1. Introduction

During the last 10-15 years, there has been a tremendous boost in computational chemistry and particularly in its biological applications. Biological systems have complex molecular structure and therefore the limitations in computational resources had been previously prohibitive for the meaningful studies of such systems. However, the accelerating speed of the growth of computer power has enabled to start to model biological molecular systems within realistic time frame.

On the other hand, huge amount of experimental data has been collected about the structure of the components of complex biological systems, such as enzymes, receptors, membranes using X-ray, neutronographic, NMR and other spectroscopic measurements. This forms a solid basis for the studies of detailed mechanisms of biomolecular interactions using the computational chemistry approaches.

In the past five years the researchers at the Chair of Molecular Technology have been heavily involved in the method development and application of the *in silico*, i.e. computational approach for the description of the molecular processes and interactions in biological systems as well as to the design of novel biologically active compounds, e.g. potential drug candidates. In general, the research carried out can be divided rationally into following directions:

1. Development of novel computational methods for the discovery of new biologically active compounds and drug candidates.
2. Modeling of chemical, physical, pharmacodynamic and pharmacokinetic properties and toxicity (ADME/Tox) of compounds.
3. Modeling of the targets related to viral diseases and prediction of novel potential antivirals.
4. Modeling of the targets related to neurodegenerative diseases and prediction of novel potential drug candidates.
5. Modeling of biological properties of small peptides, especially the cell-penetrating peptides.

In the following, a concise overview of the research carried out on each of the directions is presented.

2. Development of novel computational methods

The researchers at the Chair have been active in several directions important for the theoretical description of complex biomolecular systems.

The structural characteristics of small compounds are major determinants in the early steps of drug design. Various threshold or range sets (such as Lipinski’s rule-of-five for orally bioavailable compounds and others) are based on molecular descriptors and have been designed and extensively used for *in silico* library screens. An open and active area of research is how known drugs’ and non-drugs’ populations of chemical space locate relative to each other. In
particular, the relative location of drug areas in the chemical space belonging to distinct disease categories is missing, ie. the threshold sets for disease categories. A better knowledge of these thresholds can improve profiling of chemical libraries and optimization of compounds, as well as better target compounds to reduce selectivity issues. All those aspects have been thoroughly studied by analysis of populations of drug and non-drug compounds using principal component analysis [1], probability distribution functions [2], multidimensional logistics regression [3] and Bayesian classifiers [4]. Figure 1 shows that the drugs and non-drugs do occupy different areas on the score plot (in chemical space described by molecular descriptors) with a borderline overlapping area and molecular descriptors group in a clear manner on the loadings plot according to solvation free energy components.

Figure 1. Principal component analysis plots for 1st and 2nd principal component scores (a) and loadings (b) of drug and nondrug populations.

The efficiency indices (EI-s) have been under the workgroup interest for several years [5]. They have been derived from the experimental binding affinities of drug candidates to macromolecules, are ‘two-in-one’ measures and include information on both pharmacodynamics and pharmacokinetics of the drug candidate molecules. Recently extensive analysis of different molecular docking engines and scoring functions was performed to determine the effect of different EI-s on the experimental and predicted free energy of binding [6] and their use in virtual screening of molecular libraries while ranking ligands [2, 5, 6]. Also the accurate calculation of the free energy of interaction of protein-water-ligand systems has been thoroughly studied and computational approach is provided for the systematic and rigorous prediction of the thermodynamic influence of ordered, structural water molecules on ligand modification and optimization in drug design by calculating free energy changes in protein-water-ligand systems [7].

We have first defined the unitary set of appropriate procedures for the selection of descriptors into a valid and predictive QSAR model [8]. The applicability limits of the method have been examined on the huge amount of chemical, physical and biological experimental data [9, 10]. Recently detailed review was written about molecular descriptors derived only from two-dimensional representation of chemical structure, giving insight how they can be interpreted and how they have been used to model toxicity and environmental effects [11].

Investigation of large biochemical systems in silico requires, as a rule, vast computational resources. The group has been pioneering in transferring the molecular modeling methods and algorithms into computer Grid environment. The technology developed has been applied to facilitate the discovery of potential anti-cancer agents [12, 13]. Part of this work has been carried out within consortium lead by Professor Piotr Bala from University of Warsaw to
3. Modeling of chemical, physical, pharmacodynamic and pharmacokinetic properties and toxicity (ADME/Tox) of compounds

The knowledge about the so-called ADME/Tox (absorption, distribution, metabolism, excretion and toxicity) properties is very essential in the design of new chemicals, materials and medical drugs. The computational methods have some important advantages as compared to the experimental measurements of such properties: (i) there is no need to initially synthesize compounds for determining their ADME/Tox properties and (ii) the processing of the property data using numerous different modeling tools is rapid. Consequently, carefully developed and rigorously validated in silico ADME prediction models applied in early screening and evaluation of compounds allow accurate prediction of key information of new drug candidates and to remarkably enhance the productivity in drug discovery.

A wide variety of chemical, physical and ADME/Tox properties of compounds have been modeled at the Chair of molecular technology during the past 5 years:

- Permeability in artificial and natural membranes [16, 17]
- Cytotoxicity on anthraquinones [18]
- Blood-brain barrier penetration [19]
- Solubility [20]
- Ostwald Solubility and partition Coefficients in Ionic Liquids [21]
- Bio-concentration factor [22, 23]
- Toxicity of fish, *Pimephales promelas* [24]
- Toxicity to water flea, *Daphnia magna* [25, 26]
- Toxicity in activated sludge process [27]
- CMC of surfactants [28, 29]
- Polarizability of Polyaromatic Hydrocarbons and Fullerenes [30]
- Antifungal activity of compounds [31]
- Various physico-chemical properties [32, 33, 34, 35]

4. Modeling of targets related to viral diseases

A part of the research in this direction has been carried out in close collaboration with Professor A. Merits’s virology group at the University of Tartu. The primary goal was the discovery of new potent inhibitors for HCV NS3/4A protease using a novel fragment-based QSAR (FQSAR) method and to test experimentally the best compounds. This rapid and efficient in silico screening led to the discovery of seven novel potential inhibitors, two of which are prospective leads for new drug candidates [36].

Another target of large interest in this direction was HIV protease and potential new inhibitors for this enzyme. The robust QSAR models developed by us for this class of compounds enable to predict substituted fullerenes (C_{60}) with significantly higher antiviral activity [37]. In addition, the physical interpretation of the model confirms the specific binding of fullerenes to the HIV protease. Another QSAR model for HIV-1 protease inhibition was derived for six- and seven-membered cyclic ureas [38].

Optimization of ligands to protein targets and off-targets or anti-targets is foreseen to be critical for compounds interacting at several simultaneous sites. For simultaneous screening of
multiple sites a combined approach has been proposed that is using ligand efficiency, cross-docking, and anti-target hits [39]. In silico hits against a set of antitargets (i.e., proteins or nucleic acids that are off-targets from the desired pharmaceutical target objective) are used to predict a simple, visual measure of possible interactions for the ligands, which helps to introduce early safety considerations into the design of compounds before lead optimization. For the study wild-type and drug-resistant Y181C HIV-1 reverse transcriptase proteins are used.

Multi-binding site inhibitors have been also screened for avian influenza H5N1 wild-type neuraminidase and of the oseltamivir-resistant H274Y variant [40]. Inhibitors that target simultaneously several adjacent binding sites of the open conformation of the virus protein were ranked using ligand efficiency indexes suggests potential inhibitors that mimic a polysaccharide and β-lactam structures.

5. Prediction of cell-penetrating peptides

Since the discovery of cell-penetrating peptides (CPPs) in 1990s, these have attracted substantial attention as promising drug delivery candidates allowing transport of pharmacologically active compounds, such as peptides, proteins, cytotoxic agents, antisense oligonucleotides, plasmid DNA, siRNA, nanoparticles, medical imaging contrast agents and other substances into mammalian cells. In collaboration with Professor Ü. Langel’s groups at the Universities of Stockholm and Tartu, we were able to develop successful artificial intelligence employing models for the prediction of new short cell-penetrating peptides [41, 42].

6. Novel drug candidates for neurodegenerative diseases

Dopamine is a crucial neurotransmitter responsible for functioning and maintenance of the nervous system that has also been implicated in a number of diseases including schizophrenia, Parkinson's disease and drug addiction. Therefore, molecules modulating dopamine receptors activity are vastly important for understanding the nervous system functioning and for the treatment of neurological diseases. We have developed novel computational models that efficiently predict binding affinity of the existing small molecule dopamine analogs to dopamine receptor [43].

The discovery of neurotrophic factor mimetics as GFRα receptor agonists (cf. Fig. 2) in close collaboration with Professor M. Saarma from the Institute of Biotechnology of University of Helsinki and their application as potential drugs against different neurodegenerative diseases such as the Parkinson’s disease, amyotrophic lateral sclerosis and neuropathic opens a new direction in the therapy of those diseases [44].

A very important property in the developing of the drug candidates for the neurodegenerative diseases is the blood-brain barrier (BBB). Research carried out by us indicates that both the multilinear regression and ANN models exhibit reasonable BBB prediction capabilities [19].


Understanding interactions between proteins has huge therapeutic potential and computational methods provide means to get insight how changes in molecular structure influence macromolecular interactions. Molecular dynamic (MD) methods are particularly useful for the study of structural changes and explaining resulting interactions in bio-molecules. In collaboration
with Professor Pärt Peterson from molecular pathology group at University of Tartu, we have been using molecular dynamics calculations in providing computational insight and complementing experimental work how gene autoimmune regulator (AIRE) interacts with histone H3 through its first plant homeodomain (PHD) finger (AIRE–PHD1) and preferentially binds to non-methylated H3K4 (H3K4me0) [45]. In collaboration with scientists from chair of molecular biology lead by Professor Jaanus Remme we have studied influence of structural changes to ribosomal RNA using MD methods, particularly the influence of single-nucleotide substitutions [46] and pseudouridines in several locations were analysed [47].

![Figure 2. Blind docking of an agonist molecule (green) to GFRα1 receptor.](image)

**8. Defended PhD dissertations**

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Year</th>
<th>Supervisor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalev Takkis</td>
<td>Virtual screening of chemical databases for bioactive molecules</td>
<td>2012</td>
<td>Sulev Sild</td>
</tr>
<tr>
<td>Dana Martin</td>
<td>The QSPR/QSAR approach for the prediction of properties of fullerene derivatives</td>
<td>2011</td>
<td>Mati Karelson</td>
</tr>
<tr>
<td>Indrek Tulp</td>
<td>Multivariate analysis of chemical and biological properties</td>
<td>2010</td>
<td>Uko Maran</td>
</tr>
<tr>
<td>Svetoslav Slavov</td>
<td>Biomedical applications of the QSAR approach</td>
<td>2007</td>
<td>Mati Karelson</td>
</tr>
</tbody>
</table>

**8. Research projects**

Funding resources and research projects listed below start of end during the period of 2007 – 2012:

European Union Framework Programs:
- Understanding Nano-Materials From the Quantum Perspective, Site PI Mati Karelson, EU FP6, 2004-2007
- Grid services based environment to enable innovative research, Site PI Uko Maran, EU FP6, 2006-2009

Estonian Target Funding:
- Molecular design of biotechnological and macromolecular systems, PI Uko Maran, Ministry of Education and Research, 2004-2008
- Application of computational chemistry in the study of complex molecular systems of large biomedical and environmental importance, PI Uko Maran, Ministry of Education and Research, 2009-2014
- Small infrastructure upgrade grant for target funding project: Application of computational chemistry in the study of complex molecular systems of large biomedical and environmental importance, PI Uko Maran, Ministry of Education and Research, 2011-2012.

Estonian Science Foundation:
- The classification of toxic chemicals and estimation of their toxicity based on chemical structure, PI Uko Maran, ETF, 2004-2007
- More effective application of structure-property relationship methods for complex molecular modeling and design problems, PI Sulev Sild, ETF, 2007-2010
- Methods for the validation and definition of applicability domain of quantitative structure activity relationships, their development and application, PI Uko Maran, ETF, 2008-2011
- Post-doctoral fellowship grant for target funding project: Molecular design of biotechnological and macromolecular systems, PI Uko Maran, ETF, 2007-2008

Estonian Biotechnology Programme:
- Development of Trk antagonists as drug candidates for the treatment of neuropathic pain, Site PI Mati Karelson 2012-2015

Industrial contracts:
- Synthesis of 4-aminated 2-(2-naphthyl)quinolines, PI Mati Karelson, 2008
- Synthesis of magnetically active compounds, Estiko AS, PI Mati Karelson, 2009
- Environmental risks of nanostructured materials, Jaan Leis, Port of Tallinn, 2010
- Design of Neublastin Agonists, Chemedest Ltd., PI Mati Karelson, 2011
- Design and synthesis of N-alkylsulphonamides, Genecode Ltd., PI Mati Karelson, 2012

Other
- Improving the international competitiveness of education and research & development in modern molecular technology, PI Uko Maran, Foundation Innove, 2006-2008


Garcia-Sosa, Alfonso T.; Maran, Uko (2012). Drugs, Non-Drugs, and Disease Category Specificity: Organ Effects by Ligand Pharmacology. SAR and QSAR in Environmental Research, in press.


Karelson, M; Dobchev, D; Tamm, T; Tulp, I; Jänes, J; Tämm, K; Lomaka, A; Savchenko, D; Karelson, G (2008). Correlation of Blood-Brain Penetration and Human Serum Albumin Binding with Theoretical Descriptors. Arkivoc, xvi, 38 - 60.


22 Piir, G.; Sild, S.; Roncaglioni, A; Benfenati, E.; Maran, U. (2010). QSAR model for the prediction of bio-concentration factor using aqueous solubility and descriptors considering various electronic effects. SAR and QSAR in Environmental Research, 21(7-8), 711 - 729.


