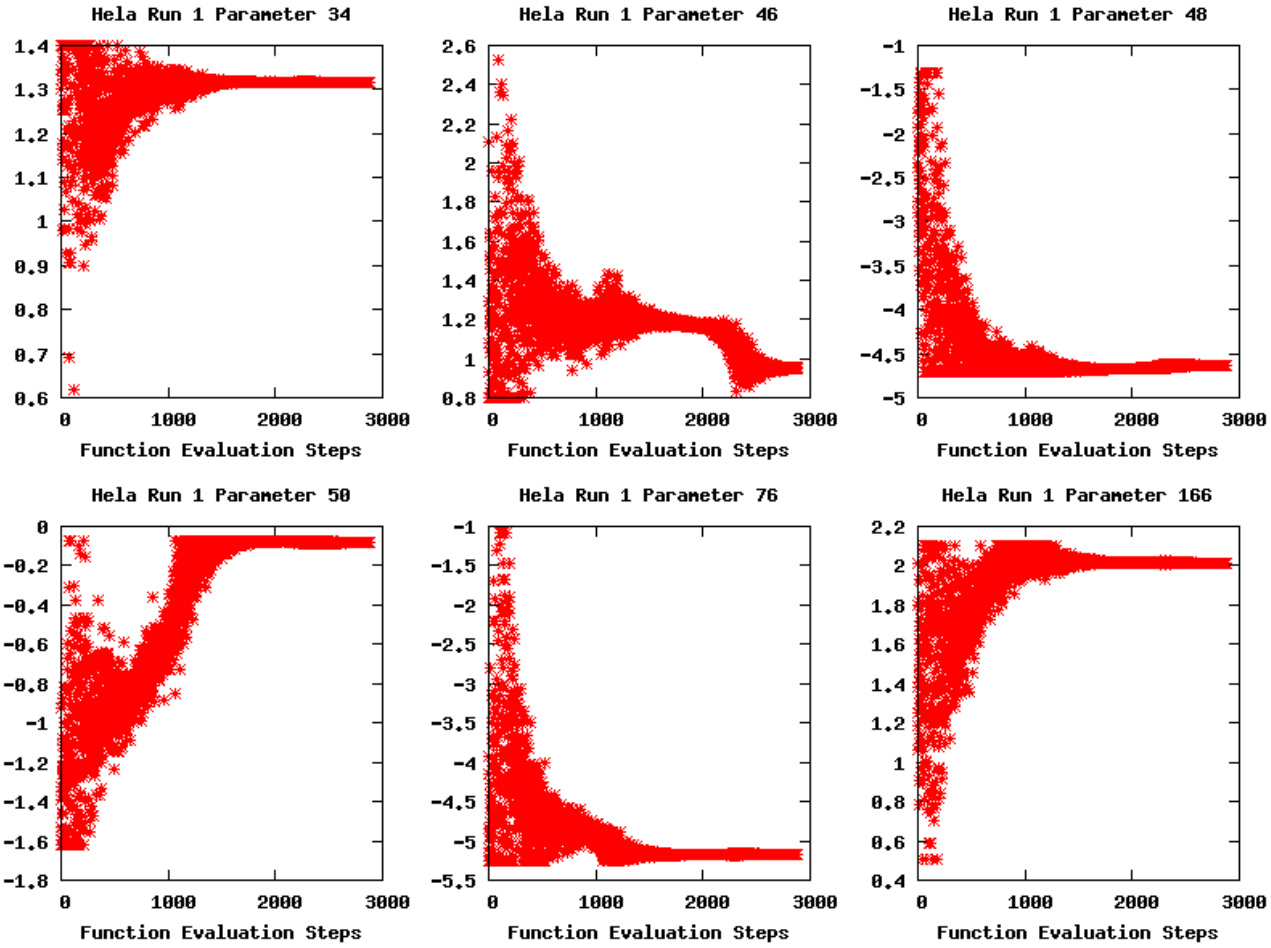
Fitness Value
(Logarithmic)



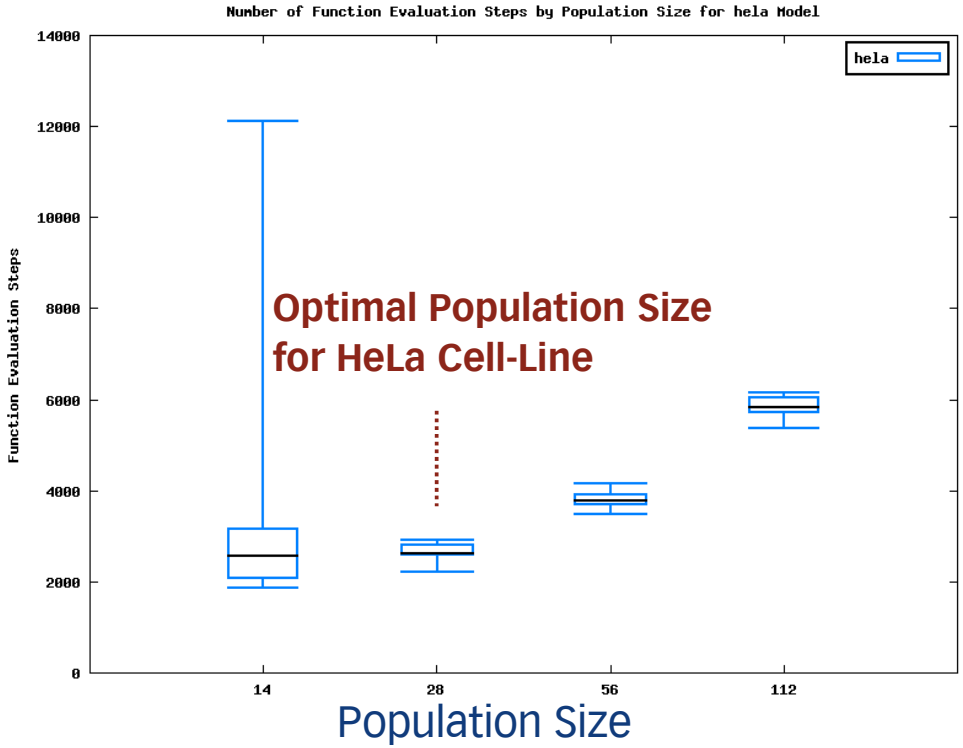
The fitness value is given by the weighted squared sum of deviations between the experimental data and the simulated data.

Progression of Parameter Values

– HeLa Cell-Line

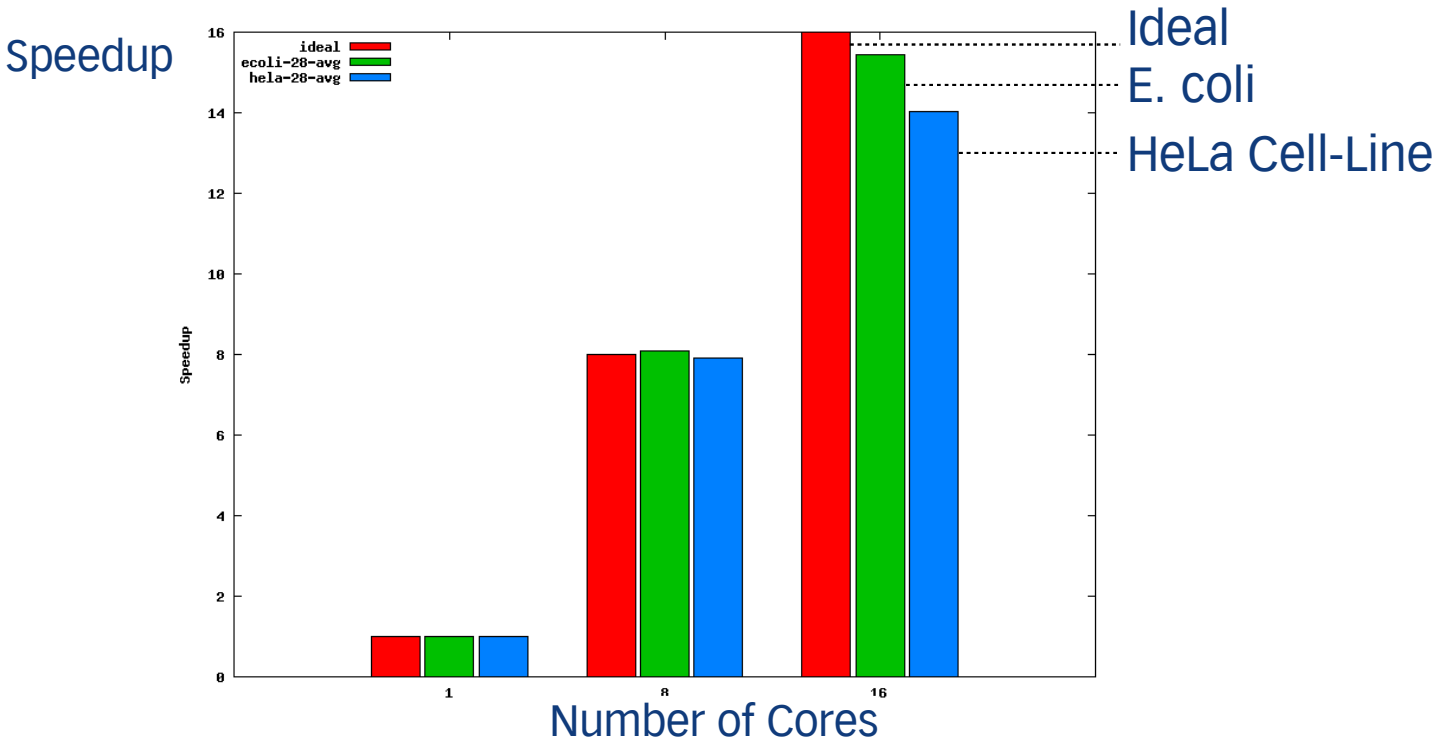


Function Evaluation Steps vs. Population Size – HeLa Cell-Line



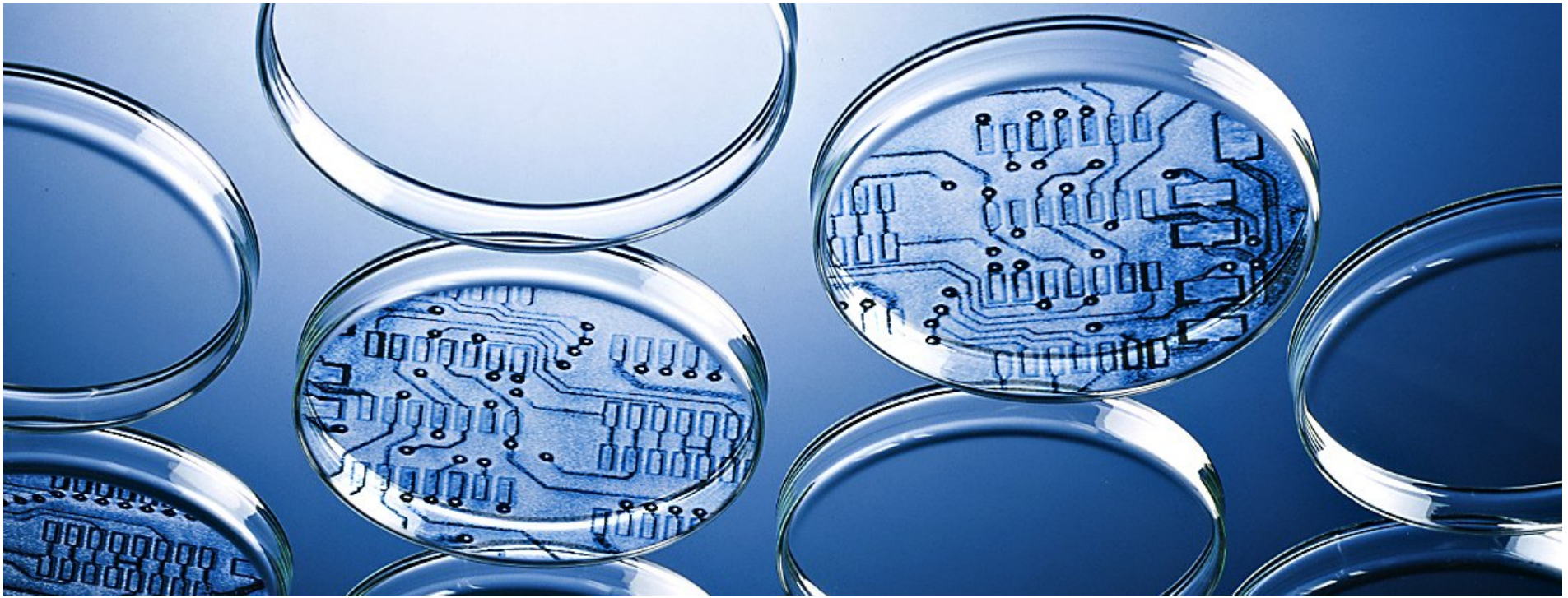
The optimal population size is given by the minimal number of function evaluations, the process robustness, and node infrastructure.

Speedup of Network Verification

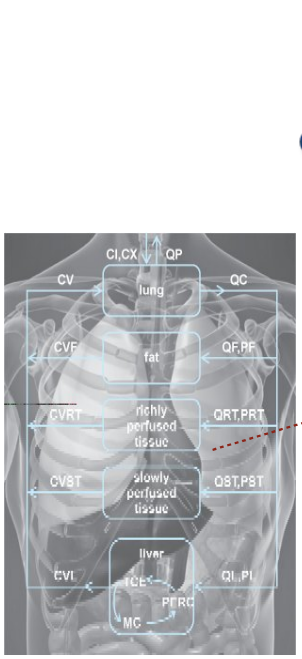


The solution provides excellent speedup, in particular for the industrially most relevant large-scale networks.

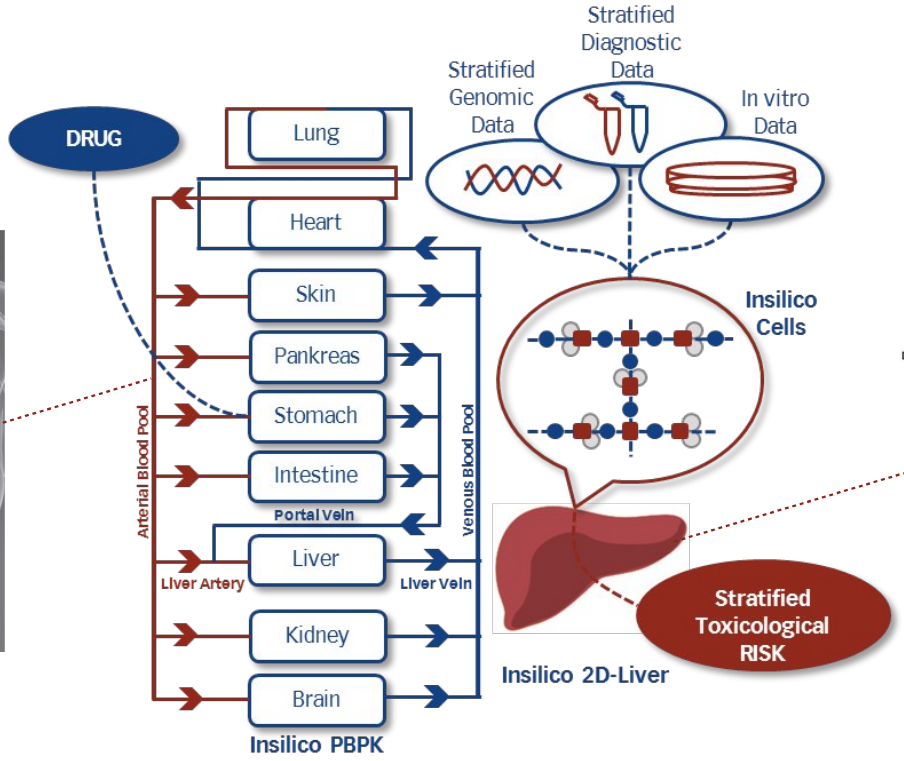
Future Perspectives of HPC at Insilico Biotechnology



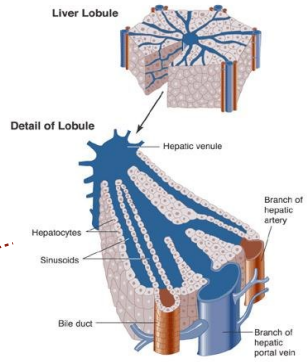
Modeling and Simulation of Genome-Based Whole-Body Systems



Body



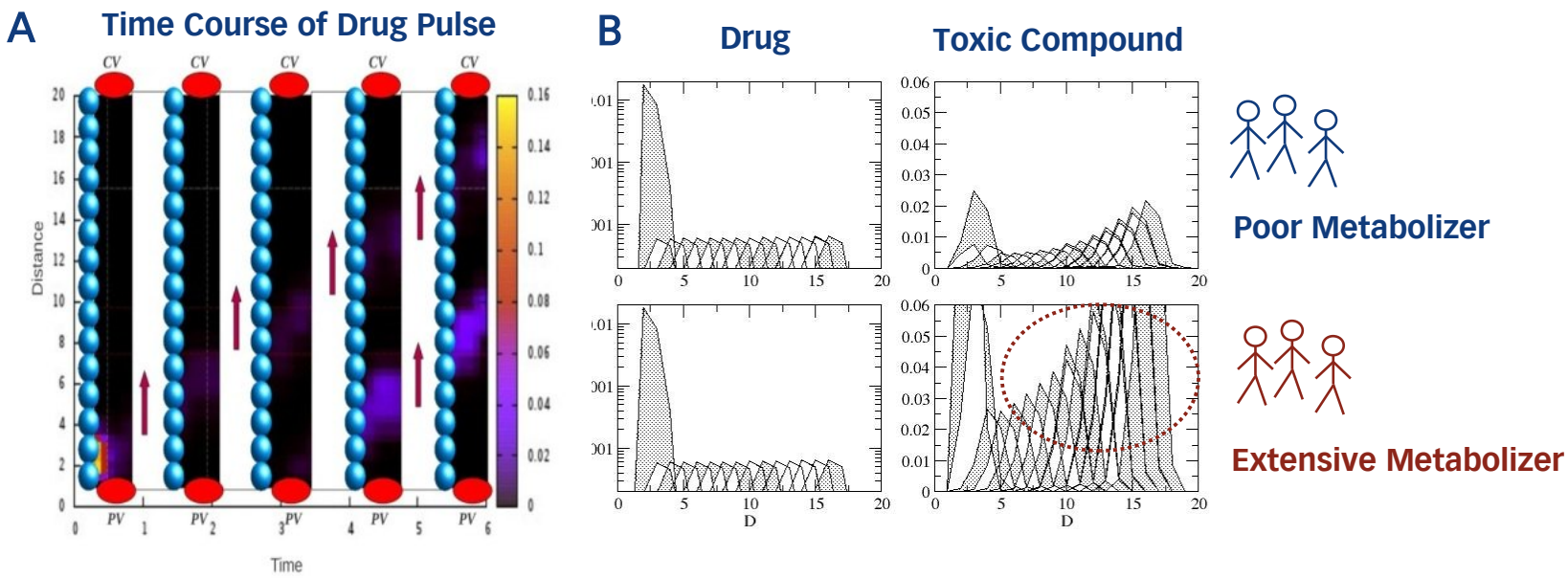
Insilico 2D-Liver



Organ (Liver)

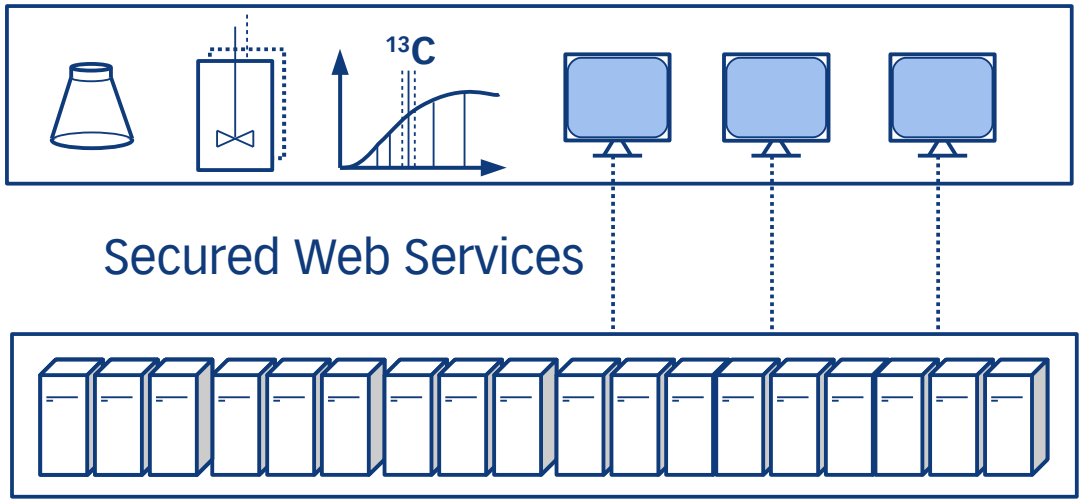
Predicting the genome-based response of drug administration implies a multi-scale approach with time constants ranging from seconds (individual cells) over minutes (organ) to hours (whole body).

Example: Spatial Resolution of Toxic Compounds in the Liver (Sinusoid)



The toxic response as well as the efficacy of the drug Acetaminophen is dependent on genetic disposition of the patient.

Next Generation
Insilico Software Services*



Data Integration
& Hypothesis Testing
at Customer Side (Thin Client)

High-Performance
Computing

Software and high-performance computing are provided online through next generation software services.

*Available October 2011

Summary

- High-performance computing is already an indispensable tool at Insilico Biotechnology
- Quantitative and predictive in silico models are necessary tools in a knowledge-based bio-economy
- High-performance computing will play a key role for speeding up hypothesis testing in the Life Sciences
- High-performance computing might become the computational engine for personalized/stratified medicine

Acknowledgement

Thomas Baumann

Ute Hofmann

Contact

Insilico Biotechnology AG
Meitnerstr. 8
70563 Stuttgart | Germany
T +49 711 460 594-0
info@insilico-biotechnology.com