Kazan Summer School on Chemoinformatics

PROGRAM
&
ABSTRACTS

KSSCI KAZAN 2013

August 26-29, 2013
KAZAN, RUSSIA

KAZAN (VOLGA REGION) FEDERAL UNIVERSITY
KFU
1804

GENERAL INFORMATION

ORGANIZERS

Kazan (Volga region) Federal University
Russian Foundation for Basic Research
Russian Section of the International Cheminformatics and QSAR Society
D.I. Mendeleev Chemical Society of Republic of Tatarstan

SPONSOR

Chemical Abstract Service (https://www.cas.org/)

ORGANIZING COMMITTEE

Chairmen of the committee:
Prof. Alexandre Varnek (Strasbourg, France)
Prof. Igor S. Antipin (Kazan, Russia)
Prof. Vladimir V. Poroikov (Moscow, Russia)

Scientific Secretary: **T.I. Madzhidov** (Kazan, Russia)

Members of the Committee:

N.I. Ivanova	M.A. Kazymova	V.A. Burilov
M.D. Misarov	P.O. Sidorov	A.R. Kurbangalieva
D.R. Chubukaeva	A.V. Bodrov	R.I. Nougmanov
T.R. Gimadiev	A.S. Petrovsky	

SCIENTIFIC PROGRAMME

The programme of the First Kazan Summer School on Chemoinformatics includes 10 lectures, 2 plenary and 6 oral reports, 4 tutorials and 24 poster presentations.

OFFICIAL LANGUAGE

The official School language is English. No translation will be provided.

OFFICE OF THE ORGANIZING COMMITTEE

The office of the organizing committee is located at **A.M. Butlerov Chemical Institute** building, **Auditorium No. 218**. Participants have the ability to use the telephone and the Internet there.

VENUE

The event will be held mainly in the **A.M. Butlerov Chemical Institute** building (Lobachevskogo St. 1). Lectures, plenary and oral presentations will take placeinthe **Hall No. 319**. Tutorials will be held in **Hall No. 401**. Only a few stand-alone computers will be

installed for the usage by participants. However Wi-Fi connection for personal laptops of the participants is provided.

SOCIAL EVENTS

Welcome Party

All the participants are invited to a welcome party that will be held **August 26 in the second floor of the Cafe of the Institute of Physics** (Kremlyovskaya St. 16A) **at 18.00**. The participation is free.

The meeting outside A.M. Butlerov Chemical Institute building at 18.00.

Excursion

The attendees from outside Kazan are welcome to take part in the excursion program. **The excursion will begin August 25 at 14.00**. Visit to Kazan Kremlin, Zilant Monastery and Raifa Bogoroditsky monastery are planned. The tour around the Kazan Kremlin includes a walk in the Kremlin, where you will see the Savior's Tower, leaning Syuyumbike Tower, will visit the Kull Sharif Mosque and the Annunciation Cathedral. The duration of the excursion will be 5-5.5 hours. During the excursion around the religious building dress code should be observed. Please note that no food (only water) will be provided during the excursion program. The excursion language will be English.

The meeting outside A.M. Butlerov Chemical Institute building at 13.30.

PROGRAM
of Kazan Summer School on Chemoinformatics

August 25, 2013

Arrival day

12:00-15:00	**Registration**	
15:00-19:00	**Excursion Program**	

August 26, 2013

8:30-9:00	**Registration**	
9:00-9:15	**Opening ceremony**	
	Lecture Session 1 **Chairman – V. Poroikov**	
9:15-10:15	A. Varnek (Strasbourg, France)	Chemoinformatics: Basic Concepts and Areas of Application
10:15-11:15	T. Madzhidov (Kazan, Russia)	Chemical Databases: Encoding, Storage and Search of Chemical Structures
11:15-11:45	Coffee-break	
11:45-12:30	V.-P. Hyttinen (CAS, Finland)	Presentation of SciFinder
12:30-14:00	**Lunch**	
	Oral Session 1 **Chairman – I. Tetko**	
14:00-14:20	V.S. Abrukov (Cheboksary, Russia)	Artificial neural networks for creation of knowledge bases in scientific and applied research
14:20-14:40	N.I. Baranova (Saint-Peterburg, Russia)	QSAR modeling of calculation IC50 for Xa coagulation factor
14:40-15:00	O.V. Galzitskaya (Moscow, Russia)	How to Determine the Size of the Nucleus of Protofibrils from the Concentration Dependence of the Lag-Time Of Aggregation? Experimental Application: Insulin and Lys-Pro Insulin
15:00-15:30	P. Polishchuk (Odessa, Ukraine)	Development of "non-classical" antagonists of fibrinogen receptors - promising anti-platelet agents
15:30-16:00	Coffee-break	
	Tutorial Session 1	
16:00-18:00	G. Marcou, D. Horvath (Strasbourg, France)	Tutorial with ChemAxon
18:00-20:00	**Welcome party**	

August 27, 2013

	Lecture Session 2 **Chairman – T. Langer**	
9:00-10:00	I. Baskin (Moscow, Russia)	Obtaining, Validation and Application of SAR/QSAR Models
10:00-11:00	A. Tropsha	SAR/QSAR Modelling: State of the Art

	(Chapel Hill, USA)	
11:00-11:30	Coffee-break	
11:30-12:30	I. Tetko (Munich, Germany)	ADMET Predictions
12:30-14:00	**Lunch**	
	Tutorial Session 2	
14:00-16:00	I. Tetko (Munich, Germany)	Tutorial with QSAR modeling
16:00-18:00	**Poster Session**	

August 28, 2013

	Lecture Session 3 **Chairman – A. Tropsha**	
9:00-10:00	D. Horvath (Strasbourg, France)	Conformational Sampling
10:00-11:00	T. Langer (Strasbourg, France)	Pharmacophores and Its Applications
11:00-11:30	Coffee-break	
11:30-12:30	G. Marcou (Strasbourg, France)	Molecular Docking Methods
12:30-14:00	**Lunch**	
	Tutorial Session 3	
14:00-16:00	S. Bryant (Vienna, Austria)	Tutorial with LigandScout
16:00-16:30	Coffee-break	
16:30-18:30	G. Marcou, D. Horvath (Strasbourg, France)	Tutorial with LeadIt

August 29, 2013

	Oral Session 2 **Chairman – K. Balakin**	
9:00-9:20	S.V. Luschekina (Moscow, Russia)	Refining Molecular Modeling Techniques: 2-Amino-5-Halomethyl-Thiazolines Esterase Profile Case Study
9:20-9:40	E. Muratov (Odessa, Ukraine)	Obscurum per obscurius: computer-aided design of novel antivirals using Simplex approach
9:40-10:00	L.A. Uroshlev (Moscow, Russia)	Algorithm for prediction ions in protein structures
10:00-10:30	V. Solovev (Moscow, Russia)	Substructural Molecular Fragments in Consensus QSPR Modeling
10:30-11:00	Coffee-break	
	Lecture Session 4 **Chairman – A. Varnek**	
11:00-12:00	K. Balakin (Moscow, Russia)	Computational Mapping Tools for Drug Discovery
12:00-13:00	V. Poroikov (Moscow, Russia)	Drug Design & Discovery in Academia
13:00-13:30	**Winner award ceremony, Closure of the School**	

Oral Reports

Artificial neural networks for creation of knowledge bases in scientific and applied research

V. Abrukov

Department of Applied Physics and Nanotechnology, Chuvash State University, Moskovsky Avenue, 15, Cheboksary, Russia
abrukov@yandex.ru

Modelling is still emerging to become the key for industrial progress in areas ranging from materials and chemical research to the development of novel and improved applications.

Let's take as an example the area of nanotechnology. Currently a lot of experimental data on properties and characteristics of various nanomaterials are obtained in all of the world. The question is how we can summarize it and present in the form of common models allowing description the characteristics of previously studied nanomaterials?

It is obvious that the characteristics of nanomaterials related to the composition of nanomaterials and type of components, manufacturing technology, the shape and size. The question is how we can generalize these links as computational models (CM) that allow determining the characteristics of the nanomaterials without carrying out additional experiments (direct task)?

Even more important question is it possible to predict what should be the nanomaterial (structure, components, and dimensions) and what technology should be used with to provide the required properties and characteristics of nanomaterials (inverse task)?

We present an example of the results of application of artificial neural networks (ANN) to create such CM. They are based on experimental results for the characteristics of nanofilms of linear-chain carbon (LCC) (carbene) with embedded into LCC various atoms (LCCA). For the first time LCCA were manufactured in the Chuvash State University, using unique technology protected by a patent, and using a variety of know-how. The direction of work can be of great interest for active and passive elements of solid-state electronics, photovoltaic elements, sensors, medical applications, etc.

To date we have developed two CM (ANN-models) that allow us to reveal all dependences between variables, to generalize them and to calculate the physical-electrical and optical properties of LCCA in dependence on amount of kind of atoms (one or two kinds) embedded in a LCCA, kind of atoms (number and group of atom in accordance with the Mendeleev's periodic table), and the thickness of the LCC MNA. It allows us to solve also the inverse task.

An analysis of results obtained has depicted that:

1. The CM correctly reveals all dependences of the current and transmission coefficient on other parameters and it is the good tool for generalization and prediction of connection between variables.
2. The CM instantly calculates a value of the necessary characteristic and it is the fast engineering calculator specialized to LCC MNA.
3. The CM easy gets any characteristics of a hypothetical sort of LCCA and it is the most cheap way for receiving of "new" "experimental" results without an experiment.
4. The CM are the Knowledge Base (KB) of LCCA as well as prototype of a future KB of nanomaterials world.

Other our examples of the creation of CM deal with the inverse and direct problems of optical diagnostics, combustion research, operation of solar electrical station. We have also the KB of family relations in Russia as well as we execute a project deals with a creation of KB in the area of educational system. Several examples are presented in WWW: http://www.chuvsu.ru/2008/proekt.html and http://mfi.chuvsu.ru/opros/ .

QSAR modeling of calculation IC50 for Xa coagulation factor

N.I. Baranova[1,2]
P.A. Alexeeva[2]
I.A. Siginevich[2]
E.V. Fedorova[3]

[1] Department of Inorganic Chemistry, [2] Laboratoty of pharmacological reseach, [3] Department of IT. Saint Petersburg State Chemical Pharmaceutical Academy, Prof. Popova St., 14, Saint Petersburg, Russia
nadezhda.baranova@pharminnotech.com

Selective inhibitors of specific coagulation factors represent a novel class of antithrombotic agents, designed to overcome the limitations of traditional anticoagulants, such as warfarin. Available clinical studies demonstrate that the most promising new anticoagulants are those selectively targeting factor Xa and thrombin.

The goal of our study is in developing a model of QSAR-model to calculate IC50 values for coagulation factor Xa and to confirm its stability in *in vitro* study.

53 compounds were selected to build the model according to the results of literature search. The theoretical model was constructed using GUSAR program which uses binary fragment descriptors to describe the chemical structure.

Consensus pattern parameters: V = 9; F = 23.077; r^2 = 0.896; SD = 0.774; q^2 = 0.838. Values of the parameters let us come to the conclusion about the stability of the model created. To confirm stability of the QSAR-model and the accuracy of prognosis two compounds were selected from class of 5-aryl-1,3,4-oxadiazole-2-thiol: 5-(2,4-dichlorophenyl)1,3,4-oxadiazole-2-thiol (**1**) (the most active compound according to the created model) and 5(4-nitrophenyl)-1,3,4-oxadiazole-2-thiol (**2**) (inactive compound). pIC50 values calculated for compound 1 and compound 2 were 0.01 and 0.92 respectively.

Anticoagulant activity of compound **1** and compound **2** was evaluated *in vitro*. A fixed amount of compound **1** and compound **2** solutions in DMSO were added to preincubated samples of platelet poor plasma collected from healthy intact rats and basic coagulation parameters – prothrombin time (PT), activated partial thromboplastin time (APTT) and thrombin time (TT) – were measured using 2-channel coagulation analyzer HumaClot Duo Plus (Human GmbH, Germany) and Hemostat reagents (Human GmbH, Germany)

Compound **1** one was tested at the concentration range from 0.0008 to 0.5 mmol/ml. A linear increase in PT, APTT and TT was detected in the concentration range from 0.1 to 0.5 mmol/ml (threefold increase compared to control plasma at 0.5 mmol/ml).

Compound **2** was tested at the concentration range between 0.00032 and 0.2 mmol/ml and did not demonstrate any significant changes in values of coagulation parameters in comparison with control plasma.

Compound **1** has showed direct anticoagulant activity in vitro, while compound **2** did not affect the coagulation cascade in the same conditions. These results were consistent with the data provided by QSAR-model. Consequently, obtained model can be used in screening compounds for finding potential new anticoagulants.

Funding was provided by Saint Petersburg State Chemical Pharmaceutical Academy.

O.V. Galzitskaya
N.V. Dovidchenko
M.Yu. Suvorina
O.M. Selivanova
I.A. Eliseeva
A.V. Finkelstein
V.V. Schmatchenko
A.K. Surin

How to determine the size of the nucleus of protofibrils? Experimental application: insulin and LysPro insulin

Institute of Protein Research, Russian Academy of Sciences, 4 Institutskaya St., Pushchino, Moscow Region, 142290, Russia
ogalzit@vega.protres.ru

The question about the size of nuclei of protofibrils formed by different proteins and peptides is yet open. By the nucleation mechanism, the formation of protofibrils begins from the thermodynamic unfavorable steps resulting in the formation of a critical nucleus consisting of n monomers. The kinetic model of the process of formation of amyloid fibrils is suggested in our work allowing us to calculate the size of the nucleus using kinetic data. In addition to the stage of nucleation, the given model includes both a linear growth of protofibrils (proceeding only at the cost of attaching of monomers to the ends) and an exponential growth of protofibrils at the cost of branching and fragmentation. Theoretically, only the exponential growth is compatible with the existence of a lag-period in the fibril formation kinetics. Insulin is a commonly used protein for studies of amyloidogenesis. There are a few insulin analogues with different pharmacokinetic characteristics, in particular the onset and duration of action. As the duration of action may be connected with the duration of the lag-phase, the challenge is to consider the process of amyloid formation for different analogs of insulin. One of them is LysPro insulin. The behavior of LysPro insulin in the process of amyloid formation has not been studied in detail yet. To quantitatively investigate the differences between the two samples in the aggregation reaction and estimate the difference in the lag-time, we used thioflavin T fluorescence assay, electron microscopy, X-ray diffraction methods, and theoretical modeling. Kinetic experimental data for both insulin and LysPro insulin samples demonstrated the increasing of the lag-time for LysPro insulin at low concentrations of monomers, particularly at 2 and 4 mg/ml, which corresponds to the pharmaceutical concentration. The obtained analytical solution and computer modeling allow us to determine the size of the nucleus from the experimentally obtained concentration dependences of the relationship between the lag-time and the time of growth of amyloid fibrils. In the case of both insulin and LysPro insulin, this relationship is independent of the protein concentration. According to the developed theory, this means that the size of the nucleus corresponds to one monomer in both insulin and LysPro insulin.

This study was supported by the Russian Foundation for Basic Research and by the Russian Academy of Sciences programs "Molecular and Cell Biology" and "Fundamental Sciences to Medicine".

S.V. Lushchekina[1]
G.F. Makhaeva[2]
N.P. Boltneva[2]
O.G. Serebryakova[2]
A.N. Proshin[2]
I.V. Serkov[2]
S.O. Bachurin[2]

Refining molecular modeling techniques: 2-amino-5-halomethyl-thiazolines esterase profile case study

[1] Emanuel Institute of Biochemical Physics, RAS, 119334, 4, Kosygina St., Moscow, Russia
[2] Institute of Physiologically Active Compounds, RAS, 142432, 1, Severny proezd, Chernogolovka, Russia
sofya.lushchekina@gmail.com

Molecular docking methods are widely used in computer-aided drug design for prediction of protein-ligand interactions and virtual screening for new potent drugs. For accurate estimation of protein ligand-binding it is crucial to use a reliable scheme of assignment of partial atomic charges. Conventionally charges are assigned according to a force-field parameterization, however it causes inaccuracies for non-classical interactions such as π-cation interactions, π-stacking, hydrogen bonding to π-systems etc. An accurate charge distribution could be obtained with the high-level *ab initio* quantum mechanics calculations, however for proteins and large ligand databases such calculations are too time-consuming. Elaboration of semi-empirical quantum chemistry methods might improve accuracy of molecular docking at the reasonable computational cost.

Determination of compounds esterase profile is a new approach was developed and is being widely applied in IPAC RAS. It implies a comparative estimation of compounds inhibitory activity towards several esterase enzymes: acetylcholinesterase (AChE), butyrylcholinesterase (BChE), carboxylesterase (CaE) and neuropathy target esterase [1-3]. This approach allows us to get a more complete view on biological effects of the compound and thereby evaluate its therapeutic potency and possible side-effects. We synthesized derivatives of 2-amino-thiazoline and demonstrated that they are able to inhibit therapeutically important serine esterases, BChE and CaE, showing a low inhibitory activity against AChE [4].

In the present study we calculated binding energies of 67 *N*-substituted 2-amino-5-halomethyl-thiazolines with three target enzymes: AChE, BChE and CaE using Autodock 4.2 program. For proteins and ligands standard Gasteiger charges were used and partial atomic charges were calculated semi-empirically using PM7 method implemented in MOPAC2012 program. It was demonstrated that using of the atomic charges derived from quantum-mechanical calculations helps to improve significantly the results of molecular docking. Different protonation states of the ligands were considered and several popular programs for pKa estimation were analyzed. Also, geometry of the enzymes was partially optimized by means of semi-empirical quantum chemistry methods for improvement of the crystallographic structures traditionally used for molecular modeling. The results explain the observed inhibitor selectivity of *N*-substituted 2-amino-5-halomethyl-thiazolines and allow us to make recommendations for their directed modification.

[1] Makhaeva G. et al. *Doklady Biochem Biophys*, 2008, **423**, 352-357.
[2] Makhaeva G. et al. *SAR QSAR Env. Res.*, 2012, **23**, 627-627.
[3] Makhaeva G. et al. *Chem. Biol. Interact*, 2013, **203**, 231-237.
[4] Boltneva N. et al. *Doklady Biochem Biophys*, 2013 (In press).

This work was supported by RAS program "Medicinal chemistry" and RFBR grants #11-03-00863 and 12-03-31039-mol and the Dynasty Foundation Fellowship (Sofya.Lushchekina).

E. Muratov[1,2]
V. Kuz'min[1]
E. Varlamova[1]
A. Artemenko[1]
V. Lozitsky[3]
A. Fedchuk[3]
T. Gridina[3]
L. Mudrik[3]

Obscurum per obscurius: computer-aided design of novel antivirals using simplex approach

[1] A.V. Bogatsky Physical-Chemical Institute NAS of Ukraine, Odessa, Ukraine
[2] University of North Carolina, Chapel Hill, USA
[3] I.I. Mechnikov Ukrainian Anti-Plague Research Institute, Odessa, Ukraine

Viruses are one of the important causes of human diseases, such as hepatitis, AIDS, or influenza. Depending on the strain and the person's state of health, they can infect almost any type of body tissue, from the brain to the skin and hit almost everyone. The burden of only influenza in the USA is currently estimated to be 25-50 million cases per year, leading to 150 thousand hospitalizations and up to 30-50 thousand deaths. Here we report on the application of the Simplex representation of molecular structure (SiRMS) for QSAR modeling in antiviral research. After introducing the field of cheminformatics, we present the SiRMS approach: SiRMS represents every molecule as a system of different simplexes (tetratomic fragments with fixed composition, structure, chirality, and symmetry). The main advantages of SiRMS are: *(i)* consideration of the different physical-chemical properties of atoms, *(ii)* high adequacy and good interpretability of developed models, and *(iii)* clear workflow for computer-aided molecular design. The SiRMS approach is implemented within the "HiT QSAR" software, which is available from the authors by request. The reliability of developed QSAR models as predictive virtual screening tools and their ability to serve as the guide for targeted drug design was validated by synthetic and biological experiments. We discuss several applications of SiRMS in antiviral research and future directions of extending Simplex approach for QSAR analysis of mixtures.

Algorithm for prediction ions in protein structures

L. Uroshlev[1,2]
S. Rakhmanov
I. Kulakovskiy
V. Makeev[1,2,3]

[1] Engelhardt Institute of Molecular Biology, Laboratory of Biophysics, 119334 Vavilova St. 32, Moscow, Russia
[2] Vavilov Institute Of General Genetic, Laboratory of System Biology and Computative Genetics, Gubkina St. 3, Moscow 119991, Russia
[3] Moscow Institute of Physics and Technology, 9 Institutskiy lane, Dolgoprudny City, Moscow Region, 141700, Russia
leoniduroshlev@gmail.com

Ions play different and important roles in folding, functions and aggregation of proteins and their complexes. Disruption of an ion binding site in a protein molecule may cause protein function loss, which in turn can result in disease development [1]. Thus, a problem of prediction of ion binding sites specificity and its location in the given protein structure or structure model is important.

We present an effective algorithm for prediction binding sites for ubiquitous ions of zinc, magnesium, and calcium. The algorithm employs the empirical potentials method, which is based on the Boltzmann equation

$$E = -kT\ln\frac{f_{obs}(d)}{f_{\exp}(d)}$$

where is E the pseudoenergy of binding, k is the Boltzmann constant, T is the temperature, $f_{obs}(d)$ and $f_{\exp}(d)$ are the observed and the expected frequencies of contacts between the atom (the query ion in our case) and the element of a structure from the learning set (the protein atom assigned to a particular amino acid). The expected frequency can be directly estimated from the learning set with using the MCRS method [3]. PDB database, source of information of protein structures with bound ions, gives a reliable statistics necessary to construct a high quality potential for all three ion types.

We compared the implementation of our algorithm (PIONCA) with two other ligand prediction programs. The results are presented in the table. It is important that in all cases the correct type of specifically bound ion was correctly identified.

ID(PDB)/Type of ion	PIONCA RMSD (A)	FINDSITE RMSD(A)	COFACTOR RMSD(A)	ID(PDB)/Type of ion	PIONCA RMSD (A)	FINDSITE RMSD(A)	COFACTOR RMSD(A)
2OEO/CA	0.53	0.6	0.89	1AUX/CA	0.9	1.61	0.57
1FA5/ZN	0.13	5.5	Not detect	1IG5/MG	0.07	1.57	1.69
1FKQ/CA	0.21	0.5	0.28	2CHE/MG	0.12	0.53	1.25
1AST/ZN	0.06	0.47	0.13	1LBU/ZN	0.13	0.2	Not detect
2FBX/MG	0.41	1.76	1.14	1B66/ZN	0.12	1.58	0.07
1BMO/CA	0.68	2.70	1.94	1BJ3/CA	0.39	0.65	1.13

[1] A.I. Bush and R.E. Tanzi Therapeutics for Alzheimer's Disease Based on the Metal Hypothesis. (2008) *Neurotherapeutics* 5:421-432.
[2] S. Ньls, T. Нцgen, N. Vassallo, K.M. Danzer, B. Hengerer, A. Giese and J. Herms. AMPA-receptor-mediated excitatory synaptic transmission is enhanced by iron-induced α-synuclein oligomers (2011). *Journal of Neurochemistry* 117:868-878.
[3] Sergei Rahmanov, I. Kulakovskiy, L. Uroshlev L, Makeev V. (2010) Empirical potentials for ion binding in proteins. *J Bioinform Comput Biol*. Jun; 8(3):427-35.

Substructural molecular fragments in consensus QSPR modeling

V. Solov'ev[1]
A. Varnek[2]

[1] Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Leninskiy avenue, 31/4, 119071 Moscow, Russia
[2] Laboratoire d'Infochimie, UMR 7177 CNRS, Université de Strasbourg, 4, rue B. Pascal, Strasbourg, 67000, France
solovev-vp@mail.ru

Here we report about Substructural Molecular Fragments (SMF) as descriptors for QSPR modeling, their application for ensemble QSPR modeling with combined applicability domain approach, and some tools for compound design and property optimization on the basis of SMF and their contributions in individual models.

QSPR modeling with SMF descriptors is performed using computer programs and tools, which were developed for Windows operating systems: the ISIDA/QSPR program for Ensemble Multiple Linear Regression Analysis for QSPR modeling, data manager EdiSDF, which is editor of MDL Structure Data Files and 2D sketcher EdChemS, which is editor of 2D chemical structures.

SMF are subgraphs of molecular graphs [1]. The counts of SMF of the graphs are descriptor values. Two principal classes of SMF are generated: sequences or topological paths and augmented atoms or atoms with nearest neighbors. The minimal and maximal lengths of sequences are varied from 2 to 15. They represent shortest paths or all paths, those with explicit representation of atoms and bonds or terminal groups as paths defined by length and explicit identification of terminal atoms and bonds. In SMF, atom has different attributes: it can be presented by element symbol only, or atomic hybridization can be taken into account, or Bensons' scheme can be used, where some atomic groups are presented as extended atoms, or atom can be labeled for indication of reaction center and a property of selected atom. Similar molecular fragments belonging to two different regents are considered as different. Hereby, labeled units can be fragments and atoms. Chemical bonds have also several attributes: types for compounds, complexes and reactions; order (single, double, triple and aromatic etc.); cyclic or acyclic for taking into account of topology of 2D structure.

The ISIDA/QSPR program with combined forward and backward stepwise variable selection techniques can generate thousands of linear relationships (individual models) between dependent variable (here, property, activity) and independent variables (here, SMF descriptors) using hundreds of SMF types and several forward variable selection techniques. The property is reliable predicted as an arithmetic mean of values obtained by individual models excluding those leading to outlying values and being outside applicability domains of individual models.

SMF and their contributions in the QSPR models are convenient building blocks for compound design and property optimization. For this aim, ISIDA Predictor (*http://infochim.u-strasbg.fr/cgi-bin/predictor.cgi*), generator of virtual combinatorial libraries, interactive compound designer [2], and a tool for coloring of atoms of 2D structure according fragment contributions of individual models have been elaborated.

The programs have been developed in the framework of the ISIDA project. It is collaborative project between the Laboratory of Chemoinformatics under the direction of Prof. A. Varnek and the Laboratory of New Physico-Chemical Problems under the direction of Academician Aslan Tsivadze. The programs are available for the end users on the web sites via the internet: http://infochim.u-strasbg.fr/spip.php?rubrique53 or http://vpsolovev.ru/programs/.

[1] Varnek A., Fourches D., Hoonakker F., Solov'ev V. *J. Comp.-Aided Mol. Design*, 2005, 19, 693-703.
[2] Solov'ev V., Sukhno I., Buzko V., Polushin A., Marcou G., Tsivadze A., Varnek A. *J. Incl. Phenom. Macrocycl. Chem.*, 2012, 72, 309-321.

VS thanks Organizers of Kazan Summer School on Chemoinformatics for support.

Poster Presentations

Computer simulation of acetylcholinesterase enzyme interaction with novel inhibitors

R. Ayupov
N. Akberova
D. Tarasov

Department of Biochemistry, Institute of Basic Medicine and Biology, Kazan Federal University, 18 Kremlyovskaya St., Kazan, Republic of Tatarstan, Russia
aurusta@mail.ru

In silico methods are often used in the development of specific inhibitors for biological molecules. In this paper, we study the interaction of pyridoxine derivatives with acetylcholinesterase (AChE). Structural properties of ligand allow us to assume that covalent bond is formed in interaction of the enzyme with ligand. This assumption is based on the similarity of radical groups of studied ligands and already known covalent inhibitors of AChE. In AutoDock experiments the similarity in the position and form of pyridoxine derivatives and covalent inhibitors have also been shown [1]. However, this is not sufficient to prove that pyridoxine derivatives are covalent inhibitors. To obtain more accurate data the molecular modeling using software Amber 99 and NAMD 2.8 was carried out. As a result it was shown that pyridoxine derivatives are capable to form covalent bonds, because the distance between the radical group of pyridoxine derivative and the amino acid residue of the enzyme has dropped to 3.5 angstrom, at which a covalent bond is able to form. This is also supported by spatial position of the ligand in the active cavity of the enzyme.

[1] Ayupov R.Kh., Akberova N.I., Tarasov D.S. Docking of Pyridoxine Derivatives in the Active Site of Cholinesterases. // *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennye Nauki.* 2011. Vol. 153, No.3. P. 107-118.

I. Borodaev
V. Zheltukhin

Simulation of scattering argentum nanoparticles by low pressure RF plasmas

Kazan Federal University, Russia, Kazan, Kremlyovskaya St., 18
igor-borodaev@yandex.ru

Radio-frequency (RF) plasma at lowered pressure is effectively used in various technological processes, including deposition of silver nanoparticle on fur [1]. Therefore mathematical model of silver nanoparticles interaction of with lowered pressure RF plasma is developed.

RF plasma at pressure p = 13.3-133 Pa, generator frequency f = 1.76 MHz, discharge power P_d = 0.5-4 kw, gas consumption $G < 0.2$ g·s^{-1} possesses following characteristics: ionization degree less than 10^{-4}-10^{-7}, electronic concentration n_e~10^{15}-10^{19} м$^{-3}$, electronic temperature T_e = 1-4 eV, temperature of atoms and ions in discharge T_a = (3-4) ·10^3 K, in plasma stream T_a = 350-700 K [2].

Diameter of silver nanoparticles is equal up 5 to 9·nm that is much less then Debay length $\lambda_D \sim 7 \cdot 10^{-5}$ m. Because of continuum conditions aren't carried out, it is necessary to use molecular dynamic methods for modeling of nanoparticles interaction with plasma [3].

Due to Ag^+ nanoparticles concentration is less then 10^5 m^{-3} one particle activation process excepting others particle influence is considered.

The model is constructed in the assumption that nanoparticles moves in plasma stream so that ions are motionless in the local system of coordinates connected with nanocorpuscle, but electrons fluctuate in phase with electric field intensity changing. Elementary cell of $n_e^{-1/3}$ by linear size, containing one charged silver particle is considered.

One nanoparticle contains approximately 10^4-10^5 silver atoms and ions. Energy ~7.5 eV is emited at electron recombination with Ag^+ ion. This energy is in almost three times more than the energy needed for evaporation of silver atom out of surface.

Nanoparticle charge is decreased at each electron collision with the particle, but its energy is increased due to recombination energy and kinetic energy of electron. Therefore nanoparticles is heated to melting temperature at first, and heated up to evaporation temperature after that.

When nanoparticle is loss positive charge, it is charged negatively. Limit negative charge of nanoparticle is found. Nanoparticles neutralization and recharge times as well as its residual volume are calculated.

[1] Y.P. Raizer, M.N. Schneider, and N.A. Yatzenko. Radio-Frequency Capacitive Discharges. Boca Raton, *CRC Press*, 1995.
[2] Abdullin I.S., Zheltukhin V.S, Kashapov N.F. Radio-frequency plazma-jet processing of materials at lowered pressure. Theory and Application. *Kazan: Kazan State University Publ.*, 2000.
[3] M.Mitchner, Ch.Kruger. Partially ionized gases. *Wiley-Intersci. Publ.*, 1973.

The work is supported by RFBR (project No 10-01-00728a, 11-01-00864a, 12-01-31458), MES RF (contract No 14.B37.21.1948), BP Russia (contract No C49-12, 13.01.2012).

Molecular Docking: Application of the Tabu Search Algorithm

G. Khensous[1,2]
B. Messabih[2]
B. Maigret[3]

[1] Computer Sciences Department, University of Mostaganem, Algeria
[2] Computer Sciences Department, USTO-MB, Algeria
[3] CNRS, LORIA, France
ghania.khensous@loria.fr

Leading a project of the pharmaceutical industry requires economical constraints due to the considerable development time as well as the high cost from the initiation of the project to the clinical tests. To face this high cost, the information technologies were appeared in order to improve considerably the molecule discovery. Thus, a new research domain baptised "*Bioinformatics*" was born.

Bioinformatics is a multidisciplinary research domain aiming the automatic processing of biological data. Since the experimental methods such as NMR (Nuclear Magnetic Resonance) or X-Rays Crystallography are very expensive, the High Throughput Screening and Virtual Screening are then used. One of the Bioinformatics challenges is the Molecular Docking.

Molecular docking methods consist on the determination of the best possible matching of two molecules. Therefore, we can consider the docking problem as the problem of the prediction of the complex structures formed from two or more molecules.

In order to predict the structure of the protein complex, several methods issued from the artificial intelligence and molecular geometries are proposed in literature. Thus, the molecular docking is composed of two steps; the first one is "*Searching*", the second step is "*Scoring*".

The scoring step allows the evaluation and the ranking of the complex conformations (Ligand – Receptor) found in the searching stage.

In the searching step, several simulation methods are used such as *systematic search*, *determinist search* and *stochastic search.*

In our experiment, we use the Tabu Search method which is a stochastic search algorithm. We have chosen this algorithm because population algorithms such as Genetic Algorithms or Ant Colony Optimisation algorithms require much computing time since these algorithms manipulate simultaneously many solutions.

On the other hand, most docking applications and software tested Genetic Algorithms [1] or Simulated Annealing [2]. Moreover, these applications deal with only the ligand's flexibility whereas ours takes into account both protein and ligand's flexibility.

Finally, the majority of the docking methods use scoring function whereas ours uses an energy function issued from classical molecular mechanics, which includes the polarization energy.

[1] J.S. TayloR and R.M. Burnett: DARWIN: A Program for Docking Flexible Molecules, The Wistar Institute, Philadelphia, Pennsylvania, 2000.
[2] K. Lee, C. Czaplewski, S.Y. Kim, J. Lee: An Efficient Molecular Docking using Conformational Space Annealing, Korea Institute for Advanced Study, 2004.

Toxicity of nanoparticles: contribution of physicochemical characteristics

N. Golovina[1,2]
L. Kustov[1,2]
I. Tetko[3,4,5]

[1] Department of Chemistry, M.V. Lomonosov Moscow State University, Leninskiye Gory, GSP-1, 1-3, Moscow, Russia
[2] N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky prospekt, 47, Moscow, Russia
[3] Institute of Structural Biology, Helmholtz Zentrum Muenchen – German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, b. 60w, Neuherberg, Germany
[4] eADMET GmbH, Ingolstдdter LandstraЯe 1, b. 60w, Neuherberg, Germany
[5] Chemistry Department, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah, Saudi Arabia
NataliMSU@gmail.com

Wide application of nanoparticles in products of daily use set up safety issue of nanomaterials. Although nanoparticles are already present in many commercial products [1], the potential harmful effects of nanomaterials on human's health or the environment have not yet been identified.

The classical way of assessing toxicity, e.g. by performing in vivo experiments, is very expensive and time consuming. Performing such tests for all possible nanoparticle types, sizes and concentrations is practically infeasible. Therefore computational models will be crucial to establish a generic understanding, in order to safeguard safety and to support regulation. Thus, researchers are investigating the potential of using Quantitative Nanostructure–Activity Relationship (QNAR) models to predict the properties of nanoparticles prior to their manufacturing. Traditional QSAR (Quantitative Structure–Activity Relationship) evaluate functional dependence between structure and activity of substances. Using QSARs for nanoparticles is a new and still developing area of research.

The previous review shows that no single particle characteristic can be a hallmark indicator of toxicity, although some particle characteristics show some role in directing the biological fate and toxicity [2]. However, Oberdцrster et al. [3] suggested that the particle size is not the only possible factor influencing on the toxicity of nanomaterials.

Within our study, we have considered contribution into toxicity of nanoparticles such factors as shape, nature of stabilizer and chemical composition. We have collected toxicity data of the number of nanoparticles for *in vivo* and *in vitro* tests. The data has been uploaded to the Online Chemical Modeling Environment (www.ochem.eu) and is publicly accessible by everyone on the Web. In our studies, we plan to use this data to develop predictive QSAR models for nanoparticles toxicity.

[1] Wijnhoven S.W.P. et al. *Nanotoxicology*, 2009, 3(2): 109-138.
[2] Shaw S.Y. et al. *Proceedings of the National Academy of Sciences of the U.S.A.*, 2008, 105 (21): 7387-7392.
[3] Oberdцrster G. et al. *Particle and Fibre Toxicology*, 2005, 6(2): 8.

The authors acknowledges the support from NATO SFP Program (project no. EAP.SFPP 984401) and FP7 MC ITN project Enviromental Chemoinformatics (ECO), grant agreement No. 238701.

V. Khayrullina[1]
I. Taipov[1]
A. Gerchikov[1]
D. Shcherbinin[2]
M. Vasiliev[1]
A. Veselovsky[2]
V. Trynkina[1]

Virtual screening of new potential 5-lipoxygenase inhibitors using SAR techniques and molecular docking

[1] Department of Chemistry, Bashkir State University, Zaki Validy St., 32, Ufa, Russia
[2] Orekhovich Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Pogodinskaya St., 10, Moscow, Russia
Veronika1979@yandex.ru

The search of new human's 5-lipoxygenase (5-LOX) inhibitors was done. The structure of 2-(3-benzoylphenyl)propanoic acid, the active substance of the non-steroidal anti-inflammatory drug "ketoprofen", were used for increasing its inhibitiory activity. Initially, using pattern recognition methods, implemented in a computer system SARD-21, "Structure-activity relationship" in the series of high and middle effective inhibitors of 5-LOX was studied It was results in several structural descriptors that important for effective inhibitors of this enzyme. These descriptors were used to design models for prediction and recognition of effective 5-LOX inhibitors with above 80% reliability of the prediction for two methods of pattern recognition theory. Using the discovered patterns structural modification of 2-(3-benzoylphenyl)propanoic acid was performed. It was result in 17 potentially effective inhibitors of 5-LOX. To verify the these potential structure the molecular docking of selected ligands in the active site of 5-LOX was done. Molecular docking in the active site of the molecule 3o8y (chain A, http://www.rcsb.org) using Autodock 4.2 based on Lamakrian genetic algorithm was performed. As a result, 15 potentially effective inhibitors of 5-LOX were found. These compounds have a high predicted affinity to the 5-LOX active site compared with binding energy value of arachidonic acid, the natural substrate of thise enzyme. Three compounds had predicted binding affinite higher than parrent compound, 2-(3-benzoylphenyl)propanoic acid.

The results of the visual analysis indicates that the most structure of ligands located in the same cluster as 2-(3-benzoylphenyl) propionic acid. It was established that the decisive role in the orientation of ligands in the 5-LOX active site plays a polar region, which contain non-heme iron ion, which forms coordination bonds with three polar histidine residues HIS372, HIS550, HIS367, asparagine ASN554, water molecule and COO-group of C-terminal ILE673. All ligands form one or more hydrogen bonds with the amino acids ALA424, ASN425 and HIS367. Also the position of the 2-(3-benzoylphenyl)propionic acid and the most of its derivatives allowed to form the □-□-interaction with aromatic ring of PHE421. Thus, we can conclude that proposed molecules can interact in the same way as parrent compound, 2-(3-benzoylphenyl)propionic acid. These compounds can be interest for the further synthesis and biochemical testing as a potentially effective 5-LOX inhibitors.

V. Kljashtorny
A. Shkliaeva
A. Gabdulkhakov

Molecular dynamics studies of the mutant forms of photoreaction center from *Rhodobacter Sphaeroides*

Institute of Protein Research, Instituskaya St., 4, Pushchino, Russia
vladkljasch@gmail.com

At least from an energetic point of view, the conversion of light energy to chemical energy in photosynthesis is, one of the most important biological process on Earth. Light is first absorbed by light-harvesting antenna complexes and the energy is then passed to the so-called photosynthetic reaction centres (PRCs), where the primary separation of charges and subsequent electron transfer across the photosynthetic membranes occur with almost 100% efficiency. For many years, well-characterized PRCs from purple bacteria have been serving as a structural and functional model for studing the evolutionarily related photosystems of plant. PRC of purple bacterium *Rhodobacter sphaeroides* represents a membrane associated protein that consists of three protein subunits and ten cofactors of electron transfer. The cofactors count four bacteriochlorophylls (BChl), two bacteriopheophytines, two ubiquinones, a molecule of carotenoid and a non-heme iron atom. Four BChl cofactors in the *Rba. sphaeroides* PRC with their central Mg atoms are pentacoordinated. In addition to the four in-plane bonds with tetrapyrrol nitrogens, This Mg is also coordinated by His residue from the adjacent protein molecule. The microenvironment of the BChls formed by the axial ligands and closest protein residues can significantly influence their photophysical and redox properties. Thus, these properties could be modified by site-directed mutagenesis of protein subunits. Comparison of the crystal structures of mutant and intact PRCs provides the information on how the interactions between protein and cofactors may affect the spectral properties of BChls and the stability of the PRC complex.

Isolation, purification and crystallization of the biological systems is a time consuming process. Thereby, the Molecular Dynamics (MD) simulation method could be used to perform preliminary analyses of mutations in the PRCs in order to evaluate their influence on stability of the complex allowing to construct the most reasonable mutant forms for subsequent biochemical and crystallisation experiments. Using such approach we calculated 50ns MD trajectories for the wild-type PRC and its different mutant forms. The all-atom simulations were performed with GROMACS software using NPT anssemble and excplicit water enviroment. Taking into account the results obtained, several mutants were chosen for futher spectral analysis and crystallisation.

This work was supported by Russian Academy of Sciences within the program 'Molecular and Cell Biology', the Russian Foundation for Basic Research (grant 13-04-01148a).
All the MD simulations were performed using sources of the Joint Supercomputer Centre of the Russian Academy of Sciences (JSCC RSA http://www.jscc.ru/).

Study of tylosin and its derivatives binding to *E. coli ribosome* by molecular dynamics simulation method

G. Makarov[1]
A. Golovin[2]
A. Shishkina[3]
G. Korshunova[3]
N. Sumbatyan[3]

[1] Department of Chemistry,Lomonosov Moscow State University, GSP-1, 1-3, Leninskie Gory, Moscow, Russia
[2] Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, GSP-1, 1-73, Leninskie Gory, Moscow, Russia
[3] A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, GSP-1, 1-40, Leninskie Gory, Moscow, Russia

ftorazyne@mail.ru

By new protein biosynthesis inhibitors development tylosin and its derivatives, namely 5 O mycaminosyltylonolide (OMT), 23 O (*tret*-butyloxycarbonil glycyl) OMT (BocGlyOMT), 23 O (*tret*-butyloxycarbonil β alanyl) OMT (BocβAlaOMT), 23 O (*tret*-butyloxycarbonil γ aminobutyryl) OMT (BocγAbuOMT), binding was studied by molecular dynamic simulation. These OMT derivatives are promissory as precursors in peptide-macrolide conjugates, which are probes for ribosome functioning examination, synthesis and new antibiotics. In 2000s molecular mechanisms of ribosomal antibiotics, including macrolides, action were clarified by means of obtaining of atomic structures of bacterial ribosome and its complexes with antibiotics. But structure of tylosin–*E.coli* ribosome complex as well as structure of tylosin derivatives complexes is still unknown. Molecular dynamic simulation models these complexes in solution in time, that enables us to compare different antibiotics behavior and deduce about its comparative activity.

23 O (*tret*-butyloxycarbonil γ aminobutyryl) OMT was synthesized from OMT.

OMT derivatives inhibitory activity was compared in system of cell-free firefly luciferase translation.

Molecular dynamic simulation of tylosin and its derivatives complexes was performed and explanation of listed macrolides inhibitory activity difference based on obtained trajectories analysis was proposed.

Calculations were performed on Moscow State University “Lomonosov” supercomputer.

Many-dimensional model of water-salt system's solubility isotherm-isobar in case of simple evtonical type

M.S. Moiseenkov
S.V. Rusakov

Department of Applied Mathematics and Informatics,
Perm State University, Bukirev St., 15, Perm, Russia
moiseenkov.ms@gmail.com

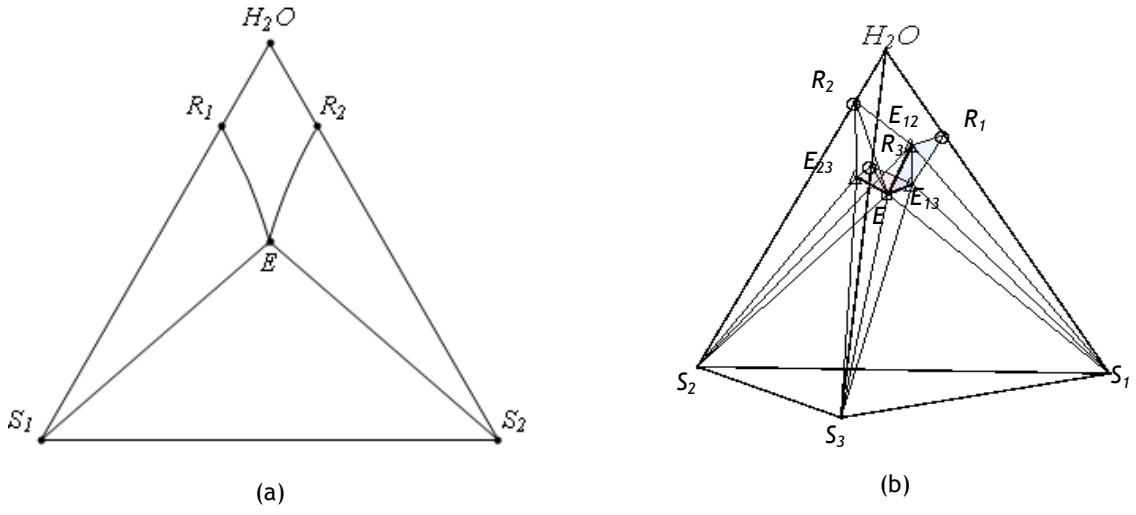
Water-salt system's solubility isotherm-isobar is a phase solubility diagram that used to predict physicochemical parameters of water-salt system with constant temperature and pressure [1]. It is actual for scientific research of water-salt systems and industrial production mineral fertilizer and salts.

At present time phase diagrams are constructed using experimental data by the instrumentality of Gibbs-Roozeboom method in baricentric coordinate system [2]. Graphical data representation is limited in that dimension of the phase diagram depends on number of components of water-salt system. Furthermore increasing number of components complicates geometric structure of diagram. Geometric structure of phase solubility diagram of water-salt systems in case of simple evtonical type is depicted in Figure.

To automate the processing of experimental data a universal model of solubility isotherm-isobar for water-salt systems in case of simple evtonical type was developed. Data presentation is based on regionalization of phase solubility diagram. Every phase region is divided into disjoint set of simplexes obtained after construction of the Delaunay triangulation on the experimental point lying at the interface boundary.

Unified data storage based on usage of generalized geometrical methods in physical-chemical analysis allows reducing the processing of multi-dimensional experimental data to solution of simple geometric problems.

Geometric structure of solubility isotherm-isobar which is invariant with respect to spatial dimension and is implemented in the object-oriented paradigm allows developing applied class library which can serve as the basis for domain-specific computational service.



(a) (b)

Figure. Geometric structure of water-salt system's solubility isotherm-isobar:
(a) three components; (b) four components

[1] Kurnakov N.S. "Vvedeniye v Fiziko-Khimicheskii Analiz" (Introduction to Physicochemical Analysis), 4-e izd., *Akad. Nauk S.S.S.R. Moskva*, 1940.
[2] Mazunin S.A. "Osnovy Fiziko-Khimicheskogo Analiza, Chast 2. Mnogokomponentnyje sistemy" (Bases of Physicochemical Analysis, Part 2. Many-component water-salt systems), *Perm*, 2000.

E.G. Mokshina
P.G. Polishchuk
V.I. Nedostup
V.E. Kuz'min

Critical properties: QSPR of binary organic mixtures

Department of Molecular Structure and Chemoinformatics,
O.V. Bogatsky Physico-Chemical Institute NAS of Ukraine,
86 Lustdorfs'ka doroga, 65080 Odessa, Ukraine
mokshinaelena@ukr.net

Critical properties (i.e., critical temperatures, pressures and volumes) are one of the most important thermodynamic characteristics of compounds and their mixtures. On the one hand, modern experimental measurements are distincted by high precision and accuracy, on the other hand, for complex organic compounds experimental results are frequently ambiguous. Therefore, development of the novel fast prediction techniques for critical properties is subject of great interest from the point of expert system development as well as validation of QSPR methodologies.

Experimental data for the current study was taken from the comprehensive review [1], containing data for more than 300 mixtures of different composition.

Simplex representation [2] was used to describe molecular structure. Novel descriptor approaches, aiming to describe intermolecular interactions, were developed. To describe interactions between two molecules pair potential-based descriptors were introduced. For all the possible pairs between two molecules values of Lennard-Jones and Coulomb potential's constants are used. Depending on the absolute value of the constant two-atom fragment is labeled as strong-interacting, medium-interacting, weak-interacting, etc. Despite the simplicity of such approach it corresponds to fundamental physical background, being strongly theoretically based.

To built QSPR models range of state-of-art modern statistical techniques were involved, such as PLS, averaged neural networks, Random Forest, SVM methods. For validation purposes were used "compounds-out" and "mixtures-out" approaches.

Obtained statistical models are robust and show average prediction error about 15 K, which is comparable with experimental.

[1] C.P. Hicks, C.L. Young. The Gas-Liquid Critical Properties of Binary Mixtures. *Chemical Reviews*, **1975,** *75***,** *2*, 119-175.
[2] Kuz'min V. E. et al. Recent Advances in QSAR Studies: Methods and Applications. *Springer+*, **2010**.
[3] E.N. Muratov, E.V. Varlamova, A.G. Artemenko, P.G. Polishchuk, V.E. Kuz'min Existing and Developing Approaches for QSAR Analysis of Mixtures. *Mol. Inf.* **2012**, *31*, 202-221.

QSAR modeling of nanoparticles' properties: structural-based or experimental-based?

N. Novoselska[1]
V.E. Kuzmin[1,2]

[1] Department of Chemistry, I.I. Mechnikov Odessa National University, Dvoryanskaya St., 2, 65082, Odessa, Ukraine
[2] A.V. Bogatsky Physico-Chemical Institute of the NAS of Ukraine, 86 Lustdorfskaya Doroga, 65080, Odessa, Ukraine
natalia.novoselska@gmail.com

Unfortunately, traditional QSAR approaches are insufficiently effective for the decision of such problems. In traditional QSAR the structure of individual molecules is analyzed, but nanoparticles' properties are caused by interactions between great number of single molecules.

Thus, the main problem is absence universal methods of the description nanoparticles for the follow-up QSAR. Nanoparticles are characterized by high structural complexicity, because they are complex assemblies of inorganic and organic elements. But classical molecular descriptors were developed originally for small organic molecules.

In particular, in work [1] was supposed that besides classical descriptors, for the description nanoparticles, it is necessary to consider such empirical characteristics, as the size of particles, the form, porosity, the surface area, surface chemistry, a superficial charge, and also crystal structure. However, in the literature there is no data about these empirical characteristics. Thus, there is a necessity of the theoretical analysis of nanoparticles' structure.

In the current study was held comparative analysis of different methods of QSAR modeling of nanoparticles and was investigated applicability of develpoped QSAR technique, which is based on symbiosis liquid drop model (it used to description QDs wholly) and Simplex Representation of Molecular Structure [2] (it used to description molecules' individual physicochemical properties).

[1] Oberdцrster, G. et al. Nanotoxicology: An Emerging Discipline Evolving from Studies of Ultrafine Particles // *Environ.l Health Pers.* 2005. 13. P.823-839
[2] Kuz'min V.E. et al. Virtual screening and molecular design based on hierarchical QSAR technology. // *Challenges and Advances in Computational Chemistry and Physics.* 2010. 8. p.127-176.

Your acknowledgement or funding description, min 12pt, italics, row interval 1.0, Times New Roman.

MM simulation of phosphopantetheine adenylyltransferase of mycobacterium tuberculosis upon binding of ATP

D. Podshivalov
V. Timofeev
L. Chupova
R. Esipov
I. Kuranova

[1] Laboratory of X-ray Analysis Methods and Synchrotron Radiation, Shubnikov Institute of Crystallography, Russian Academy of Sciences, Leninsky Prospect 59, Moscow 119333, Russia
[2] Laboratory of Biotechnology, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya 16/10, Moscow 117997, Russia
komunelen@mail.ru

Phosphopantetheine adenylyltransferase from Mycobacterium tuberculosis (PPAT Mt) is a hexamer composed of identical subunits and catalyzes the penultimate five-step reaction of the biosynthesis of coenzyme A (CoA), the reversible transfer of an adenylyl group from ATP to 4'-phosphopantetheine, resulting in the formation of 3'-dephosphocoenzyme A (dPCoA) and pyrophosphate. Reaction, is catalysed by PPAT, is crucial for the biosynthesis of CoA – a metabolite, that necessary for the life of mycobacteria, so PPAT Mt is a suitable target for the synthesis of anti-tuberculosis drugs.

It is known from the of X-ray studies that when binding ligands such as ATP and dPCoA with this enzyme, the enzyme molecule undergoes significant conformational changes [1]. But in this case we know only initial and final states of composition. Used in this study, the method of molecular dynamics has allowed to create a temporal model of the conformational changes upon ligand binding.

Model of PPAT Mt hexamer in the complex with ATP was created based on known structural data (PDB ID: 4E1A, 3UC5) with use the coordinates of unbound protein. This first approximation of protein-ligand system we used to develop the minutes of molecular dynamic for simulation the conformation change, which follow the binding of ATP. Thus, the simulation of dynamics of molecule was carried out from initial to final state. The temporal model of conformational change was received. This information will be useful in the development of innovative anti-TB drugs.

[1] Timofeev, V. I., Smirnova, E. A., Chupova, L. A., Esipov, R. S, & Kuranova, I. P. (2012) Crystallography Reports. 57, 96-104.

This work was supported by the Russian Foundation for Basic Research, grant 12-02-12053-ofi_m.

UNIFAC based information system as a tool for simultaneous chemical and phase equilibria calculations

N.S. Tsvetov
V.I. Rakhimov
M.V. Zamorianskaya

Department of Chemical Thermodynamics and Kinetics, Faculty of Chemistry, St.Petersburg State University, Universitetskiy prospect, 26, Peterhof, Saint-Petersburg, Russia
tsvet.nik@mail.ru

Study of chemical and phase equilibria is a pressing task in both applied and fundamental natural science. In particular, such investigations are important for the organization of processes of reaction extraction and reaction rectification.

Relevant experimental results require adequate theoretical processing and interpreting including thermodynamic process modeling and implementation of reliable thermodynamic calculation algorithms. For this purpose we have developed an information system that enables one to perform calculations of manifolds of chemical and phase equilibria in a wide range of reaction systems. In this information system calculation of activity coefficients is based on the UNIFAC group contribution model. The UNIFAC model, with its well-known limitations, favorably differs from other models in that it can be used in predictive calculations based on a limited number of group parameters. For the development of the information system we used the MS Access runtime environment, where we created the database of the UNIFAC model parameters and the database of compounds and their properties; the latter took into account splitting molecules into groups. Calculation of different types of equilibria (chemical, phase) based on experimental data processing and predictive calculations is performed with VBA (Visual Basic for Applications) modules.

Our information system allows calculations with the data either imported from MS Excel files or generated (in case of predictive calculations) directly in the calculation modules. Resulting data sets are exported to an Excel file specified by the user.

On the basis of the implemented algorithms we have calculated chemical equilibrium diagrams in reacting systems involved in n-propyl acetate and ethyl acetate synthesis. We also have calculated liquid-liquid-vapor and liquid-vapor phase equilibria for these systems taking into account the vapor nonideality and the effect of acetic acid dimerization in the gas phase. We have predicted the evolution of the reacting system with the liquid-vapour equilibrium in the course of a chemical reaction. Comparison of experimental and calculated data shows that the information system can be successfully applied to both predictive and approximate calculations. In addition, it is a convenient tool for visualization of thermodynamic regularities in multicomponent systems.

A.A. Repin[1]
E.M. Chernova[1]
V.V. Turovtsev[2,1]
Yu.D. Orlov[1]

Interactive database for enthalpies of formation of free organical radicals

[1] Tver State University
[2] Tver State Medical Academy
andreyrepin@live.ru

Providing quantitative information, for example, physics-chemical properties of compounds, which is necessary condition of modern fundamental and applied research. Intermediats – free radicals ($R^{\bullet}$) play great role in chemical and biochemical processes. However, $R^{\bullet}$ has high chemical activity and it's complicated to get their characteristics, but current thermodynamic data are meager and contradictory.

It possible to calculate corrective value of enthalpy of formation using bond dissociation energy $D(R_1\text{-}R_2)$:

$$D(R_1\text{-}R_2) = \Delta_f H^{\circ}(R^{\bullet}_1) + \Delta_f H^{\circ}(R^{\bullet}_2) - \Delta_f H^{\circ}(R_1R_2),$$

where $\Delta_f H^{\circ}(R^{\bullet}_1)$, $\Delta_f H^{\circ}(R^{\bullet}_2)$, and $\Delta_f H^{\circ}(R_1R_2)$ – standard enthalpy of formation of radicals and original molecule respectively. For a few years we have being carried systematically refreshing and enlarging database of $\Delta_f H^{\circ}(R^{\bullet})$ and $D(R_1\text{-}R_2)$ with analysis and verification published values. Received characteristics are used for parameterization of molecular fragments of additive models.

Firstly the database for $\Delta_f H^{\circ}(R^{\bullet})$ was created and published [1] with expert opinion for 471 radicals from different homologous series. Using of the new literary values $D(R_1\text{-}R_2)$ allowed to expand and clarified current database for $\Delta_f H^{\circ}(R^{\bullet})$ and $D(R_1\text{-}R_2)$ [2-5].

At present we carry out the work by connecting database and procedures of calculation unknown values $\Delta_f H^{\circ}(R^{\bullet})$ and $D(R_1\text{-}R_2)$, and also providing access using Internet. Using interface web-site a remote user can find in database necessary compound by several characteristics such as: brutto-formula, compound's name, literary sources. By the way, there will be the program allowing to draw molecules and search $\Delta_f H^{\circ}(R^{\bullet})$ after analyzing the image. Including parameters and relations of additive-group method, that adapted for free organic radicals, lets calculate $\Delta_f H^{\circ}(R^{\bullet})$ and $D(R_1\text{-}R_2)$ using structure formula, also for matters without this characteristic.

[1] Yu.D. Orlov, Yu.A. Lebedev, and I.Sh. Saifullin. *Thermochemistry of Organic Free Radicals* (Nauka, Moscow, 2001) [in Russian].
[2] Yu.D. Orlov, Yu., V.O.Lavrov, A.Yu. Lebedev. *Russian Chemical Bulletin*, 2001, Volume 50, Issue 6, pp 963-969 [in English].
[3] Yu.D. Orlov, E.M. Chernova, V.V. Turovtsev. *Russian Chemical Bulletin*, 2010, Volume 59, Issue 10, pp 2009-2013 [in English]
[4] Yu.D. Orlov, E.M. Chernova, V.V. Turovtsev. *Chemistry of Heterocyclic Compounds*, 2011, Volume 47, Issue 1, pp 29-35 [in English].
[5] Yu.D. Orlov, E.M. Chernova, V.V. Turovtsev. *Russian Journal of Physical Chemistry*, 2012, Volume 86, Issue 2, pp 175-183 [in English].

Ab initio study of energy transfer processes in lanthanide(III) complexes with β-diketonates and lewis bases

K.A. Romanova[1]
A.Ya. Freidzon[2]
A.A. Bagaturyants[2]
M.V. Strelkov[1]
Yu.G. Galyametdinov[1]

[1] Physical and colloid chemistry department, Institute of Polymers, Kazan National Research Technological University, Karl Marx St., 68, Kazan, Russia
[2] Photochemistry Center, Russian Academy of Sciences, Novatorov St., 7a, Moscow, Russia
ksenuya@mail.ru

The growing interest in the lanthanide(III) complexes can be explained by their remarkable emission properties: long luminescence lifetimes, large Stokes shifts and narrow emission bands. They are widely used in light-emitting devices, displays, lasers, luminescence bioprobes, sollar cells and etc. Their emission efficiency depends on the relative positions of the excited levels in the emitting ion and in the ligands. Liquid-crystalline lanthanide(III) complexes, due to their unique magnetic and spectroscopic properties as well as their possibility to align in an external electric or magnetic field, are promising luminescent materials with controlled polarization [1]. Computer simulation can help one to find the ligands that provide the most efficient energy transfer to the lanthanide(III) ion and noticeably simplify the design of highly efficient luminescence materials.

In this work, ab initio XMCQDPT2/CASSCF approach [2] is used to calculate the energies of the ground and excited states of europium(III), gadolinium(III) and terbium(III) complexes with different β-diketonates and Lewis bases (2,2′-bipyrimidine and 1,10-phenantroline). For lanthanide(III) ions we use scalar relativistic 4f-in-core pseudopotentials (ECP52MWB for europium(III), ECP53MWB for gadolinium(III), ECP54MWB for terbium(III)) with the associated valence basis sets. The 6-31G(d,p) basis set is used for other atoms. The triplet and singlet excited states, which are localized on each of the four ligands, are optimized by the CASSCF method together with the ground state. The vertical triplet and singlet excitation energies are calculated at the optimized geometries using the state-averaged CASSCF method and corrected by the XMCQDPT2 method. XMCQDPT2 calculations are performed separately for singlet and triplet excited states. The excited state optimization showed the structural changes in the ligand that carries the excitation.

It is found that during the photoexcitation the greatest contribution comes from the β-diketones whose geometry considerably changes in comparison with Lewis bases. On the basis of the calculated data the main intramolecular energy transfer channels are determined. In order to estimate the efficiency of energy transfer the energy transfer rates are calculated. Correlations between the positions of the excited levels and measured values of absolute quantum yield are established. The theoretical results are in good agreement with experimental data. It is shown that during the photoexcitation of lanthanide complexes both energy transfer process from organic ligand to ion and singlet-triplet conversion can be considered as rate-limiting steps.

[1] Galyametdinov Yu.G. et al. *Advanced Materials*, 2008, **20**: 252-257.
[2] Freidzon A.Ya. et al. *The Journal of Physical Chemistry A*, 2011, **115**: 4565-4573.

This work was supported by the Russian Foundation for Basic Research (project no. 11-03-00597-a). The calculations were performed using the facilities of the Joint Supercomputer Center, Russian Academy of Sciences.

N.P. Rusakova[1]
A.A. Repin[1]
V.V. Turovtsev[2]
Yu.D. Orlov[1]

Database electronegativity of functional groups of organic connections of sulfur

[1] Tver State University
[2] Tver State Medical Academy
d002186@tversu.ru

Recently in pharmaceutical and biological chemistry there is a search of the adequate models, allowing to predetermine required biological activity of substance and a way of it synthesis. Designing of macromolecules *in silico* should lean on a database (DB) containing the information on an electronic structure (a charge, volume, energy, etc.) and electronegativitis (□) various functional groups *(R)*. The information incorporated in DB will allow to estimate redistribution of a charge inside a molecule, to establish the possible reactionary centers and mutual spatial distribution of charges in reacting connections of different classes.

We create the database of initial molecular blocks, including atoms of polyvalent sulfur, and the interface of construction of bioorganic connections, including potentially possessing is developed by biological activity. The choice components containing sulfur is caused by that polyvalent sulfur is part modern medicinal sulfanilamide of preparations and various intermediates endocellular biological synthesis. Further expansion of a database is planned due to entering the settlement data on other functional groups.

At construction of a database the analysis carried out earlier □ groups in molecules [1, 2] with an estimation and their comparison partial charges was taken into account. On the basis of the received data laws in electronegativitis formally identical groups in various gomologous lines have been revealed. Calculation of equilibrium structures and their electronic structure has been made with use of program GAUSSIAN 03 [5] by method B3LYP/6-311++G (3df, 3pd). Charges *q* topological atoms Ω have been found in frameworks QTAIM by [3] numerical integration within the limits of internuclear surfaces and isosurface electronic density 0.001 a.e. with the help of program AIMALL [4]. All sizes received for Ω, have been related to functional groups *q (R)*. The error of calculation partial charges *q (R)* made no more than 0.001 a.e. (1 a.e. = $1.6 \cdot 10^{-19}$ C).

As a result of mass calculations the analysis more than 325 both organic, and inorganic molecules and the radicals containing also alkyl functional groups and other groups, including atom of sulfur has been lead. The base on the found data of functional groups is created and work on it expansion proceeds. The submitted database will allow not only to store and analyze available fund of an electronic structure and electronegativity of functional groups, but also to carry out on demand of the user designing of the new, not synthesized preparations.

[1] Rusakova N.P., Turovtsev V.V., Orlov Yu.D. // The Bulletin of the Tver state university. A series: Chemistry. 2010. № 10. pp. 4-8
[2] N.P. Rusakova, V.V. Turovtsev, Yu.D. Orlov // Zhurnal Prikladnoi Khimii, 2011, Vol. 84, No. 9, pp. 1578–1580.(Russian Journal of Applied Chemistry, September 2011, Vol. 84, I. 9, pp. 1626-1628)
[3] Bader R. // Atomy in molecules: the Quantum theory. M.: the World, 2001.532 with.
[4] AIMAll (Version 10.03.25), Todd A. Keith, 2010 (http: // aim.tkgristmill.com)
[5] Frisch M.J., Trucks G.W., et al. Gaussian 03 (Revision E 0.1). Gaussian Inc., Pittsburgh PA, 2007.

Your acknowledgement or funding description, min 12pt, italics, row interval 1.0, Times New Roman

Evolutionary role of the hydrophobic segment of bacillar ribonucleases

F.V. Shirshikov

Department of Microbiology, Institute of Basic Medicine and Biology, Kazan (Volga Region) Federal University, 18, Kremlyovskaya St., Kazan, Russia
shrshkv@ya.ru

By modern computational methods, such as the water-to-bilayer transfer free energy and hydrophobic moment values calculations, was shown that some bacillar ribonucleases (RNases) with antitumor action, namely barnase and binase, have a hydrophobic segment located in helix II. Interestingly, physicochemical properties of the hydrophobic segment of above-mentioned RNases different significantly. The aim of this study is to explain the evolutionary significance of bacillar secretory RNases using the methods of structural bioinformatics.

Although attention has long been known to antitumor action of binase [1], a detailed study of its natural destination in the structure biology aspect had been not observed. Secretory RNases involved in the metabolism of inorganic phosphate, however, the data about the hydrophobic segment suggest that RNases can play subtle biological role. Remarkably, the biosyntesis of binase, not barnase, can be stimulated by addition of dactinomycin [2].

The hydrophobic segment of barnase contains a glutamate residue (Glu-29). Water-to-bilayer transfer free energy of the hydrophobic segment of barnase, containing anionic Glu-29 residue, have a positive value, which means that favorable hydrophobic interaction is impossible. On the adsorption stage the hydrophobic interaction apparently unimportant, because the adsorption of barnase mostly depends on cationic character of this toxic RNase. But if we try to calculate the free energy of the hydrophobic segment, containing protonated glutamate residue, we can get a negative value, which suggests that favorable the interaction of hydrophobic segment with a lipid bilayer can be possible only in acidic environment.

Only eukaryotic cells have a compartment with acidic lumen – an endosome. Bacilli and other bacteria have not such compartments, therefore the hypothesis about the role of secretory barnase as a antibacterial agent or bacteriocin is possible in acidic environment, which can be created by fungi or acidic soils.

The hypothesis that the evolutionary role of barnase and binase as a membrane-lysing factors consists in the antagonistic relationships with other bacteria and microscopic fungi was proposed. This hypothesis based on the chemical properties of hydrophobic segment, namely on the hydrophobic indexes of amino acid residues. It is should be noted that dactinomycin can be a signal molecule for the producent of binase, which will be inform bacterial cells about the presence of the producents of dactinomycin-like molecules in environment, but barnase is more universal antibacterial and antifungal RNase in acidic environment. Such antibiotic-sensitivity of secretory toxic RNases can be an effective weapon in the struggle for inorganic phosphate. Moreover, the hypothesis that the RNase cytotoxicity toward both bacteria, as natural function, and tumor cells can be based on the complex of membrane-lysing and catalytic activity was proposed. The results obtained can be useful for the development of the new generation membrane-lysing and cell-penetrating peptides. In this case the hydrophobic segment sequence is basis for the development of above-mentioned peptides.

[1] Makarov A.A. and Ilinskaya O.N. *FEBS Letters*, 2003, 540: 15-20.
[2] Gabdarakhmanova L. A. et al. *Antibiotiki i khimioterapiia*, 1997, 42: 15-21.

Thanks go to Gelfand M.S. for useful advices and the Russian Society of Biochemistry and Molecular Biology for support.

Physicochemical properties of the antitumor ribonucleases

F.V. Shirshikov

Department of Microbiology, Institute of Basic Medicine and Biology, Kazan (Volga Region) Federal University, 18, Kremlyovskaya St., Kazan, Russia
shrshkv@ya.ru

Ribonucleases (RNases) are ubiquitous enzymes, which able to catalyse biochemical reactions of the degradation of ribonucleic acids. Some of these enzymes are secretory proteins with a wide spectrum of biological activities [1]. The bacterial RNases, such as RNase from *Bacillus amyloliquefaciens* (barnase) and RNase from *B. intermedius* 7P (binase), were proven to be a toxic proteins with antitumor, antiviral, and bactericidal actions. The precise mechanism and general molecular determinants of the biological effects of RNases are unclear. However, all biological effects should be connected with influence on the lipid bilayer structure of target cells.

Subjects of the research were secretory RNases from bacilli as follows: barnase and binase. The three-dimensional structure of these proteins was solved previously.

Here, the homology modeling approach was used to predict the three-dimensional structure of RNase from *B. circulans* (Bci-RNase), RNase from *B. coagulans* (Bco-RNase), and RNase from *B. thuringiensis* (Bth-RNase). These protein structures have not been experimentally determined. Although the cytotoxic action of above-mentioned RNases is not yet proved, their close homology to barnase and binase suggests that these proteins should be tested as an antitumor drugs.

The aim of this study is to determine the physicochemical properties of toxic RNase and their close homologs. Few important characteristics, which are necessary for the unrevealing of the mechanism of RNase cytotoxicity, namely charge, dipole moment, water-to-bilayer transfer free energy, and hydrophobic moment values were calculated. The molecular electrostatic maps of toxic RNases were also obtained. The peculiarity of this investigation consists in the complex view on the lipid-protein interaction, which includes not only the electrostatical component, but also the hydrophobic component.

It was shown that barnase and binase have similar electrostatic and hydrophobic patterns, which allow the thermodynamically favorable interaction with tumor cell surface and bacillar plasma membrane, both of them have anionic compopents. Moreover, all RNases investigated have a helical hydrophobic segment, which probably acts as a membrane-binding domain. By modern *in silico* approach was demonstrated that the surface of anionic lipid bilayer, which contains a dioleoylphosphatidylserine, possesses a mosaic hydrophobic-hydrophilic nature [2]. The phenomenon results in the forming of interfacial hydrophobic clusters, which may be as sites for the thermodynamically favorable insertion of amphipathic proteins, like RNases. The latter fact suggests that the tumor cell surface and the toxic bacillar RNases have complementary physicochemical properties.

[1] Benito A. et al. *Molecular BioSystems*, 2005, **1**: 294-302.
[2] Polyansky A. A. et al. *Journal of Physical Chemistry B*, 2009, **113**: 1120-1126.

Thanks go to the Russian Society of Biochemistry and Molecular Biology for support.

Virtual screening, synthesis and activity of chloramphenicol peptide derivatives

A.G. Tereshchenkov

Chemistry Department of Moscow State University, 119991, Moscow, Russia.
tereshchenkov@list.ru

Chloramphenicol is a widely used antibiotic that inhibits protein synthesis by binding to bacterial ribosomes in the A-site of the peptidyl transferase center. X-ray data of the ribosome – chloramphenicol complexes showed that dichloroacetic moiety of the antibiotic is directed to the ribosome tunnel [1], and, consequently, if being replaced with amino acid or peptide residue could mimic nascent peptide chain. In continuation of our previous studies in which we have designed and synthesized a number of peptide derivatives of macrolides where the peptide part modeled the growing chain, while the antibiotic served as an "anchor" for positioning the peptide at the specific site of ribosomal tunnel [2, 3], now we report chloramphenicol peptide derivatives as analogues of the peptidyl-tRNA. These derivatives are of interest both as antibacterial agents and as tools for investigation of the interactions of nascent peptide chain with the specific sites of the ribosomal tunnel and their influence on the translation. Previously [4] some amino acid and peptide derivatives of chloramphenicol were described.

Virtual screening for binding of all amino acid, di- and tripeptide chloramphenicol derivatives to the*E. coli* ribosome was performed. These derivatives were constructed by removing of dichloroacetic moiety from chloramphenicol structure and attaching of amino acid or peptide by its carboxyl group to vacant amino function of chloramphenicol residue. Optimized 3D structures of the chloramphenicol peptide derivatives were generated from SMILES strings using the OpenBabel program. Docking-based virtual screening was carried out by means of the AutoDock Vina software [5], using the rigid structure of the ribosome [6] and flexible structures of ligands. Molecule conformations obtained by means of the molecular docking were ranged both by their binding energies, and by the RMSD of chloramphenicol moiety from corresponding X-ray data. The most interesting compounds including chloramphenicol derivatives modified with short "stop-peptides" (MRL, IRA and TRP) were synthesized. The new chloramphenicol derivatives were examined for their binding ability to bacterial ribosomes by displacement of fluorescently labeled erythromycin from their complexes with *E. coli* ribosome and for their power to inhibit translation *in vitro* and *in vivo*.

[1] Dunkle J.A. et al. *Proc Natl Acad Sci USA*, 2010, **107**: 17152-17157.
[2] Sumbatyan N.V. et al. *Biochemistry (Mosc)*, 2003, **68**: 1156-1158.
[3] Starosta A.L. et al. *Chem Biol*, 2010, **17**: 504-514.
[4] Kostopoulou O.N. et al. *Open Enzym Inhib J*, 2011, **4**: 1-10.
[5] Trott O. et al. *J Comput Chem*, 2010, **31**: 455-461.
[6] RCSB PDB – 3OFC.

This study was supported by the Russian Foundation for Basic Researches (grants 12-04-31558-mol_a, 13-04-00986-a).

Hierarchical clustering of large databases and classification of antibiotics at high noise levels

S. Trepalin[1]
A. Yarkov[2]

[1] Institute Physiologically Active Compounds, Russian Academy of Science, Chernogolovka, Russia
[2] Institute Physiologically Active Compounds, Russian Academy of Science, Chernogolovka, Russia
trep@chemical-block.com

Algorithm for divisive hierarchical clustering of chemical compounds based on 2D structural fragments is suggested. The suggested algorithm involves the following steps:

1. Generating hash codes for all chemical structures in the initial dataset. At this stage all topological duplicates are removed automatically – they have identical hash codes.
2. Setting the initial number of clusters (K = 1) and starting the similarity threshold (R = 0.0).
3. Requesting the number of steps (Nst) for further calculation of the difference between similarity thresholds (Rd = 1/Nst). The Nst value equals the number of levels in the resultant hierarchy.
4. Diversity sorting in each cluster C for similarity threshold R. Forming the set of probe molecules S, namely:

a. Selecting the compound most dissimilar to all molecules in the cluster and putting it into set S.

b. Selecting compound M which is the most dissimilar to structures in set S (MinMax diverse selection). At first step, the compound M is most dissimilar to single compound, selected in step a). At next steps M has maximal average distance to all compounds in set S.

c. Calculating similarity of compound M to all compounds in set S. If maximal similarity ratio is greater than R+Rd, set S remains unchanged. Otherwise, compound M is added to set S. The algorithm then proceeds to Point b for selecting the next compound.

The number of compounds in probe set S is equal to the number of clusters plus the number of singletons. Thus, the number of resultant clusters is not determined beforehand-it is calculated from diversity of dataset.

5. Calculating similarity of each remaining compound to compounds in set S. Assigning compounds to the cluster with maximal similarity.
6. Setting similarity threshold (R = R+Rd). Assigning this value to all clusters generated from cluster C. The algorithm then proceeds to Step 4. It is repeated until similarity threshold reaches value 1.

The algorithm is deterministic, and given a random ordering of the input, will always give the same clustering and can process a database up to 2 million records on a standard PC. The algorithm was used for classification of 1.183 antibiotics mixed with 999.994 random chemical structures. Similarity threshold, at which best separation of active and non active compounds took place, was estimated as 0.6. 85.7% of the antibiotics were successfully classified at this threshold with 0.4% of inaccurate compounds. A .sdf file was created with the probe molecules for clustering of external databases.

Fast chemical structure search in large WWW databases with "*InChI key*" approach

S. Trepalin[1]
Yu. Bessonov[2]
L. Koroleva[3]

[1] Institute Physiologically Active Compounds, Russian Academy of Science, Chernogolovka, Russia
[2] VINITI Russian Academy of Science
[3] VINITI Russian Academy of Science
trep@chemical-block.com

International Chemical Identifier (InChI) was recommended by IUPAC [1] to save chemical structure as linear notation in computer-readable form. InChI product is free for any use and contains many frameworks of chemical structure processing – normalization and canonicalization. *InChI key*, derived from InChI is a hash, created by SHA-2 256 algorithm from InChI line notation. It was designed for exact chemical structure search. It contains two blocks: connectivity block and block of stereochemistry. Each block is a string and to determine if two chemical structures are identical, it is necessary to determine if these strings are identical. *InChI key* has the constant length – 14 letters for the connectivity block and 8 for stereochemical+protonation block (constant values from *InChI key* are not taken into consideration).

On the set of strings *InChI key* can be defined the metric and, thus, the relationships of greater than, less than and equal to. Therefore, it is possible to sort and one can use the quick bisection algorithm to search for a string. Suppose one has n strings, sorted in ascending order. Then string with index $n/2$ is extracted and is compared with the target string. If the string from the array is greater than target, second half of the array is rejected and the procedure is performed on the first half. On the contrary, if the string is less than target, then the first half of array is rejected and the procedure is performed on the second half. The speed of convergence is 2^k where k – number of iterations. It is necessary only 27 comparisons to find a target record in the largest chemical database (PubChem, 108 million records with duplicate). Assuming the number of iterations = 32, we will get full search in 4 billion database, which is much greater than the total number of known chemical substances.

Another problem is the using of binary representation of *InChI key* string. Capital Latin letters from A to Z are used in the *InChI key* – 26 letters in total. It will be enough only 5 bits to cover this range, while byte contains 8 bits. So, conversion of *InChI key* string to binary would require 9 bytes instead of 14 to store the connectivity block and 6 bytes instead of 8 to store stereochemistry block – 15 bytes in total. So, *InChI key* string will occupy 1.5G RAM for a database with 100M records – the largest known one.

The chemical WEB server was created based on open Indy Internet components as 64-bit application. JME Java Applet [2] was used as query builder. Start search page is available at www.viniti.ru.

[1] Stein S. and all Proceedings of the 2003 International Chemical Information Conference (Nimes), Infonortics, 2003, 131-143.Lagunin A. et al. *Bioinformatics*, 2000, **16**: 747-748.
[2] http://www.molinspiration.com/jme

Current job has the finnancial support of RFFI grant № 13-07-00488

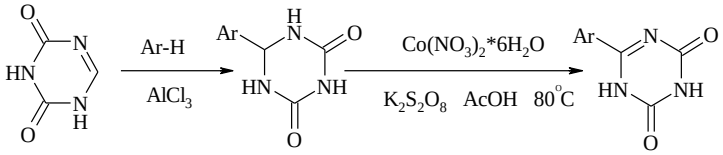
The synthesis of novel annelated and conjugated 1,3,5-triazine-2,4-dione derivatives

T.A. Tseitler[1]
G.V. Zyryanov[1,2]
V.L. Rusinov[1,2]
O.N. Chupakhin[1,2]

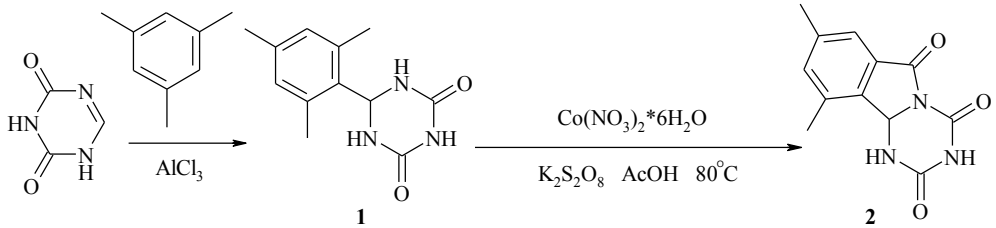
[1] Organic Chemistry Department, Ural Federal University named after the First President of Russia B.N. Yeltsin, 19 Mira St., Yekaterinburg 620002, Russia
[2] I.Ya. Postovsky Institute of Organic Synthesis, Ural Branch of the Russian Academy of Sciences, 22 S. Kovalevska St./20 Academic St., Yekaterinburg 620990, Russia
tseitler85@mail.ru

In lines of π-deficit azines, for example pyrimidines [1] or 1,2,4-triazines [2], the S_N^H methodology is most approachable and it allows to realize one-stage direct functionalization of unsubstituted C-atom from azine by fragments of π-deficit or CH-active compounds to form the products of S_N^H, sometimes with possibility to condense them into polyheteroaromatic systems [3]. Although 1,3,5-triazines are the most π-deficit among azines and they interacts with different C-nucleophiles easily, just recently for the first time we obtained the products of direct arylation for 1,3,5-triazinones when reacting 1,3,5-triazine-2,4(*1H*,*3H*)-dione with arenes in the presence of $AlCl_3$ [4].

Reaction of 1,3,5-triazine-2,4(*1H*,*3H*)-dione with *o*-, *m*- and *p*-dimethylbenzenes (Ar-H) leads to formation of corresponding 6-(dimethyl-phenyl)-[1,3,5]triazinane-2,4-diones. Following oxidative aromatization allow to receive 6-(dimethyl-phenyl)-*1H*-[1,3,5]triazine-2,4-diones with good yields.



Using the excess-electron mesitylene, arylation of 1,3,5-triazinone carried out most efficiently with 75% yield of adduct **1**. Oxidation [5] of adduct **1**, instead of aromatization, leads to formation of 5,7-dimethyl-4,4*a*-dihydro-2,4,9*a*-triaza-fluorene-1,3,9-trione **2**. Likely, the oxidation of one methyl group of mesitylene fragment to carboxylic group take place in this case with following condensation into N-atom of *s*-triazine.



1 **2**

Hence, we obtained and described novel polynuclear heterocyclic compound 5,7-dimethyl-4,4*a*-dihydro-2,4,9*a*-triaza-fluorene-1,3,9-trione **2** for the first time.

[1] Egorov I.N. et al. *Arkivoc*, 2011, **x**, 312-323.
[2] Kozhevnikov V.N. et al. *J. Org. Chem.*, 2003, ***68***, 2882-2888.
[3] Chupakhin O.N. et al. *Uspekhi Khimii*, 2002, **71**, 9, 803- 818.
[4] Egorov I.N. et al. *Tetrahedron Letters*, 2010, **51**, 13, 1717-1718.
[5] Shanmugam P. et al. *Tetrahdron*, 2007, **63**, 12215-12219.

Knowledge extraction from subject-oriented science intelligent system on physical chemistry of radical reactions

V. Tumanov
A. Prokhorov

Laboratory of Information Support for Research, Institute of Problems of Chemical Physics RAS, Semenov ave 1,142432 Chernogolovka, Russia
tve90@yandex.ru

Subject-oriented science intelligence system in physical chemistry of radical reactions is regarded as an intelligent system in the Internet, whose purpose is the collection, storage, verification, retrieval, distribution and production of new subject-oriented knowledge of the physical chemistry of radical reactions [1].

The system is designed to solve the following tasks: search for the experimental data on the reactivity of reactants in bimolecular radical reactions in the liquid phase; search for the calculated data on the reactivity of reactants in the bimolecular radical reactions in the liquid and gas phases; search for the values of bond dissociation energies of organic molecules, as well as enthalpies of formation of radicals and molecules; search for bibliographic references in the database; evaluation of the reactivity of the reagents of radical bimolecular reactions in liquid and gas phases by the thermochemical and kinetic data; evaluation of bond dissociation energies of organic molecules by kinetic data.

Production of new knowledge within the subject domain of the science intelligence system being considered, is performed by means of expert systems, clustering, artificial neural network, fuzzy knowledge bases. Expert Systems, designed for operation in the Internet environment, are constructed as a set of intellectual software agents [2] (Figure).

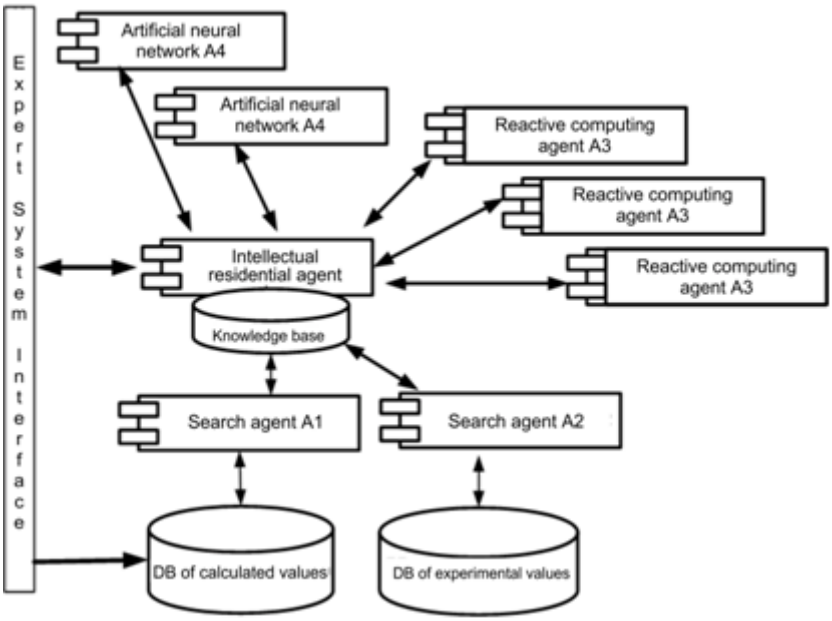


Figure. Schematic multi-agent architecture of the expert system for the evaluation of bonds dissociation energy of organic molecular by the kinetic data of radical reactions

The development and publication on the Internet of subject-oriented science intelligence system will allow the scientific community to create distributed networks for the collection, storage, retrieval, mining, distribution and production of new knowledge in the specialized areas of research and technologies.

[1] Tumanov V. et al. *Information Resources of Russia,* 2010, **5**: 16-21.
[2] Wooldridge, M., Jennings, N. *Knowledge Engineering Review*, 1995, **10** (2): 115-152.

The work is executed at support of the grant of the Russian Foundation for Basic Research 09-07-00297-a.

CONTENT

Отпечатано 16 августа, 2013.
Формат А5. Офсетная бумага мелованная 80 г/м^2.
Печать – Цифровой офсет. 40с.
Тираж 80 экз. Заказ 11.

Ул. Бондаренко, 33–44. Казань, 420066. Республика Татарстан. Россия.
Сот. тел.: 8 917 891 2622.