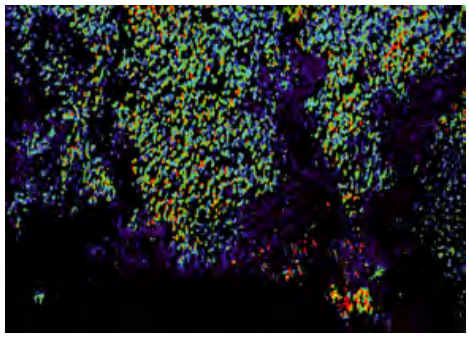
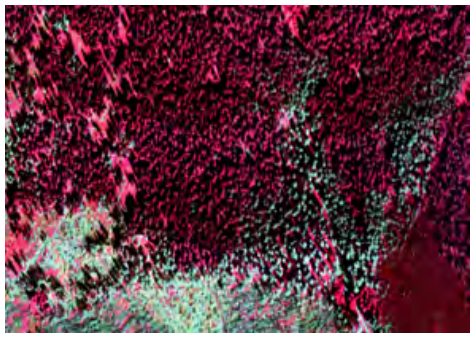
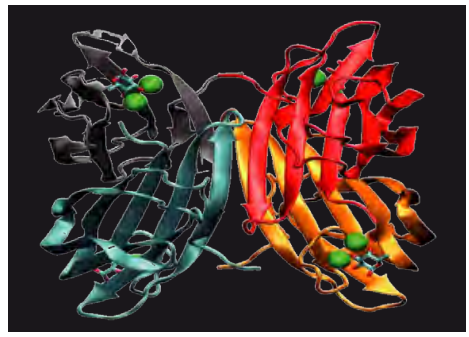
MetaCentrum







Yearbook 2010

MetaCentrum Yearbook 2010

Ivana Křenková, Petr Hanousek,
and Luděk Matyska (Eds.)

MetaCentrum Yearbook 2010

Editors

Ivana Křenková, Petr Hanousek, and Luděk Matyska
CESNET, z. s. p. o.
Zikova 4
Prague, Czech Republic

ISBN 978-80-904689-3-1

Preface

This Annual Report describes research enabled or enhanced through access a distributed computing infrastructure (DCI) as developed and operated by MetaCentrum, the National Grid Initiative (NGI). The Annual Report also covers an overview of the national Grid, the distributed computing and short to midterm storage environment coordinated by CESNET in collaboration with several Czech universities.

The 2010 was the last year of the seven year long CESNET Research Intent that helped to shape national e-Infrastructure in most of the last decade. Lot of activities were therefore dedicated to prepare the new era, as CESNET was in early Spring 2010 declared by the Czech government as operator of Large Research Infrastructure, to be funded since 2011. CESNET position is in a general way defined in the *Roadmap of Large Infrastructures for Research, Development and Innovations of Czech Republic*. Three subjects are listed as completing the future digital infrastructure of the Czech Republic —the Czech e-Infrastructure. They are CESNET, responsible for networks and coordination of distributed computing infrastructure, Center CERIT-SC responsible for cloud infrastructure and research and development of novel ways of use e-Infrastructures, and Center IT4Innovations, responsible for High Performance Computing (Supercomputing) and related research and development. As a result, MetaCentrum underwent in 2010 a transformation to become a respected coordinator of the national grid infrastructure.

New large international projects, co-funded by the European Union under the 7th Framework Program, started in May 2010 to shape the EU distributed computing landscape. CESNET is a member of EGI InSPIRE, a project dedicated to coordinated management and further development of the EU Grid—the EGI (European Grid Infrastructure). CESNET contributes to most workpackages and recently became a task leader of the Distributed Middleware Support Unit (DMSU). CESNET is providing the full backoffice technology and support for the whole EGI collaboration. Also, since project beginning, CESNET represents the Central Europe Grouping at the Project Management Board. While EGI InSPIRE covers operational aspects of EGI, its middleware is developed in the EMI (European Middleware Initiative) project, where CESNET is also contributing. There, further development of the Logging and Bookkeeping service, as well as care of some aspects of the security are CESNET primary focus. Seemingly exotic is CESNET participation in the CHAIN project—Coordination and Harmonization of advanced e-INfrastructures—where it leads a separate workpackage, but this involvement provides a wider view on e-Infrastructures in general and serves as an excellent feedback for development plans at the national level. MetaCentrum users were encouraged to use the large European grid infrastructure, through access to the Virtual Organization for Central Europe (the

VOCE). CESNET also helped to manage the astrophysics virtual community Auger.

The Annual Report is however not about the MetaCentrum only, but its primary goal is to demonstrate the potential of free access and proper use of distributed computing infrastructure. The scientific results described in this Annual Report, in 12 chapters covering work of individuals to very large research teams and even institutions, would not be achieved without access to computing and storage resources, provided through MetaCentrum. The excellent scientific results, journal and top proceedings publications, are a proof of the added value that coordinated distributed infrastructure provides to research teams from very different scientific areas. All this confirms that users benefit from MetaCentrum transformation into the Czech NGI and its wide involvement in the international activities. MetaCentrum is fully prepared to serve the national research community also in the next period, where the bulk of computing resources will not be owned by CESNET, and the role of the coordinator of distributed computing infrastructure will become even more important.

Brno, March 2011

Luděk Matyska
MetaCentrum Director

Table of Contents

Part I

The MetaCentrum Infrastructure

The MetaCentrum Infrastructure in 2010

1 Introduction

The first part of this Yearbook describes the infrastructure operated by the National Grid Initiative MetaCentrum. We give an overview of available hardware resources and software, developments in cluster operation, resulting in significant resource utilisation increase in 2010, and an overview of information resources available to the users. Statistical information about resource utilisation is also provided.

2 Summary

Throughout 2010, the task of the MetaCentrum was to run and further develop the national distributed computing infrastructure—the National Grid Infrastructure—and integrate it into the analogous European Grid Infrastructure, mainly with corresponding international (7RP EU) projects.

The total computing and storage capacity consists of a combination of CESNET sources and computing systems provided by partners, notably Supercomputing Centre Brno at the Masaryk University, Charles University in Prague, and University of West Bohemia in Pilsen. However, other systems were successfully connected to the national grid during the project; Bohemian University in České Budějovice, MZLU in Brno, BUT in Brno, etc. The entire computing capacity and related storage space is available to the students, academic workers and other researchers of the universities, Academy of Sciences institutes and other research organisations free of charge.

Long-term goal of MetaCentrum, grid activity supported by CESNET, is to build and operate the National Grid Infrastructure in the Czech Republic. MetaCentrum includes operation of distributed production grid, development of grid middleware and involvement in international grid infrastructure through participation in corresponding international projects. In recent years, this activity has been covering not only distributed computing resources, but also operation of storage capacities, serving as repositories for computational as well as general data.

With the consideration of the continuing MetaCentrum activity transformation into full featured National Grid Infrastructure (NGI) and the direct succession of CESNET Large Infrastructure, the main 2010's effort was focused on further improving of user support and finishing the actual part of virtualization and planning development. We focused on further development and emphasizing the role of national coordinator of Grid activities. Logical division of MetaCentrum into the national coordinating organisation and the resource provider parts was finished in 2010. MetaCentrum web portal has been transformed accordingly to

these disposals. Details about MetaCentrum as the coordinator of grid activities in the Czech Republic are located at `http://metacentrum.cz`, while information about resources provision—covering also the MetaCetrum role of one of the national grid providers—are located at `http://metavo.metacentrum.cz` web pages. Users and the cooperating institutions have the adequate information about roles of MetaCentrum and the internal responsibilities the organisation have been defined in a more precise way.

The development in 2010 was focused on finishing of conditions for transfer to the new infrastructure of schedulers which provides more autonomy for separate nodes of the national grid. This concept also eliminates the dependence on commercial scheduler PBSPro and replaces it with Torque system and custom-developed components. Development in the field of virtualization also continued towards the principles of cloud computing, providing better reactions on immediate user requests (flexible computing power).

In the area of resource utilisation, the increasing trend of previous years continued. Total usage of almost all resources with unrestricted access was about 75%, where new powerful machines reached long-term utilisation over 85%. Those numbers represent heavy saturation of computational capabilities. Concurrently with high usage of provided capacities, the amount of enrolled publications with acknowledgment to MetaCentrum also grows. High usage and the growing trend of amount and quality of publications gives proof of MetaCentrum's direct user cooperation and support and confirms also the viability of virtualization concept in the environment of national computational capacity. Strong saturation of MetaCentrum's resources shows the deprivation of computational capacities in the Czech Republic. It is a serious argument for the necessity of building new centres like IT4Innovations and CERIT-SC.

3 International Grid Projects

MetaCentrum is fully integrated into international activities and related projects. In 2009–2010, MetaCentrum participated in the preparation of many of projects under the 7th EU Framework Programme and some of them were chosen by committee to be realized. MetaCentrum was a major partner to pan-European grid infrastructure projects throughout the research plan. In the EGEE project series (2004–2010), the MetaCentrum manager held the office of Member of the Project Management Board, representing all of Central Europe. The MetaCentrum director was also co-ordinating the key project EGI_DS (European Grid Initiative Design Study). The latter project designed the organisational and functional structure of the future pan-European grid infrastructure, which is currently being implemented under EGI InSPIRE, in which the MetaCentrum is again involved as a partner. The EUAsiaGrid project lasted till the end of June 2010 and the part of its activities has been continuing since December 2010 under the CHAIN project. With relation to positive valuation of EUAsiaGrid project, the eSCAPE project was prepared and submitted in the second half of 2010 (with CESNET coordination, together with partners from Asia-Pacific region).

The project focused on development of tools, procedures and environments for effective utilization of distributed infrastructure. We used the openings of new projects for acceleration of connections among MetaCentrum activities and those international subjects to take profit from synergies that follow on national and international operations. MetaCentrum connected the CESNET association with the main international activities at the European level so the association holds its position as an important partner on the field of distributed computing infrastructures.

Besides preparation of new projects, MetaCentrum has been participating in the whole range of national and international projects in the area of grid infrastructure development. A list of the most significant projects is given in Table 1.

4 Hardware and Software Resources in MetaCentrum

This section describes hardware and software portfolio available in MetaCentrum.

4.1 Computation Resources

Users of MetaCentrum VO were able to compute on more than 1500 CPU cores. The following table shows the number of CPU cores in MetaCentrum during the past years.

In the first half of 2010, the number of CPUs used in MetaCentrum raised. The new Tarkil cluster (CESNET) was put into operation, being the replacement for Skurut cluster (CESNET) in Prague. We renewed the Ajax server in Pilsen (UWB) and expanded clusters Alela in Brno (FEEC VUT) and Hermes in České Budějovice (JCU). We put off the operation clusters Perian17-68 and Perian69-76 in Brno (both PřF MU). In comparison with the first half of the year, number of CPUs negligibly decreased. Perian77-86 and Perian87-96 clusters were dislocated to the new server room under administration of cluster owners, although staying connected into MetaCentrum.

New institutions were interested in connecting their resources into the MetaCentrum infrastructure, namely University of Tomáš Baťa in Zlín and University of West Bohemia in Pilsen. A new cluster equipped with graphical coprocessors (GPU) for computing acceleration will be put into operation in the UWB in 2011.

MetaCentrum offers a wide range of hardware resources. The portfolio includes popular SMP servers with 16 or 32 cores and 64 to 256 GB memory (`eru`, `aule`, and `manwe`). Typical "high density" clusters consist of many 4- or 8-CPU servers connected with fast networks: 1 Gb/s (Gigabit Ethernet), 2.5 Gb/s (Myrinet) and 20 Gb/s (Infiniband). List of clusters and their parameters as of the end of 2010 is shown in Table 3. Up-to-date number of machines, CPUs and their utilisation is displayed on the MetaCentrum portal[1].

[1] `http://meta.cesnet.cz`

EGI InSpire (5/2010–4/2014) http://www.egi.eu/	European Grid Initiative Integrated Sustainable Pan-European Infrastructure for Researchers in Europe. Project continues the transition towards the sustainable pan-European e-Infrastrusture initiated within series of EGEE projects through support of Grids of high-performance and high-throughput computing. The primary partners for EGI-InSPIRE are the national grid infrastructures from practically all European countries, Russia, Southeast Asian countries and the US. The project co-ordinator is EGI.eu, an organization established in Amsterdam in February 2010 and controlled by a consortium of NGIs. The Czech Republic is part of the Central European group, which elected Prof. RNDr. Luděk Matyska, CSc., the Association representative, as a member of EGI InSPIRE Executive Committee.
EMI (5/2010–4/2013) http://www.eu-emi.eu/	The European Middleware Initiative. Under the closely related EMI project, the Association continues developing grid middleware, specifically the Logging and Bookkeeping service, as well as certain components associated with operational security. It associates representatives of three most important grid middleware systems being developed in Europe—ARC, gLite and UNICORE. The aim of the project is to make and further develop a consolidated set of middleware components designed for the EGI grid, PRACE and possibly also other DCI (Distributed Computing Infrastructures).
CHAIN (12/2010–11/2012) http://www.chain-project.eu/	Co-ordination and Harmonisation of Advanced e-INfrastructures. The objective is to connect regional grid infrastructures with the EGI grid. Here, regions refer to areas outside Europe, such as Asia, Latin America and Africa. The EU has supported the development of EGI-compatible infrastructures in all those regions as separate projects (including the EUAsiaGrid, for instance). Under the CHAIN project, the activities will be integrated and co-ordinated at a higher level in order to ensure truly global, boundary-free cooperation of scientific teams making use of distributed computing infrastructure.
EPIKH (3/2009–3/2013) http://www.epikh.eu/	Exchange Programme to advance e-Infrastructure Know-How. The main aim of the project is to reinforce the impact of e-infrastructure in scientific research defining and delivering stimulating programme of educational events, including Grid School and High Performance Computing courses.
EUAsiaGrid (4/2008–6/2010) http://www.euasiagrid.org/	Towards a common e-Science Infrastructure with the European and Asian Grids. It aims to promote awareness of the EGEE infrastructures, middleware and services in the Asian countries, based on results of the EGI_DS projects.
EGEE III (5/2008–4/2010) http://www.eu-egee.org/	Enabling Grids for E-science in Europe. EGEE III is the third in a series of grid infrastructure projects. Its goal is to extend and maintain pan-European grid infrastructure available to European scientific community. MetaCentrum representative served also as the representative of the Central Europe region at the Project Management Board.

Table 1. International Grid Projects with MetaCentrum (CESNET) participation.

year	2004	2006	2008	2009	2010/1	2010/2
# of CPU	262	600	1200	1468	1560	1522

Table 2. Numbers of CPUs.

Machine	**CPU cores**	**description**	**memory per node**	**owner**
Acharon	16	SGI Altix 350	48 GB	UK
Ajax	8	SGI Altix 350	72 GB	ZČU
Alela	96	PC Cluster	8 to 32 GB	VUT
Dali	10	SGI Onyx 350	16 GB	UK
Eru	64	Cluster SUN	132 to 256 GB	CESNET
Hermes	100	PC Cluster Xeon	16 GB	JČU
Hydra	12	PC Cluster Xeon	1 GB	KIV/ZČU
Konos	120	PC Cluster AMD	2 to 8 GB	KMA/ZČU
Loslab	24	PC Cluster AMD	4 GB	LL/MU
Manwe	112	Cluster SUN X4600	32 to 128 GB	MU, CESNET
Nympha	160	PC Cluster HP	16 GB	CESNET
Orca	72	PC Cluster AMD	8 GB	NCBR/MU
Perian	190	PC Cluster SGI/HP	2 to 8 GB	NCBR/MU
Quark	62	PC Cluster Xeon	2 to 18 GB	MU
Skirit	212	PC Cluster Xeon	1 to 8 GB	CESNET, FI/MU
Tarkil	232	PC Cluster	24 GB	CESNET
Clusters only available to their owners				
Wood	16	PC Cluster	4 GB	MZLU
Other		dedicated clusters	n/a	various

Table 3. Machines and clusters connected in MetaCentrum, 1522 CPU cores in total.

MetaCentrum resources are physically hosted in the following locations: Masaryk University in Brno (Supercomputing Center Brno), Charles University in Prague (Supercomputing Center UK), University of West Bohemia in Pilsen (Westbohemian Supercomputing Center ZČU), University of South Bohemia in České Budějovice (Faculty of Science's cluster JČU), University of Technology in Brno (Faculty of Electrical Engineering and Communication VUT), Institute of Wood Science of the Mendel University in Brno (Supercomputing Center MZLU), and CESNET headquarters in Prague.

The clusters are connected with a network based on the CESNET2 backbone. The clusters accept jobs submitted through the common scheduling system and share data with distributed file systems.

The network has following independent international connections:

- 10 Gbps to GÉANT, used for academic traffic
- 2.5 Gbps to Telia, used for commodity traffic

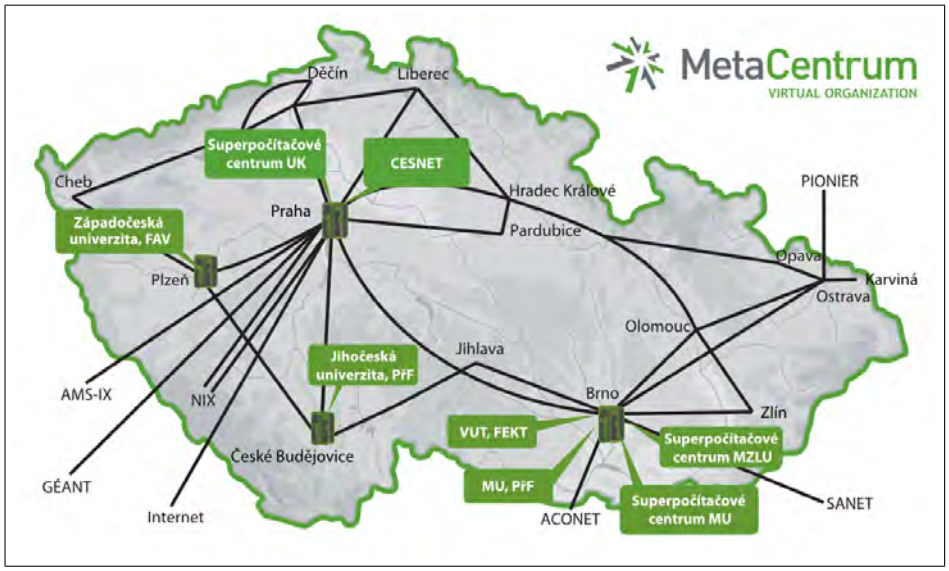


Fig. 1. MetaCentrum major sites

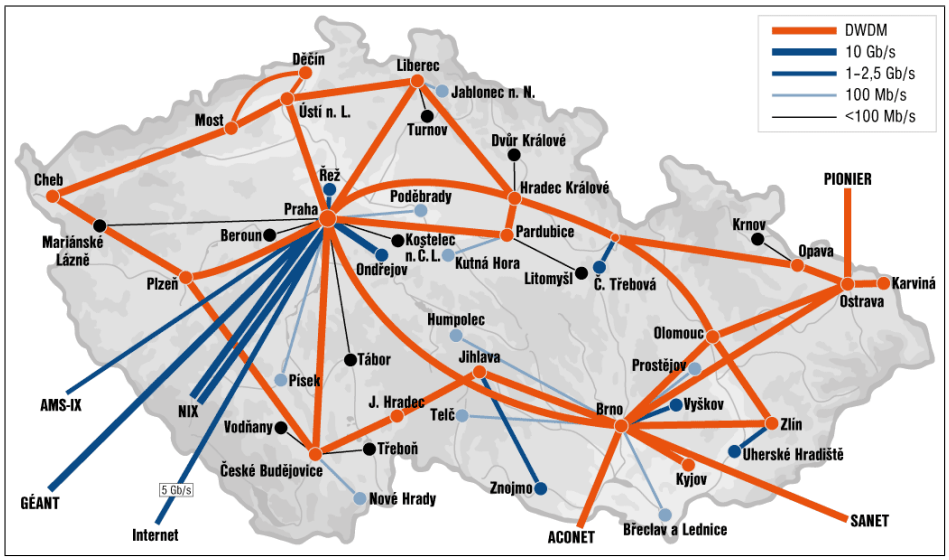


Fig. 2. CESNET2 network topology in December 2010

– 10 Gbps to SANET, academic network of Slovakia
– 10 Gbps to ACOnet, academic network of Austria
– 10 Gbps to PIONIER, Polish optical academic network

– 10 Gbps to NetherLight/GLIF for experimental traffic; this line is connected to our experimental optical network named CzechLight, so it is not depicted in the map of CESNET2 above

In addition to these international lines, external connectivity of CESNET2 is enhanced by powerful peering connections:

– 2×10 Gbps to NIX.CZ
– 1 Gbps to AMS-IX

Network provides multicasting capability and is a part of the MBone network. Videoconferencing services based on the MBone technology are offered to the network users. IP version 6 is provided as a standard service.

4.2 Data Storage

MetaCentrum runs several data storage systems with both local and global access. Each cluster node has a /scratch space (sized in tens to hundreds GBs) to store intermediate results of currently running computations. Home directories are usually local for a cluster and realised by a local NFSv3 or NFSv4 server.

Main storage capacities are available through the NFSv4 protocol. In cooperation with Supercomputing Centre Brno (SCB, Masaryk University) financing, we have extended the disk array capacity of 24 TB to whole current capacity of 124 TB. Formatted capacity of disk storage was raised from 44 TB to 77 TB. Measurements show that the NFSv4 central server is not suitable to centrally provide cluster home directories; data access from Pilsen to Brno is about 10-times slower than local access which is caused by latencies of protocol control messages. Despite this drawback, the NFSv4 volume is suitable to keep long-term data that is staged in computation nodes.

MetaCentrum still offers a global filesystem based on AFS. Software distributions are accessed via AFS, but it can be deployed directly by users to benefit from fine-grained access control possibilities.

For backup purposes, MetaCentrum uses tape libraries Overland Storage NEO8000 with 500 tape positions based on LTO-3 technology. Two identical libraries are installed with total on-line capacity of 400 TB (uncompressed), one at the University of West Bohemia and one at the Masaryk University, Supercomputing Centre Brno. The libraries are managed with a combination of EMC Legato NetWorker suite and in-house developed software. The distributed backup environment provides a robust and fault-tolerant solution that could survive even catastrophic events such as a complete destruction of one of the hosting sites.

4.3 Application Programs

Software portfolio of MetaCentrum covers wide spectrum of application programs. The following list of installed software is not complete, it serves merely as an overview for both current and potential users. It is roughly divided into several problem areas.

Computational Chemistry & Molecular Modelling

- *Amber*
- *Babel*
- *Gaussian/GaussView*
- *Molden*
- *Gromacs*
- *VMD*
- *Gamess*
- *deMon*
- *PC GAMESS/Firefly*
- *MolPro*
- *Tinker*
- *NAMD*
- *VASP*
- *Autodock Vina*

Technical and Material Simulations

- *ANSYS (including module LS-DYNA)*
- *Fluent*
- *MSC.Marc*
- *Open Foam*

Mathematical and Statistical Modelling

- *Maple*
- *Matlab*
- *SNNS*
- *R*

Development Tools and Environments

- *SGI Development environment*
- *PGI CDK*
- *TotalView*
- *Vampir*
- *Paradise*
- *SICStus Prolog*
- *Wine*
- *GCC*
- *Grace*
- *Numpy*
- *Python*
- *Scientific Python*
- *Scipy*
- *JDK*
- *Lisp-CMUCL*

Structural Biology and Bioinformatics

- *MrBayes*
- *QUEEN*
- *X-PLOR*
- *CS-Rosetta*

Software packages in MetaCentrum can be installed on user request. More details about the application programs and licence policies can be found at the MetaCentrum web site.

The new version of popular Gaussian09 software (A.02) was obtained in 2010. It is the second most popular commercial application in MetaCentrum. We also have improved the documentation of JDK application and installed new software RAxML, Molden 64b and new versions of Python and SciPy on user requests.

4.4 Security

We have continued active development of Pakiti service for monitoring of security updates. Pakiti is a free software for monitoring software packages, mainly for identifying security problems with insufficiently applied patches. During the year we succeeded to upgrade Pakiti from pilot operation to a regular service which is used for daily monitoring of MetaCentrum and also for the whole EGI infrastructure. Mainly in the EGI heterogeneous environment, Pakiti was the crucial step to decrease update times and therefore for better overall security, as attacks to unpatched vulnerabilities represent serious risks in IT systems.

The federation identity was the next realm we took into presence. We continued in technical support of electronic catalogue of pathological images. This service is recently connected to 14 international academic federations whose users can use for the access to catalogues their common identity. The service is worldwide unique in the number of supported international federations.

We have been invited to the Moonshot project which is also targeted into usage of federations. Moonshot is GN3 and JANET (UK) supported activity with main purpose of supporting federate mechanisms in the world outside web. Typical example of the support is the implementation of federated SSH where user does not need a unique password for SSH server but can use his identity in federation. Proposed architecture is similar to Eduroam system. The standardization in Moonshot project has a great emphasis and its participants actively develop the IETF specifications in recently created group IETF ABFAB (Application Bridging for Federated Access Beyond). MetaCentrum was invited to this project on the basis of previous works on the HTTP authentication field. During the year, we developed the Apache and Firefox modules which supports Moonshot federated infrastructure and we plan to focus on using this architecture in grid environment.

We also continued development of the authorization service. ACL administration was rewritten in the scope of allowing easier and more intuitive manipulation with privilege lists. Resource access administration was also one of the most important tasks due to developing of the system for user and resource administration Perun3.

We continued operation of CESNET CA registration authority and moved this service for new CESNET CA3 authority. In operation is also the RADIUS server for MetaCentrum Eduroam users and for identity providers. We are active members of EGI CSIRT team and we closely cooperate with CESNET CSIRT team.

5 Operation of the Infrastructure

5.1 New Services for Users

New services have been introduced in 2010.

– We established the user wiki with kerberos authentication where users can share their knowledge, tips, manuals, howtos and recommendations with

other users. Publicly accessible beginner's guides for using MetaCentrum resources, data storage tutorial, VOCE basics, and many other materials have been prepared.
- A tool for data archiving to the tape library is accessible via web interface. Users can backup important data on their own, completion of backup is notified by e-mail. Archived files can be restored to disks on demand.
- A service of SCP access to NFSv4 shared volumes was introduced. Direct SCP access to the data storage is comfortable for users (it is not necessary to wait for an interactive job run) and it does not require installation of NFSv4 clients on the user's end station, which is especially suitable for Windows users.
- New version of popular Gaussian09 software (A.02) was installed.
- The MetaCentrum web portal was split into the NGI and virtual organisation parts.
- User account renewal rules were unified in MetaCentrum VO and VOCE.

5.2 RT System

User interactions and solving of their problems proceeds mainly through the trouble ticketing system (Request Tracker, RT). In the first half of 2010, we installed the new version of RT system, which solved the problem with attachments and improved the security. Due to spam filter tuning we realized the decrease of amount of improper requests, we do not send users anti-receipt messages and spams are quarantined.

We provided special queues for interested user groups, for example NCBR group or ESFRI ELI project.

During the transformation of MetaCentrum into the Czech NGI we connected our RT with the system required in EGI—GGUS. About one fifth of all requests in 2010 was received through this connection.

In the observed period, 1469 new tickets were created by end users (usually the question or a problem report) or by using of MetaCentrum's portal on the basis of users needs (establishing or prolonging of an account, quota raising requests, etc.) or by administrators. There were 249 requests concerning international grid environment.

5.3 Cooperation with Important User Groups

Positions for direct user support with the aim of finding user groups with excellent scientific results were founded. These user groups were provided with support for optimizing their jobs running in MetaCentrum and optimizing of data transport and storage. We also provided them the support for migration from national to international grid environment. The factual result of such cooperation with one of the groups is the technical report [4].

The user support group was focused on solving two basic user problems—processing large number of jobs and manipulating of large amount of data. With

the experience from users the support group developed the manual pages covering the frequent problems with running jobs and dealing with file systems available in MetaCentrum.

We supported four important user communities. Two of them acknowledged each other and they exchanged their experience with digitizing and processing of the extensive picture data (JPEG2000 conversion) and also they shared the necessary tools. In cooperation with other two communities we established in MetaCentrum the Diane framework which is used for optimization of running of large number of jobs in grid. After testing in MetaCentrum we plan the deployment of Diane framework in EGI (VOCE) as well.

5.4 Information Services

MetaCentrum, a CESNET department responsible for coordinationg and managing grid activities has been astablished as fully-fledged National Grid Initiative (NGI) in the Czech Republic with connection to the international environment. Information presented on the portal has been split to the two following portals:

- Basic information gateway of MetaCentrum NGI[2] with general information about EGI and NGI objectives, services, events, virtual organizations, etc.
- MetaCentrum VO operation portal[3] including operation information about catch-all virtual organization MetaCentrum, rules of usage a documentation, accounting information, etc. The portal provides both general information for casual visitors and specific information for registered users (authentication may be required).

[2] `http://metacentrum.cz/`
[3] `http://metavo.metacentrum.cz/`

In addition to general documentation on portal, new MetaCentrum user wiki[4], started in Spring 2010, serves for sharing knowledge, tips, manuals, howtos and recommendations among MetaCentrum users.

5.5 User Courses, Workshops and Dissemination

User support encompasses wide range of related tasks: informing users about the current state of the centre, addressing user problems, maintaining collaboration with user communities, monitoring resource availability, providing evaluated feedback to the operation activities, and many others.

- One day's Grid Computing Workshop was held in Prague to discuss current grid status and the implementation of new services. The meeting was held on 15 October and was attended by more than seventy experts. The main goal of the meeting was to inform current and potential users of high-performance/high-throughput computing about the possibilities available for solving research problems at national as well as international level.
  The seminar included a keynote lecture on the challenges of scaling high-performance computing applications to peta-scale levels, by Jean-Pierre Panziera, director of High Performance Engineering at Bull Extreme Computing, which co-organised the event.
  The MetaCentrum seminar was an excellent opportunity for users and administrators to meet together in a face to face discussions. It helped both sides to get much better understanding of what the infrastructure is capable of and what its users are expecting to get from it.



- 8th Discussions in Structural Molecular Biology (March 2010, Nové Hrady)
- ISGC 2010 (March 2010, Taipei, Tchaj-wan)—International Symposium on Grids and Clouds—project EUAsiaGrid
- Science Café grid project (May 2010, Prague)—EGI, NGI
- Training activity (May 2010, Pardubice)—quantum chemical methods and modelling of properties of new materials on Grid
- Annual report 2009 (July 2010) with more than 22 user contribution written by individual users as well as large scientific teams covering practically all the scientific areas that use the distributed computing infrastructure.

[4] `http://meta.cesnet.cz/wiki/`

- European AFS & Kerberos Conference 2010 (September 2010, Pilsen)
- EGI technical forum (September 2010, Amsterdam)—major event within the EGI community (EGI InSPIRE project). It brings together European distributed computing projects and their collaborators in academia and businesses, from around Europe and around the world. The major theme of the meeting, achieved through technical sessions, a demonstration and exhibition area, networking space and events, is to establish collaborations between the new and the current European Distributed Computing Infrastructure projects to meet the needs and requirements of the research community. The European Grid Infrastructure today supports thousands of researchers and scientists around the world, helping them to meet their daily e-science challenges.
- Training activity (November 2010, Pilsen)—usage of MATLAB Parallel Toolbox for static and dynamic optimization.
- Contribution to the autumn issue of EGI Newsletter Inspired (November, 2010).

Many presentations were also given to the grid development communities at various international and national conferences.

6 Infrastructure Usage—Operational Statistics

We have continued in further development of computing the operational statistics about resource usage in virtualized environment. Data are mined from records of PBS job scheduling system, from the user identity management system Perun and from the logs of computational nodes. The details about the methodology of statistics generation are described in wiki.

Data collected from monitoring and accounting tools is available at the MetaCentrum portal[5]. The statistical data collecting must take infrastructure complexity into account. It is necessary to keep records on clusters, machine membership in clusters, physical and virtual machine outages, user reservation of resources, coexistence of virtual machines on a single physical computer etc.

6.1 Basic statistics—MetaCentrum VO

To the end of 2010, MetaCentrum recognizes 420 active users (it was 371 users in 2009). User accounts were prolonged for 281 users and we gained 139 new users. At least one job was run by 243 users. Some of users uses their involvement in MetaCentrum only for access to storage capacities and other services, some of them were never active.

Strong and continuously growing user interest in MetaCentrum is demonstrated by significant increase in usage of computational resources managed by MetaCentrum. During the year 2010 almost 600 thousand of jobs were computed (500 thousand in 2009), consuming more then 6.4 millions of CPU-hours (4.5

[5] `http://metavo.metacentrum.cz`

millions in 2009). For the first half of year it was more than 3 millions CPU hours in 190 thousands of jobs. The average workload of all MetaCentrum's machines oscillated for the whole year around 75%—this is the weighted average for MetaCentrum VO in relation with the number of CPUs.

period	# of jobs	CPU hours
2009/1-6	125 thous.	1,5 mil.
2009/7-12	364 thous.	3,0 mil.
2010/1-6	190 thous.	3 mil.
2010/7-11	400 thous.	3,4 mil.

Table 4. Significant increase in usage of computational resources.

The purpose of the infrastructure is to allow users to do research, therefore, scientific papers referring to MetaCentrum have been collected since 2008 and users with registered scientific results are prioritised. Total of 64 papers published in international journals and prestigious international conferences, with acknowledge to MetaCentrum, were registered in 2010 (37 papers in 2009).

Disk array is available for 388 MetaCetrum's users and 289 users have their data stored there. User data lies on approximately 60 TB (80%) in 80 millions of files (last year it was only 30 millions of files). The average size of one file is 741 kB.

250 users have stored less than 100 thousands files, 23 users have stored less than 1 million files and 16 users have more than one million files. At least 1 GB of data are stored for 144 users which is the double amount than in 2009. Anonymized chart of users with most data are shown in tables 5 and 6. The user with the highest amount of files is not identical with user with the highest amount of data.

Compared to last year, the archives of Bohunice IT section campus grew from 3.8 to 9.6 TB and the archive of Moravian library from 4.6 TB to 6.8 TB.

# files on disk array	
user A	21 578 425
user B	12 691 831
user C	8 655 936
user D	4 884 061
user E	4 449 573

Table 5. Chart of users according to number of stored files.

data volume on disk array	
user U	10,6 TB
user V	9,6 TB
user X	6,8 TB
user Y	4,8 TB
user Z	4,0 TB

Table 6. Chart of users according to data amount.

6.2 Influences of Changes in Planning System and Queue Management

Significant increase in usage of computing resources was observed in 2010. Users consumed about 6.4 million CPU-hours in 600 thousands jobs (4,5 million CPU hours in 2009).

We adapted the PBS planning system to be more effective with using available computational resources to satisfy the most users. All of these adaptations are concurrently built in new environment of cooperating scheduler Torque and will still be available for users. New scheduling system was put into operation for experimental purposes during the year. The stable use of Torque is planned to the beginning of 2011.

In the beginning of the year, we listened to users with high number of jobs who complained about the limits for maximum number of running jobs for one user. We established the new queue wit low priority called "backfill" suitable for large amount of jobs with duration time less than 24 hours. Jobs in this queue are filling the vacant space of machines and can be in case of need suspended or broken from our side. This queue is becoming popular among users despite the fact that jobs in this queue are just for filling the machines vacant time and have a low priority of running and the possibility of suspending. In the second half of the year this queue became the one with the highest number of computed jobs (Fig. 3). We suppose that this queue is used mostly by users at first places in computing because of no limit in the number of running jobs.

We have adjusted the conditions for accessing privileged queues "privileged" and "privileged@arien". New rules allow users to gain access to queues for submitting of the user manual for applications. Old rules required only three publications with acknowledgement of MetaCentrum. Currently there are 25 users with access to these queues. According the statistics are these queue used mainly for multiprocessor jobs with high memory demands which would wait for a long time in a different queue. Machines sufficient to run such jobs are quite rare.

In the first quarter of year we observed long waiting times of jobs in both privileged queues, reaching up to 25 days. The whole time of computed CPU time in these queues was in rank of ones of percent of MetaCentrum's computed time. At the beginning of May we raised the maximum amount of concurrently running jobs in both privileged queues and we limited the access of "backfill" queue on powerful machines (ajax, aule, eru, manwe). Due to these disposals

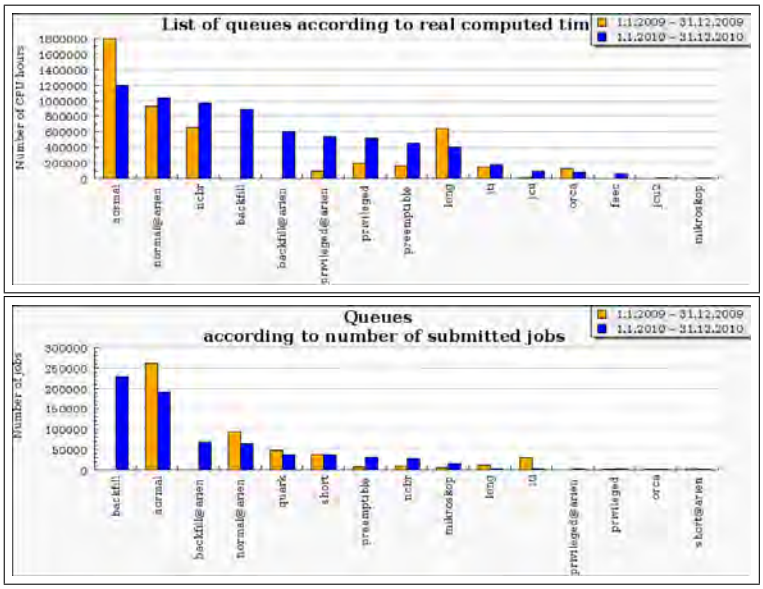


Fig. 3. Queues according to number of jobs and computed CPU time

has significantly increased the rate of jobs launched till one minute, what is shown in Fig. 6.

In April, we unlocked due to a new policy access to Quark cluster for all of the MetaCentrum users using the low priority queues "short" and "backfill".

During the second half of 2010, utilisation of most of MetaCentrum clusters with open access reached 75–90% (Fig. 7), which is an excellent result for an infrastructure running heterogeneous jobs, meaning practically a complete saturation, indicating strong user interest in the infrastructure. This also corresponds to job waiting times which generally increased. In order to prevent job waiting times to grow into unacceptable levels, it is obvious that further investment into the infrastructure is necessary.

The overview of time usage in various applications is shown in Fig. 8.

Fig. 6 displays numbers of jobs according to waiting times from submitting to the start of execution, dividing the halves of the year for comparison. While majority of the jobs waits no more than a couple of minutes, it can be seen that the infrastructure utilisation increase in the second half of the year caused several thousands of jobs wait days.

Running times of jobs Fig. 9 offer an interesting comparison, too. Extremely short jobs usually indicate a configuration and/or job submission error (users with excessive amount of such jobs are offered help by the User Support). The graph doesn't distinguish sequential and parallel jobs, but in general, jobs running more than several hours tend to be parallel (4–32 CPUs). Despite seemingly long job waiting times caused by high resource utilisation, MetaCentrum still makes user (real time) waiting for results shorter: a five-hour 16-CPU job would

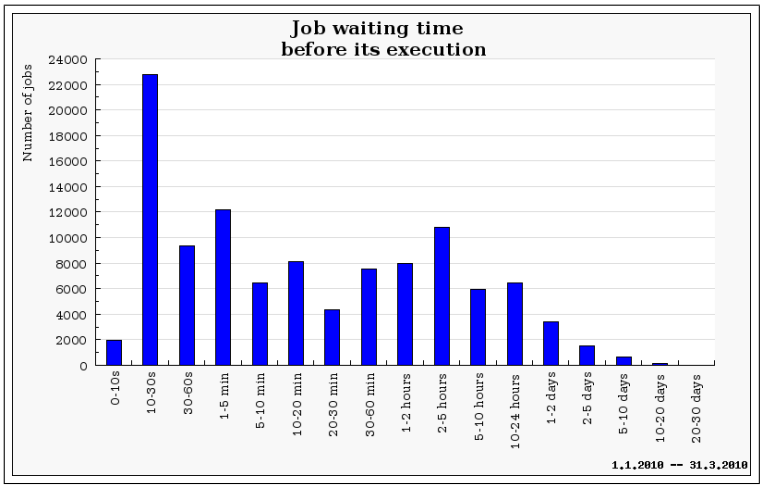


Fig. 4. Jobs waiting in privileged queues in the first quarter and in the rest of year.

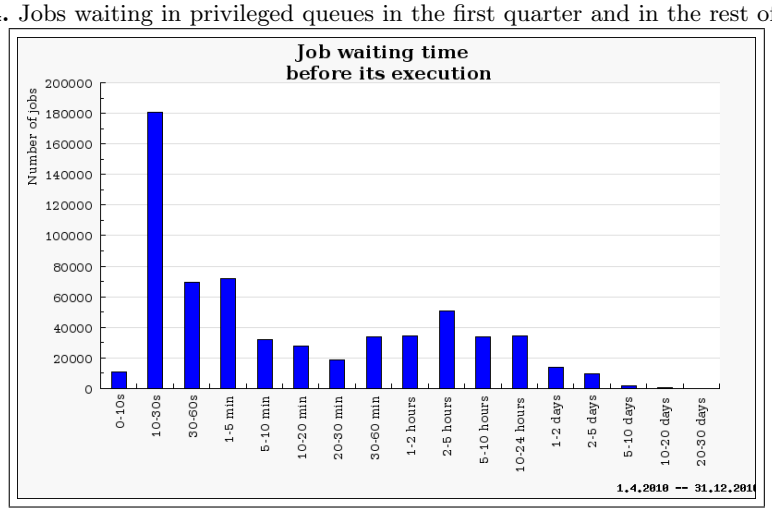


Fig. 5. Jobs waiting in privileged queues in the first quarter and in the rest of year.

last more than 3 days on a single CPU system and nearly a day on a 4-CPU system.

The participation of multi processor jobs is shown at charts in figure 10. While there is a domination of single processor jobs, most of the computed time is used by multi processor jobs. These results are the proof of MetaCentrum orientation on multi processor machines.

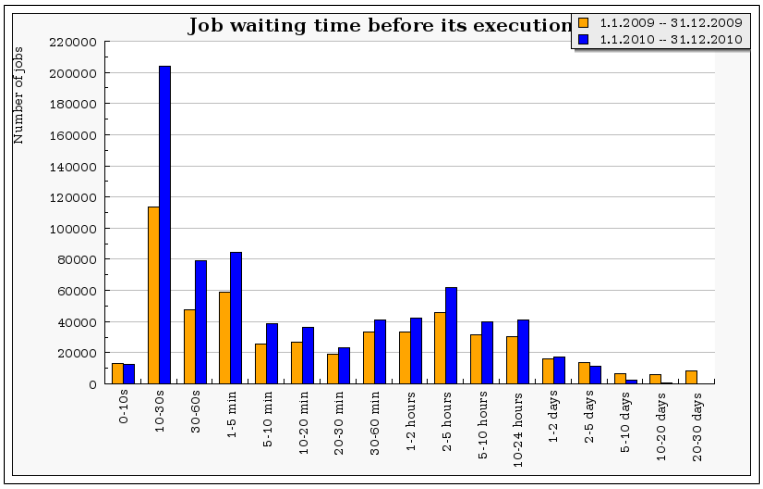


Fig. 6. Comparison of job waiting time in 2009 and 2010

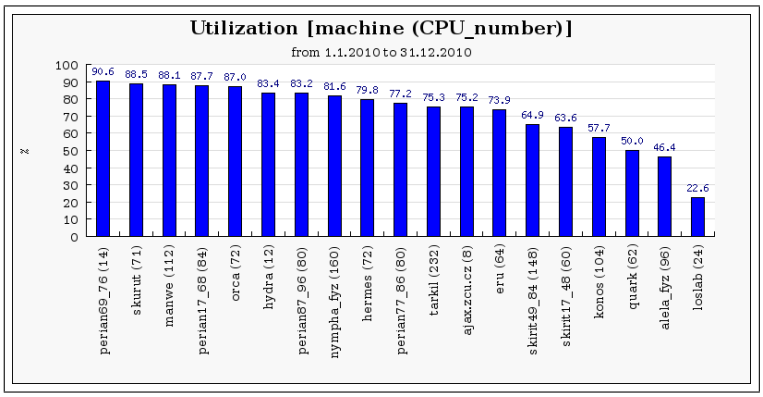


Fig. 7. The average cluster workload

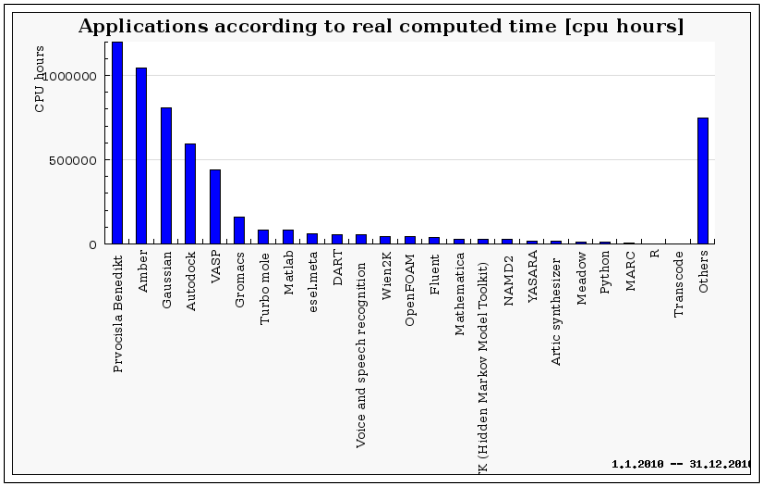


Fig. 8. CPU time in applications

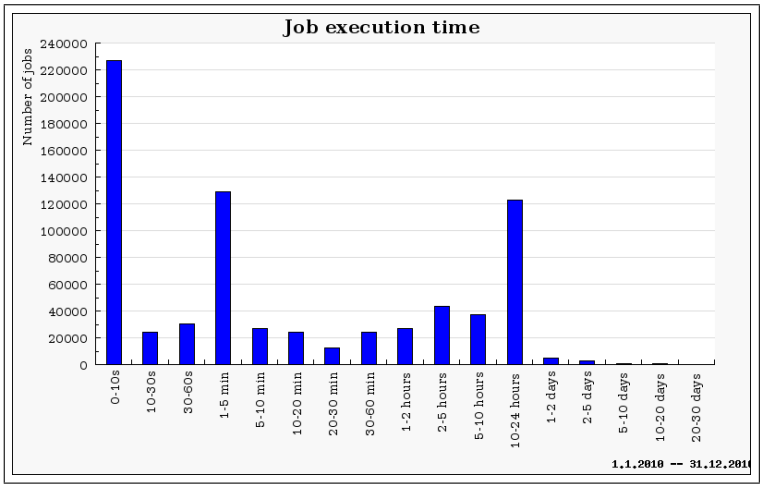


Fig. 9. The running time of jobs in CPU hours

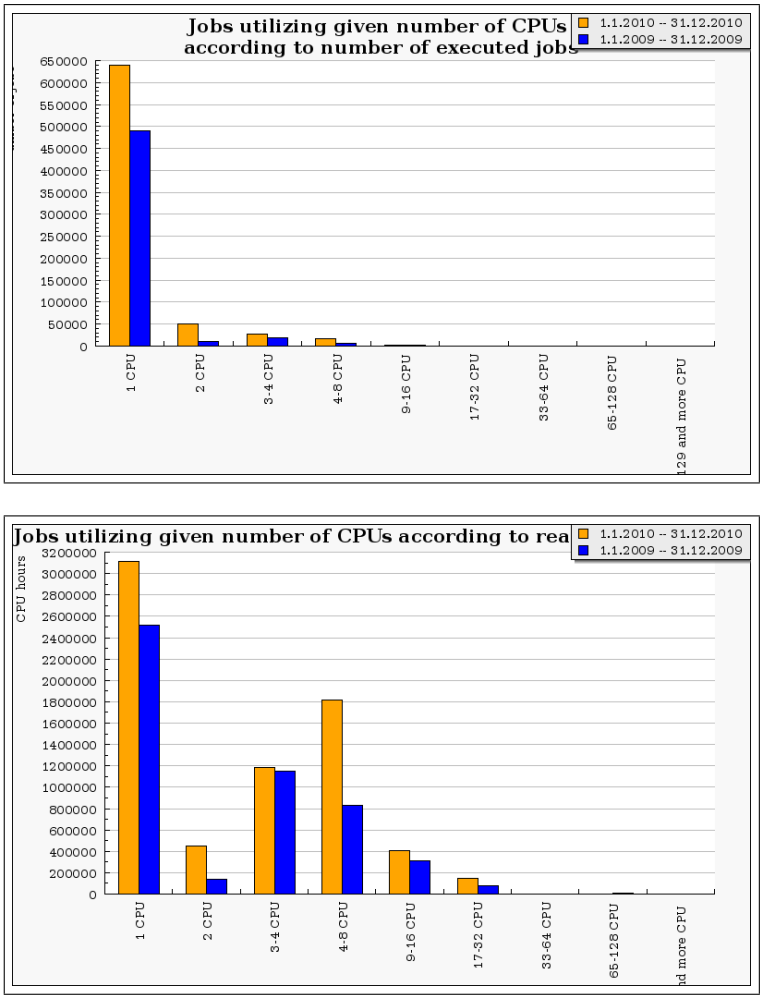


Fig. 10. The number of jobs and counted time with differentiation of number of used CPUs

6.3 Resource Utilization Graphs

Graphs in figures 11 – 20 shows hardware utilisation in 2010. The base (100%) is the total number of available CPU-core-seconds minus the CPU-core-seconds of machines under maintenance. The values are CPU-core-seconds of running jobs and reservations. Complete statistics can be found at the MetaCentrum portal.

The most demanded clusters are the most powerful multiprocessor ones with sufficient memory like Nympha, new Skirit, Tarkil and Manwe. Tarkil cluster (28 eight core units) was put into operation in March. Despite this late time of operation it took the first place in computed CPU hours in 2010. Fig. 7 shows the clusters sorted according to whole computed CPU time and according to amount of computed jobs. The CPU number is not took in account.

Growing interest for support of parallel jobs can be demonstrated on workload of 16 processors Manwe cluster (Fig. 11), which oscillated during the whole year near 90% while in 2009 it was about 85%.

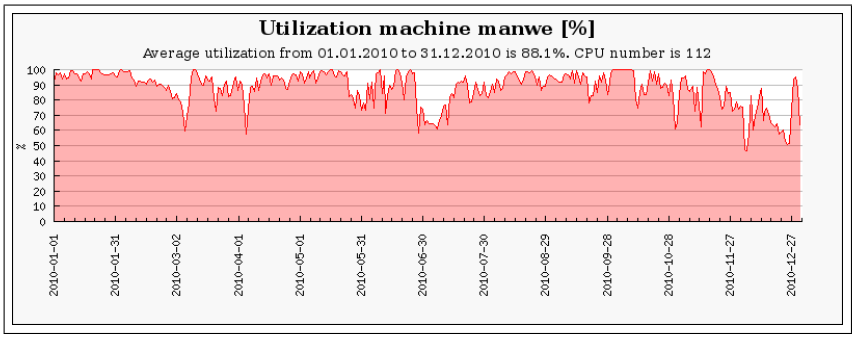


Fig. 11. Workload of Manwe cluster (SCB/MU, CESNET).

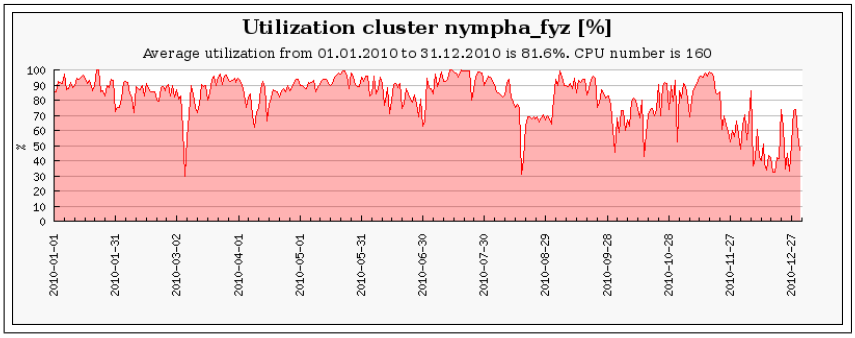


Fig. 12. Workload of Nympha cluster (CESNET).

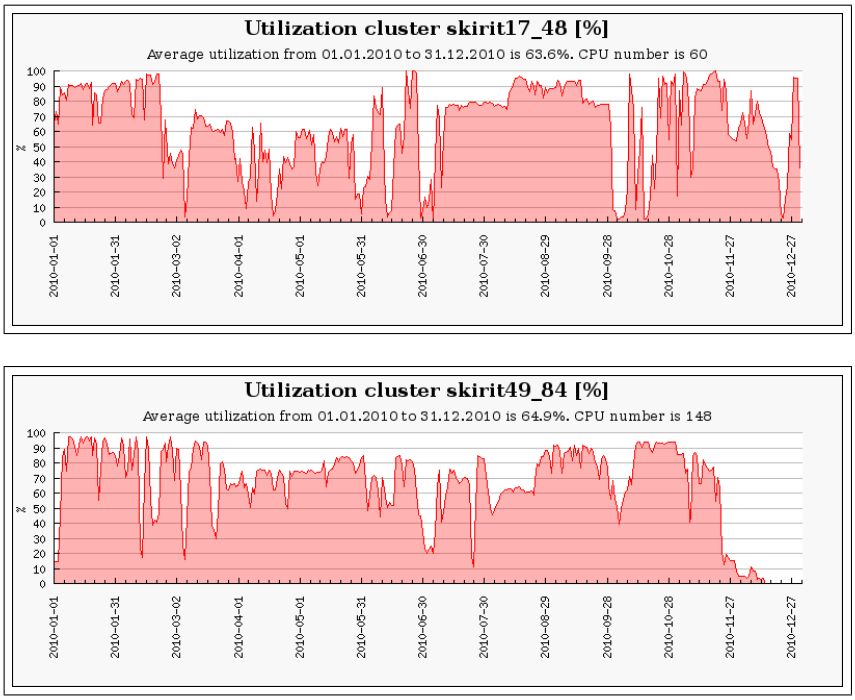


Fig. 13. Workload of Skirit cluster (SCB/MU).

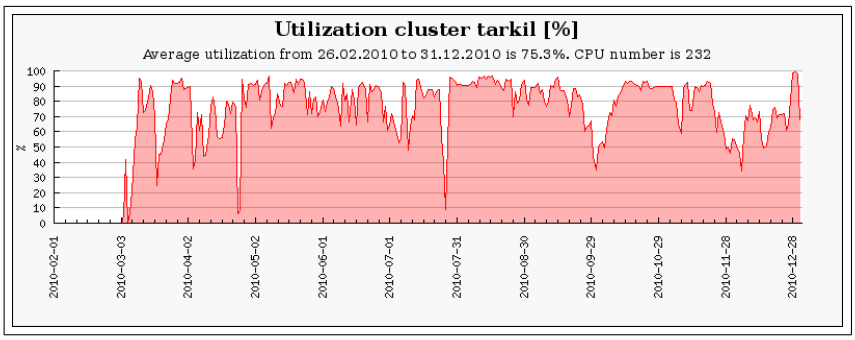


Fig. 14. Workload of Tarkil cluster (CESNET).

The advantage of open access can be demonstrated by utilisation of clusters Orca and Perian. Both of them are owned by the NCBR research centre. Originally, they were dedicated to their owners only, since summer 2009 they have been accessible to all MetaCentrum users. Jobs from priority queues accessible only by the owners can preempt task from the other queues. Owners still have privileged access to their resources without any loss of perceived access quality. Cluster workload raised from 20% before equalization (in 2009) to approx. 80%

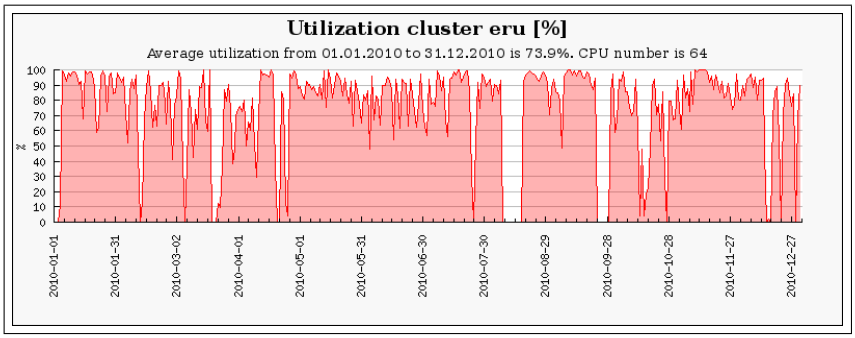


Fig. 15. Workload of Eru machines (CESNET).

after that (Fig. 16). Similar results can be seen at Perian cluster (18) belonging also to NCBR.

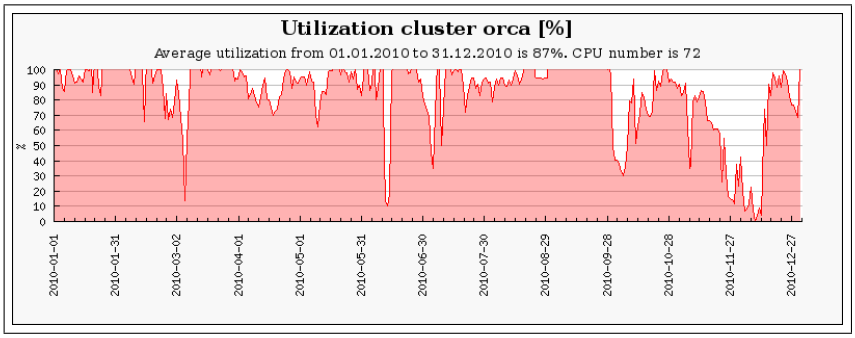


Fig. 16. Workload of Orca cluster (NCBR/MU).

In April, a part of the Quark cluster (Fig. 17) was made accessible to all MetaCentrum users in less priority queues. Also with this cluster we noticed significant growth of its average workload which exceeded 60% in the second half of year.

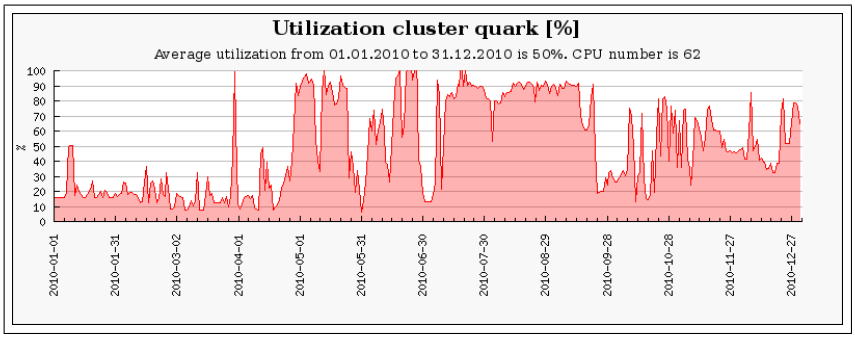


Fig. 17. Workload of Quark cluster (FI/MU).

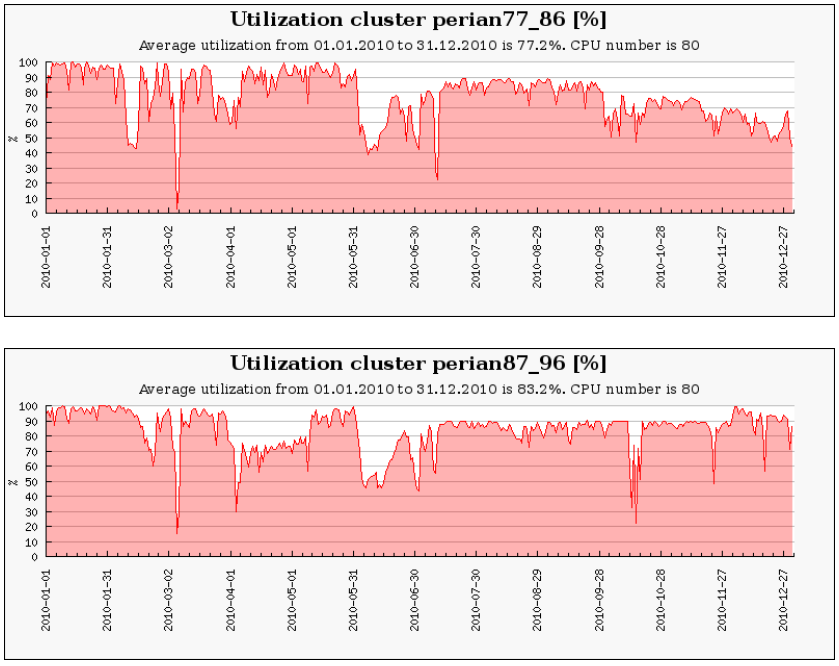


Fig. 18. Perian cluster utilisation (NCBR/MU).

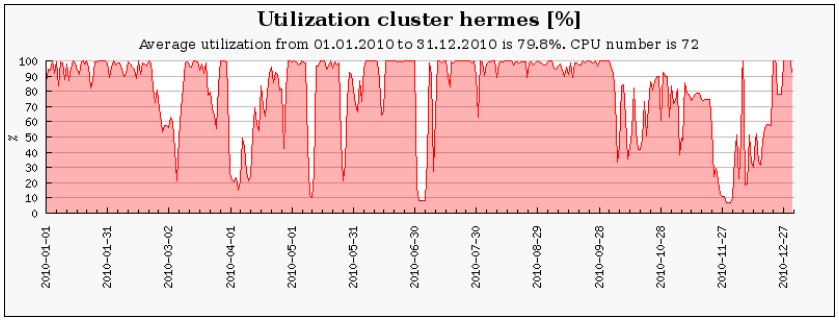


Fig. 19. Hermes cluster utilisation (JCU).

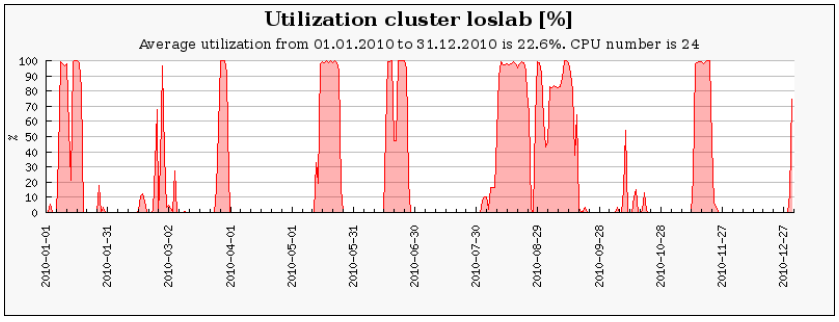


Fig. 20. Loslab cluster utilisation (Loschmidt Laboratories/MU).

6.4 Top Users

Graphs in Figs. 21 and 22 illustrate quantitative differences of number of submitted jobs and consumed CPU time among users. The top user runs more than 250 thousand jobs and consumed more than 1.4 million CPU-hours (about 160 CPU-years).

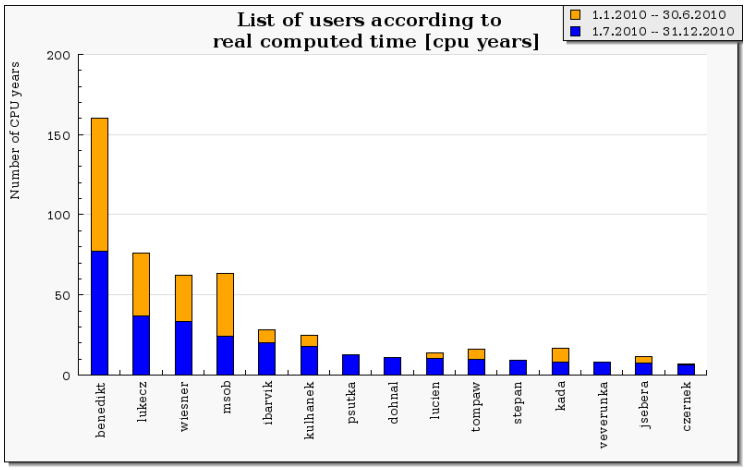


Fig. 21. CPU time by top users

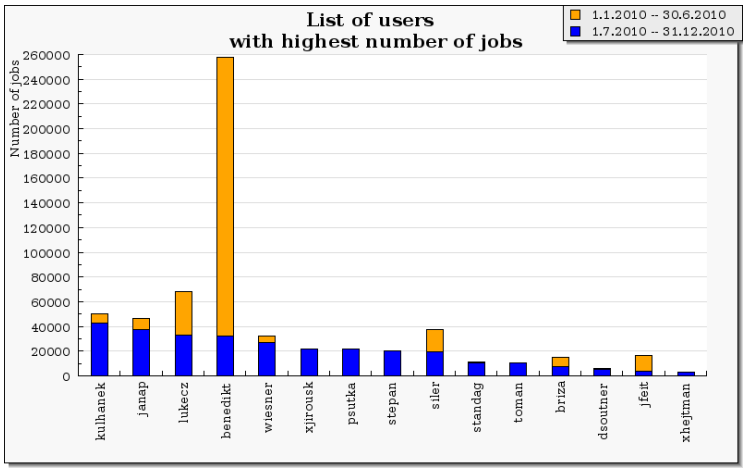


Fig. 22. No. of jobs by top users

User with longest computed time took more than one fifth of whole computed CPU time in MetaCentrum. The first four users computed in 2010 together nearly 60% of whole computed CPU time.

More detailed preview of computer time using by institutions is shown at chart in Fig. 23. The most activity is perceived at the Masaryk University (44% of whole CPU time), University of West Bohemia (25%) and Academy of Sciences (18%).

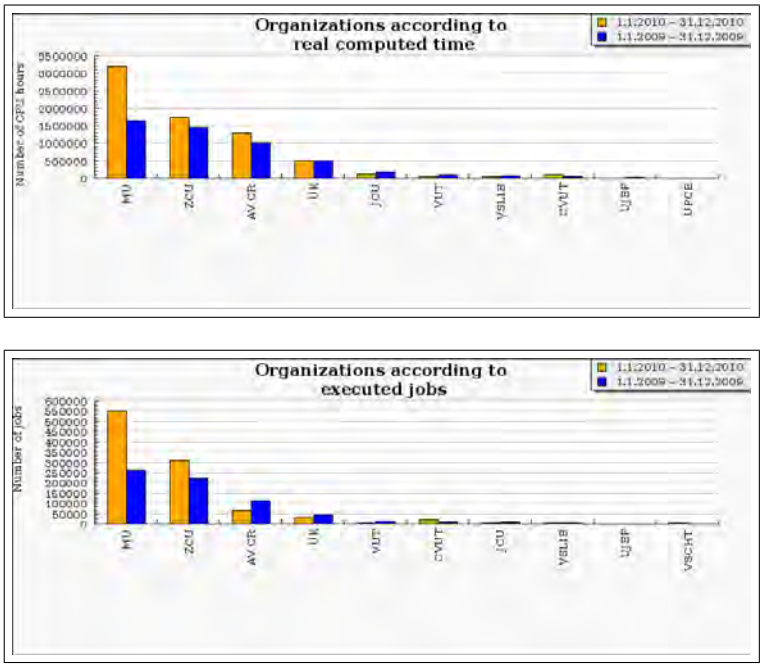


Fig. 23. Institutions according to computed CPU time and the number of jobs.

The applications consuming most of CPU time are shown in Fig. 8. The "Others" column represents sum of commercial applications with smaller usage, undistinguished home-grown user scripts and applications.

6.5 Support of International Virtual Organizations (VO)

Users in international VOs used more than 10 millions CPU hours in 2.4 millions jobs in praguelcg2 on FZU AV CR and prague_cesnet_lcg2 on CESNET.

MetaCentrum provides resources for EGI through two centres—praguelcg2 at FZU AV CR and prague_cesnet_lcg2 at CESNET (2720 + 80 CPU cores). These resources are used by Auger and VOCE virtual organizations (under our maintenance) and Belle, SuperNEMO, EUAsia, Atlas, Alice and others (with our support).

Auger virtual organization (in maintenance of MetaCentrum, 57 users) of Pierre Auger Observatory project states in EGI official statistics on 11th place according to submitted jobs and on the 14th place according to computed CPU

time (of 266 virtual organizations). Also is the largest user of CESNET's gLite resources with 82% share on normalized CPU computing time. Simulated data are transferred to Lyon computing centre where they are available via SRB also for users without grid certificate.

Virtual Organisation of Central Europe (VOCE, under MetaCentrum administration, with 196 active users) is so called catch-all virtual organisation providing its resources (gLite middleware) for users who does not belong into special VO or who want to try grids in middle European region.

EUAsia is so called catch-all virtual organisation representing users in Asia and Pacific region. This VO was established as a part of EUAsiaGrid project.

Japanese experiment Belle (KEK B-factory) run in 2010 two simulation waves on the grid platform. CESNET resources were part of both. The most of computing resources provides for Belle the KEK centre in Japan, which is outside the grid. The next generation of Belle project (Belle II) plans to move the majority of simulation activities on distributed grid sources. Preparation and tests are being solved in simulations inside current Belle project. This project support required installation of its own database on dedicated virtual server. It consumed 14% of normalized computing time on CESNET gLite working nodes.

SuperNEMO project from particle physicist of Manchester uses CESNET's resources mainly in batch actions with needs of mass simulations. They run nearly 1500 jobs on our clusters. The whole amount of jobs was 89 thousands in EGEE/EGI part.

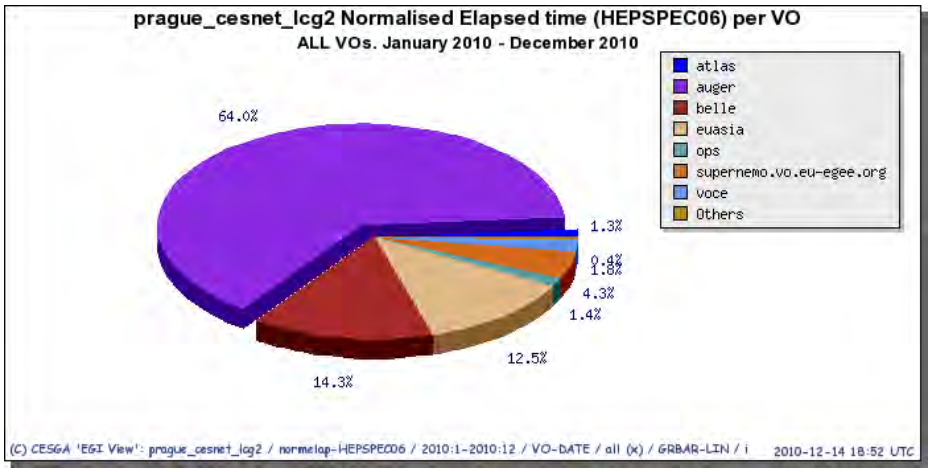


Fig. 24. prague_cesnet_lcg2 – normalized computed CPU time (elapsed HEP-SPEC06) by different VOs

EGI recognizes on the whole 266 of virtual organizations. According to official accounting data (computed time and number of jobs) from the whole EGI are

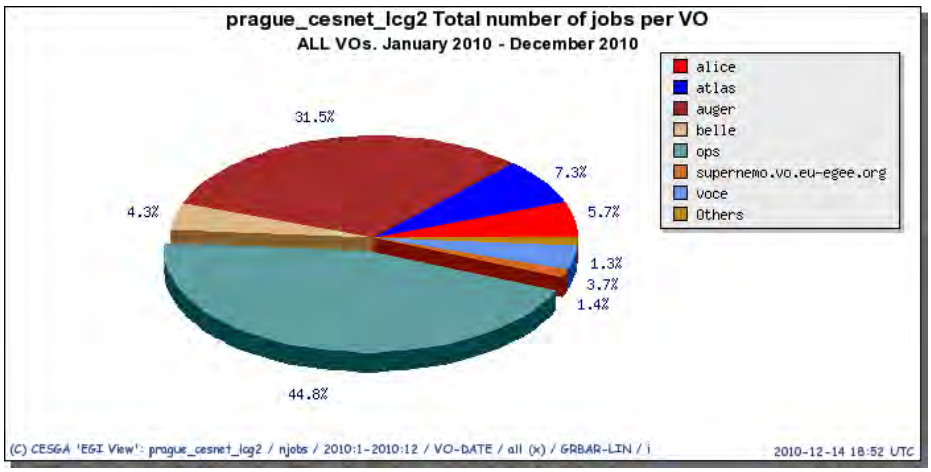


Fig. 25. prague_cesnet_lcg2 – amount of jobs from different VOs

our maintained and supported virtual organisations on the top of the charts with computed CPU time (table 7) and with number of computed jobs (table 8).

rank	VO	CPU hours
1	atlas	479 mil.
2	cms	146 mil.
3	alice	120 mil.
4	lhcb	62 mil.
13	auger	4,8 mil.
23	belle	2,1 mil.
44	voce	0,6 mil.
46	euasia	0,6 mil.
73	supernemo	0,2 mil.

Table 7. Rank of VOs in comparison with whole EGI according to computed CPU time.

rank	VO	# jobs
1	atlas	175.478 thous.
2	cms	37.879 thous.
3	alice	17.703 thous.
4	ops	14.320 thous.
11	auger	712 thous.
32	euasia	166 thous.
42	voce	105 thous.
50	supernemo	89 thous.
55	belle	78 thous.

Table 8. Rank of VOs in comparison with whole EGI according to number of jobs.

7 Special Services

Czech NGI maintains services like "back office" for large international projects ELI and EGI InSPIRE.

EGI project:

- webs,
- wiki,
- document server,
- Indico,
- mailing lists.

ELI project:

- wiki,
- document server,
- single sign on,
- RT support.

More detailed information for mentioned services are available at MetaCentrum's wiki[6].

8 MetaCentrum Personnel

Most of the people contributing to the Metacentrum success in 2010 work only part time as CESNET employees or are members of the international projects teams. We list all the people who in some way worked for CESNET (usually in a part time job) in 2010 and contributed to either the MetaCentrum activities or were involved in international projects.

[6] `https://meta.cesnet.cz/metawiki/index.php5/EGI` and `https://meta.cesnet.cz/metawiki/index.php5/ELI`

David Antoš
Jiří Chudoba
František Dvořák
Jiří Filipovič
Lukáš Hejtmánek
Petr Holub
Jana Hrdličková
Martin Klubal
Jan Kmuníček
Tomáš Kouba
Daniel Kouřil
Gabriela Krčmářová
Iva Krejčí
Aleš Křenek
Ivana Křenková
Igor Krnáč
Martin Kuba
Jan Kundrát
Slávek Licehammer
Miloš Liška
Miloš Lokajíček
Luděk Matyska
Miloš Mulač
Jan Pospíšil
Michal Procházka
Miroslav Ruda
Zdeněk Salvet
Zora Sebestianová
Jiří Sitera
Zdeněk Šustr
Jan Švec
Šimon Tóth
Michal Voců
Vlasta Žáková

9 End-User Papers

All the infrastructure activities do not have much sense without users that unleash the potential of the shared computing and storage infrastructure. The statistics provided earlier confirm that the CPUs are not staying idle, the actual scientific value of the computing power consumption is demonstrated on the following pages that are the bulk of this Yearbook.

This report includes contributions written by individual scientists as well as large scientific teams covering practically all the scientific areas that use the distributed computing infrastructure. The papers were selected to include all the major infrastructure users together with contributions that did not consume enormous amount of computing power but were of interest due to the methods they deployed, tools they used or results they achieved.

Each paper references one or more publications, usually from high quality journals or international conferences, as a confirmation of the scientific achievements done by the MetaCentrum user community with the direct use of provided resources.

References

1. Jiří Sitera, Miroslav Ruda, Petr Holub, David Antoš, Luděk Matyska MetaCentrum Virtualization – Use Cases CESNET technical report 30/2010, CESNET, z. s. p. o., 2010.
2. Šimon Tóth, Miroslav Ruda, Luděk Matyska Towards Peer-to-Peer Scheduling Architecture for the Czech National Grid CESNET technical report 27/2010, CESNET, z. s. p. o., 2010.

3. Petr Holub, Miloš Liška 4K Video and Audio Packet Format for UltraGrid CESNET technical report 24/2010, CESNET, z. s. p. o., 2010.
4. J. Kmuníček, L. Hejtmánek, J. Brezovský, V. Kaplan, T. Hnízdil, L. Matyska Adoption of Automatic Distributed Analysis Environment in MetaCentrum (Czech National Grid Initiative) CESNET technical report 13/2010, CESNET, z. s. p. o., 2010.
5. Colling David, Martyniak Janusz, McGough Andrew, Křenek Aleš, Sitera Jiří, Mulač Miloš, Dvořák František Real Time Monitor of Grid job executions In *Journal of Physics*. Conference Series, 6/219 ISSN: 1742-6588.
6. Šustr Zdeněk, Sitera Jiří, Dvořák František, Filipovič Jiří, Kouřil Daniel, Křenek Aleš, Matyska Luděk, Mulač Miloš, Pospíšil Jan, Ruda Miroslav, Salvet Zdeněk, Voců Michal Something you may have wanted to know about L&B In *Journal of Physics*. Conference Series, 6/219 ISSN 1742-6596.
7. Michal Procházka, Daniel Kouřil, and Luděk Matyska. User Centric Authentication for Web Applications. In *Proceedings of the 2010 International Symposium on Collaborative Technologies and Systems*. IEEE, 2010. ISBN 978-1-4244-6620-7.

Part II

MetaCentrum User Reports

Thermodynamic, Magnetic and Mechanical Properties of Advanced Materials

Mojmír Šob[1,2]

[1] Department of Chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic
sob@chemi.muni.cz
[2] Institute of Physics of Materials, Academy of Sciences of the Czech Republic, v.v.i., Žižkova 22, 616 62 Brno, Czech Republic
mojmir@ipm.cz

1 Introduction

In 2010, our scientific activities utilizing computational facilities of MetaCenter continued in theoretical studies of mechanical, magnetic and thermodynamic properties of advanced materials. The particular subjects are listed below, together with the most important results achieved.

2 Principal Results Obtained in the Year 2010

2.1 Thermodynamics and Phase Diagrams in Advanced Metallic Materials

We have performed first-principles electronic structure calculations of total energy differences between the sigma phase and Reference States (RS) of pure constituents in Cr-Fe and Cr-Co systems; the calculated results were compared with enthalpies of formation measured by calorimetry. Both measurements and calculations provide positive values of enthalpy of formation with respect to the RS. Negative values can be obtained when the pure constituents in the sigma phase structure are taken as the RS. Total energy differences of all sigma phase configurations involved are calculated at equilibrium volumes, reproducing well the experimental energy of formation of the sigma phase. The magnetic configurations in Cr-Fe and Cr-Co are also investigated and the stabilizing effect of magnetic ordering in sigma phase at 0 K is demonstrated. It turns out that the magnetic moment depends on the type of the occupied sublattice and total composition of alloy [1].

Further, we have studied the energetics of Cr-Hf and Cr-Ti systems which exhibit the existence of all polytypes of Laves phases, i.e., lower-temperature cubic C15 and higher-temperature hexagonal C14 and C36. Comparison of total energies of these structures calculated from first principles with the total energy of the ideal mixture of elemental constituents revealed the relative stability of Laves phases in these systems. The effect of magnetic order in the Laves

phases was also analyzed. The calculated total energies of formation of all the three polytypes were employed in two- and three-sublattice models to revise the thermodynamic description of both the systems published recently. New remodeled Gibbs energies of Laves phases require less fitting parameters than those obtained in previous treatments and corresponding phase diagrams provide an excellent agreement with the experimental phase data found in the literature. The proposed procedure allows us to compare the optimised heat capacity differences with those determined experimentally or theoretically and to use them in phase diagram calculations [2].

We have also started theoretical studies of energetics of C14, C15 and C36 Laves-phases formation in the system Ta-V. We have also performed the structural analysis of the above mentioned phases. The results obtained were used for assessment of thermodynamic properties and phase diagram calculations.

The above results were presented also in the invited talk [9], in oral contributions [26, 27] and a poster presentation [20] at important international conferences and at two academic institutions abroad [31, 34].

2.2 Phase Stability and Physical Properties of Polonium

Employing full-potential linearized augmented plane-wave method, we investigated the stability of Po in its ground-state simple cubic structure alpha-Po with respect to the trigonal spiral structure exhibited by Se and Te and to the displacive phase transformations into either tetragonal or trigonal phases. The origin of the phase stability of alpha-Po was analyzed with the help of densities of states, electronic band structures, and total energies of competing higher-energy structures corresponding to selected stationary points of the total energy. The electronic structures and total energies were calculated both within the generalized gradient approximation and local-density approximation (LDA) to the exchange-correlation energy as well as with and without inclusion of the spin-orbit (SO) coupling. The total energies were displayed in contour plots as functions of selected structural parameters and atomic volume. It turned out that the LDA calculation with SO interaction incorporated provided best agreement with existing experimental data and that the simple cubic structure of alpha-Po was stabilized by relativistic effects of core electrons. High elastic anisotropy of alpha-Po was explained as a consequence of its simple cubic structure and was compared with elastic properties of other crystal structures. Finally, an uniaxial tensile test for loading along the [001] and [111] directions was simulated; the corresponding theoretical tensile strengths calculated within the LDA+SO approach amount to 4.2 GPa and 4.7 GPa, respectively, which are the lowest values predicted in an element so far. According to Pugh and Frantsevich criteria, alpha-Po was predicted to be ductile. Also a positive value of the Cauchy pressure confirmed the metallic type of interatomic bonding [3].

2.3 Magnetism, Electronic Structure and Atomic Configuration of Grain Boundaries

Analyzing experimental data and calculating corresponding energy barriers from first principles, we elucidated the feasibility of the experimentally observed phase transformation between the hcp and double hcp (dhcp) structures in the (11-20) oriented hcp Pd thin films grown on W(001) and Nb(001) substrates and absence of the hcp-fcc transformation in those films. The hcp-dhcp transformation can be modeled by a transformation path which preserves the existing domain topology of the films and exhibits a sufficiently low energy barrier. On the other hand, this orthogonal pattern of rectangular domains induced by the fourfold symmetry of the substrate surface hinders the hcp Pd phase to convert back to the ground-state fcc phase, although there exists a transformation path exhibiting a very low energy barrier between the hcp and fcc structures. This path, however, would break the domain arrangement and, therefore, it cannot be accomplished. In this way, the hcp crystalline phase is locked inside of nanograins. Our study constitutes an example how a higher-energy configuration can be stabilized by the topology of defects (here grain architecture/grain boundaries) and suggests a route for technological applications to prevent the destabilization of certain desirable properties (e.g., possible ferromagnetism) induced in the higher-energy (hcp Pd) phase [5].

Further, we have studied in detail segregation and embrittling energy of sp elements of the 3rd, 4th and 5th period (Al, Si, P, S, Ga, Ge, As, Se, In, Sn, Sb and Te) at the $\Sigma 5(210)$ grain boundary (GB) in fcc nickel and the segregation of these impurities at the (210) free surface (FS). The effect of impurities on the distribution of magnetic moments has been analysed. We determined the embrittling energy from the difference between the GB and FS binding energies on the basis of the Rice-Wang model and separated embrittling energy into the chemical and mechanical part [11, 12].

The results were further presented in one invited talk at a prestigious international conference [6], presented in the conference presentations [14, 15, 22-24] and at the Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany [28].

2.4 Ab Initio Study of Structure and Extended Defects in $MoSi_2$

We have shown that due to a lower symmetry of the tetragonal $C11_b$ structure when compared with the cubic BCC lattice, the 1/2<331] dislocation cores are split asymmetrically contrary to the 1/2<111> BCC dislocations. This has essential impact on their behaviour and, consequently, on mechanical properties. Various types of dislocation dissociations were analyzed in the frame of anisotropic elasticity with the help of the data from ab initio calculations of ?-surfaces for generalized stacking faults [4]. The results on $MoSi_2$ were also presented in the conference presentations [13, 18, 19, 21].

2.5 Summary of Presentation of Results: Publications, Invited Talks, Conference Presentations

In total, in 2010 our results have been published in 5 papers in top academic journals [1-5], in two Proceedings of national conferences [11, 12] and have been presented in five invited talks at important international conferences [6-10], in 25 other contributions at conferences, scientific meetings and schools [13-27], in 6 talks at scientific institutions abroad [28, 30-34] and in one talk at a Czech scientific institution [29].

2.6 Organization of a National Meeting

Prof. M. Šob served as a co-chairman of the Doctoral Conference on Multiscale Design of Advanced Materials organized on Dec. 2, 2010 at the Institute of Physics of Materials of the Academy of Sciences of the Czech Republic, Brno. The conference was attended by about 25 doctoral students associated in the Doctoral Project of the Grant Agency of the Czech Republic (Project. No. 106/09/H035) who presented their latest results.

3 Software and Financial Support

The software for ab initio calculations of electronic structure (LMTO-ASA, WIEN2k–FLAPW, VASP–pseudopotential code) used for solution of the above-mentioned tasks was provided by the cooperating Institute of Physics of Materials, Academy of Sciences of the Czech Republic, Brno, and was implemented at the MetaCenter of the Masaryk University under the Research Project No. MSM6383917201.

Further, our research was supported by the Grant Agency of the Czech Republic (Projects No. 202/09/1786, P108/10/1908 and 106/09/H035), by Ministry of Education of the Czech Republic (Projects COST OC09011 and OC10008), Grant Agency of the Academy of Sciences of the Czech Republic (Project No. IAA100100920) and by Research Projects AV0Z20410507 and MSM0021622410. The access to the computing facilities of the METACenter of the Masaryk University, Brno, provided under the Research Project No. MSM6383917201, is gratefully acknowledged.

4 Publication and Presentations of Results

4.1 Publications in Refereed Journals and in Scientific Monographs

[1] J. Pavlů, J. Vřešťál, M. Šob: Ab initio study of formation energy and magnetism of sigma phase in Cr-Fe and Cr-Co systems. Intermetallics 18 (2010), 212-220.

[2] J. Pavlů, J. Vřešťál, M. Šob: Thermodynamic modeling of Laves phases in the Cr-Hf and Cr-Ti systems: Reassessment using first-principles results. CALPHAD: Computer Coupling of Phase Diagrams and Thermochemistry 34 (2010),

215-221.

[3] D. Legut, M. Friák, M. Šob: Phase stability, elasticity, and theoretical strength of polonium from first principles. Phys. Rev. B 81 (2010), 214118 (19 pp).

[4] V. Paidar, M. Čák, M. Šob, V. Vitek: Theoretical analysis of dislocation splittings in $MoSi_2$. Journal of Physics: Conference Series 240 (2010), 012007 (6 pp).

[5] E Hüger, T. Káňa, M. Šob: Hexagonal close-packed Pd locked inside the nanograins. CALPHAD: Computer Coupling of Phase Diagrams and Thermochemistry 34 (2010), 421-427.

4.2 Invited Talks at International Conferences

[6] M. Všianská, M. Šob: The effect of segregated sp-impurities on grain-boundary embrittlement in nickel. 139th Annual Meeting and Exhibition of The Minerals, Metals and Materials Society (TMS), Seattle, WA, Feb.14-18, 2010 (invited talk).

[7] M. Šob, M. Friák, M. Zelený: Ab initio studies of magnetism at high strains and theoretical strength of materials. Mini-Symposium on Computational Materials Science and Engineering, Department. of Materials Science and Engineering, Pohang University of Science and Technology, Pohang, Korea, May 21, 2010 (invited talk).

[8] M. Šob, M. Zelený, M. Friák: Phase Stability in Fe, Co and Ni under High Strain and Surface Phase Transitions in Overlayers. Materials Week, 2010 International Symposium on Multi-scale Modeling and Simulation of Materials, Institute of Materials Research, Chinese Academy of Science, Shenyang, July 5-8, 2010 (invited talk).

[9] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Application of ab initio Results in Modelling Phase Diagrams Containing Complex Phases. Materials Week, 2010 International Symposium on Multi-scale Modeling and Simulation of Materials, Institute of Materials Research, Chinese Academy of Science, Shenyang, July 5-8, 2010 (invited talk).

[10] M. Šob, M. Zelený, M. Friák: The effect of magnetism on strength and structural stability in ferromagnetic metals. 7th Pacific Rim International Conference on Advanced Materials and Processing (PRICM7), Cairns, Australia, Aug. 2-6, 2010 (keynote talk).

4.3 Contributions in Proceedings of international and national conferences - published

[11] M. Všianská, M. Šob: Ab initio study of effect of segregated sp-impurities at grain boundaries nickel. In: 10th Workshop of Physical Chemists and Electrochemists, ed. L. Trnková, Mendel University, Brno 2010, pp. 262-266.

[12] M. Všianská, M. Šob: Influence of segregation of non-magnetic impurities on structure and properties of grain boundaries in nickel. In: Multiscale Design of

Advanced Materials 2009 (Proceedings of Doctoral Conference), eds. I. Dlouhý, J. Švejcar, M. Šob, Institute of Physics of Materials, Academy of Sciences of the Czech Republic, Brno 2010, pp. 7-12.

4.4 Unpublished Presentations at International and National Conferences and Schools

[13] M. Friák, D. Legut, M. Šob: Ab initio study of extreme loading conditions in transition-metal disilicides with C40 structure. 139th Annual Meeting and Exhibition of The Minerals, Metals and Materials Society (TMS), Seattle, WA, Feb.14-18, 2010.

[14] M. Všianská, M. Šob: Structure and magnetism of clean and decorated grain boundaries in nickel. 2010 Frühjahrstagung der Deutschen Physikalischen Gesellschaft, Regensburg, March 21-26, 2010.

[15] E. Hüger, T. Káňa, M. Šob: Ab initio study of blocking of hcp-fcc transformation in Pd films by domain boundaries. XXXIXth Int. Conf. On Phase Diagrams and Computational Thermodynamics (CALPHAD), Jeju Island, Korea, May 23-28, 2010.

[16] M. Zelený, M. Friák, M. Šob: Ab initio studies of surface phase transitions in Fe, Co and Ni overlayers. Int. Conf. on Solid-Solid Phase Transformations (PTM 2010), Avignon, France, June 6-11, 2010.

[17] M. Šob, M. Zelený, M. Friák: The effect of magnetism on ideal strength and structural stability in ferromagnetic metals. 6th Int. Conf. On Materials Structure and Micromechanics of Fracture, Brno, June 28-30, 2010.

[18] T. Káňa, M. Šob, V. Vitek: Transformation paths in transition-metal disilicides. 6th Int. Conf. On Materials Structure and Micromechanics of Fracture, Brno, June 28-30, 2010.

[19] T. Káňa, M. Šob, V. Vitek: Ab initio study of phase transformations in transition-metal disilicides. 17th WIEN2k workshop & Int. Conf. On Advanced Materials Modelling, Institut des Matériaux Jean Rouxel, Université de Nantes, France, July 5-10, 2010.

[20] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Application of ab initio Results in Modelling Phase Diagrams Containing Complex Phases. Psi-k 2010 Conference - Ab initio (from electronic structure) calculation of complex processes in materials, Berlin, Sept. 12-16, 2010 (poster).

[21] T. Káňa, M. Šob, V. Vitek: Ab initio study of transformation paths between C11b, C40 and C54 structures in transition-metal disilicides. Psi-k 2010 Conference - Ab initio (from electronic structure) calculation of complex processes in materials, Berlin, Sept. 12-16, 2010 (poster).

[22] M. Všianská, M. Šob: Ab initio study of impurity segregation at grain boundaries in nickel. Psi-k 2010 Conference - Ab initio (from electronic structure) calculation of complex processes in materials, Berlin, Sept. 12-16, 2010 (poster).

[23] J. Kuriplach, O. Melikhova, M. Šob, P. Lejček, V. Paidar: Structure, stability and segregation at $\Sigma 19$ (331) tilt grain boundary in Ni. Psi-k 2010 Conference - Ab initio (from electronic structure) calculation of complex processes in materials, Berlin, Sept. 12-16, 2010 (poster).

[24] E. Hüger, T. Káňa, M. Šob: A mechanism of inhibition of phase transitions in nano-grained close-packed Pd thin films. 3rd workshop on ab-initio phonon calculations, Cracow, December 1-4, 2010 (poster).

[25] M. Šob, M. Zelený, M. Friák: The effect of magnetism on ideal strength and structural stability in metals and intermetallics. Int. Conf. On Applied Mechanics, Materials and Manufacturing, Muscat, Sultanáte of Oman, Dec. 13-15, 2010.

[26] J. Pavlů, J. Vřešťál, X.-Q. Chen, P. Rogl: Stability of Laves phases in the Ta-V system. XXXIXth Int. Conf. On Phase Diagrams and Computational Thermodynamics (CALPHAD), Jeju Island, Korea, May 23-28, 2010.

[27] J. Pavlů, J. Vřešťál, X.-Q. Chen, P. Rogl: First-principles study and thermodynamic modelling of Laves phases in the Ta-V system. Discussion Meeting on Thermodynamics of alloys (TOFA 2010), Porto, Portugal, Sept. 12-16, 2010.

4.5 Presentations at Scientific Institutions

[28] M. Všianská, M. Šob: Ab initio study of segregation of impurities at grain boundaries and surfaces in nickel. Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany, Jan. 21, 2010.

[29] M. Šob: Application of ab initio electronic structure calculations for multiscale materials modelling. Institute of Materials Chemistry, Faculty of Chemistry, Brno University of Thechnology, March 15, 2010.

[30] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Aplication of ab initio Results in Modelling Phase Diagrams Containing Complex Phases. Northeastern University, Key Laboratory for Anisotropy and Texture of Materials (MOE), Shenyang, China. July 9, 2010.

[31] M. Šob, M. Zelený, M. Friák: The effect of magnetism on some mechanical properties of metals and intermetallics. Northeastern University, Key Laboratory for Anisotropy and Texture of Materials (MOE), Shenyang, China. July 12, 2010.

[32] M. Šob, M. Friák, D. Legut, J. Kuriplach, I. Turek, V. Vitek: The application of ab initio electronic structure calculations in multiscale modeling of materials. Northeastern University, Key Laboratory for Anisotropy and Texture of Materials (MOE), Shenyang, China. July 12, 2010.

[33] M. Šob, M. Zelený, M. Friák: The Effect of Magnetism on Strength and Structural Stability in Ferromagnetic Metals. Metallurgy and Materials Research Institute, Chulalongkorn University, Bangkok, Thailand, Aug. 18, 2010.

[34] M. Šob, J. Pavlů, J. Vřešťál, A. Kroupa: Aplication of ab initio Results in Modelling Phase Diagrams Containing Complex Phases. Seminar of the Division of Solid-state Physics, AGH University of Science & Technology, Cracow, Poland, Oct. 20, 2010.

4.6 Proceedings of International and National Conferences

[35] I. Dlouhý, J. Švejcar, M. Šob (eds.): Multiscale Design of Advanced Materials (Proceedings from Doctoral Conference, Brno, Dec. 2, 2010), Institute of Physics of Materials of the Academy of Sciences of the Czech Republic, Brno, 2010, 131 pp. ISBN 978-80-87434-02-4.

Radiative Transfer Study—DART Simulations

Věroslav Kaplan, Petr Lukeš, and Jan Hanuš

Global Change Research Center, AS CR
Bělidla 4a, Brno, 60200
kaplan.v@czechglobe.cz, lukes.p@czechglobe.cz, hanus.j@czechglobe.cz

Abstract. Chlorophyll is a green foliar pigment, which plays an important role in photosynthesis. It absorbs and transports the energy of incoming sunlight. Five different chlorophyll types have been identified, the highest chlorophyll concentration in tissues is of C_a and C_b types. Concentration of C_{a+b} varies in different plant species during the year. Chlorophyll concentration also responds to leaf physiological stress and overall concentration of C_{a+b} can be used as a proxy of vegetation health status.
Indirect estimation of chlorophyll content by the remote sensing techniques became possible with the introduction of hyperspectral remote sensors. Imagery acquired by hyperspectral instruments offers possibility of narrow band evaluation, which made possible more precise evaluation of chlorophyll absorption peaks (430 and 662 nm for C_a; 465 and 642 nm for C_b).

Remote Sensing workgroup on CzechGlobe (Global Change Research Center AS CR) uses MetaCentrum computing resources mainly to prepare supplemental data for estimation of chlorophyll content of Norway spruce (Picea abies Karst.) tissues. The retrieval of chlorophyll content is done from hyperspectral image date using radiative transfer inversion techniques with DART radiative transfer model.

Discrete Anisotropic Radiative Transfer model (DART, `http://www.cesbio.ups-tlse.fr/us/dart.htm`) was developed in CESBIO laboratory (Center for the Study of the BIOsphère) for simulation of radiative transfer in the "Earth-Atmosphere" system. The input of the model is a 3D model of landscape including optical, geometric and biophysical parameters of model elements—i.e., optical properties of the ground, positions and dimensions of trees on the landscape, distribution of the tree biomass, tree leaves types and many other properties.

The DART model can be applied in these two modes:

- in forward mode, the DART model simulated transfer of energy in the simulated scene. Results of the model are images which would be (theoretically) acquired by airborne/satellite imaging sensors flown over the simulated scene.
- In indirect inverse mode, the images generated by DART forward runs are compared to images acquired by imaging sensor. The matching of acquired images to simulated images with known parameters allows estimation of parameters from set-up of the simulated scene (e.g., chlorophyll content).

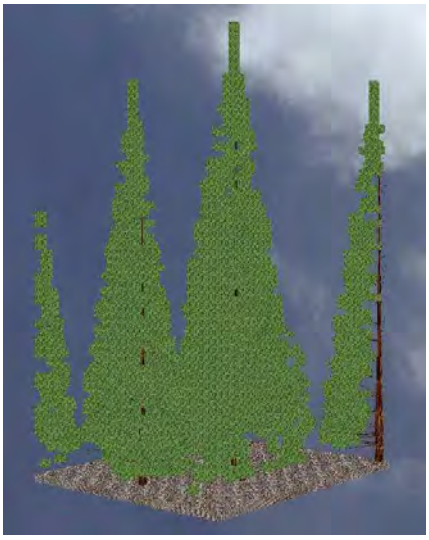


Fig. 1. 3D visualisation of DART mock-ups.

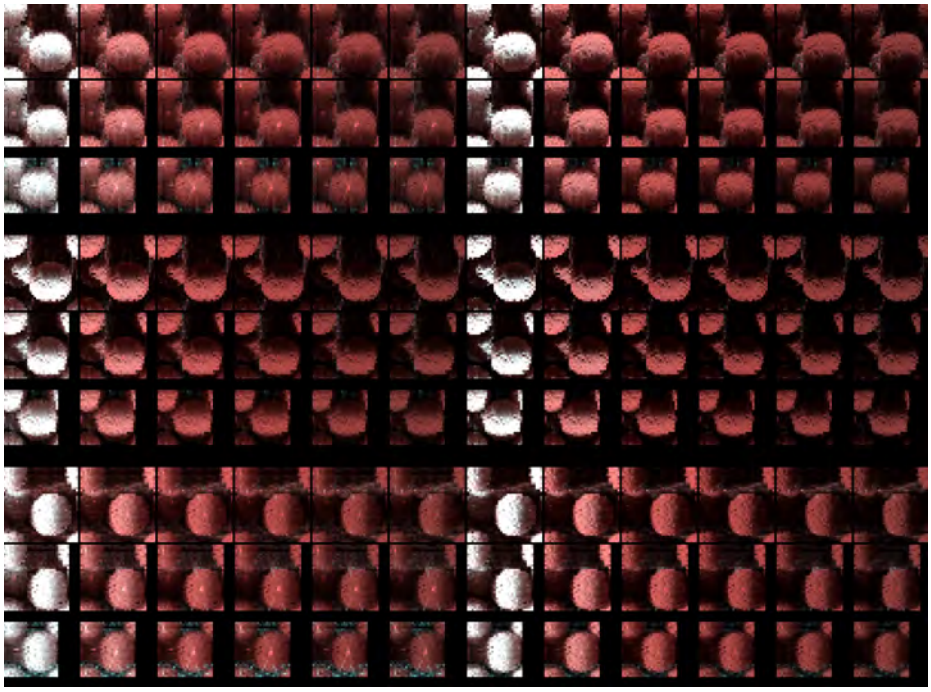


Fig. 2. Aggregated simulation results, arranged for further processing. Each simulation results in one tile.

In order to perform a radiative transfer model based retrieval, configuration of trees and parameters which are expected to appear in a studied area have to be prepared. The more precise parametrization of input scenes means the more precise retrieval results, so appropriate care should be taken to achieve better results.

Preparation of simulation parameter consists of field-work measuring of tree parameters (as mean trunk dimensions, mean live and dead crown height, ...), optical properties (needle reflectance and transmittance) and some other supplemental measurements (vegetation under-story optical properties, ...).

All measured properties are summarized, evaluated and appropriate simulation mock-ups are built. Unknown properties (chlorophyll content, leaf area index, water content, ...) are used in several expected values.

Depends on amount of known and unknown properties, number of simulations varies from several hundreds to several thousands. Prepared simulations are submitted to (super-)computing facilities.

Each of the simulations runs are independent on each other, making this type of computation load suitable for computing grid environment.

Results of all simulations are stored to the database (sometimes called lookup tables) with respect of their input properties. The prepared lookup tables are then used to train neural network or some other form of optimization tool. Trained neural network can be used to perform the retrieval.

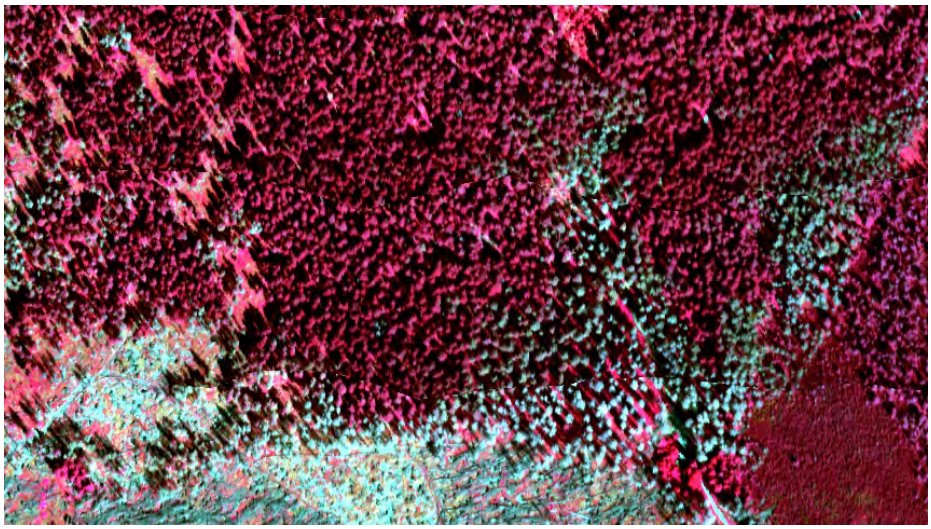


Fig. 3. Image acquired by AISA/Eagle, CIR colors. Černá Hora, Šumava National Park.

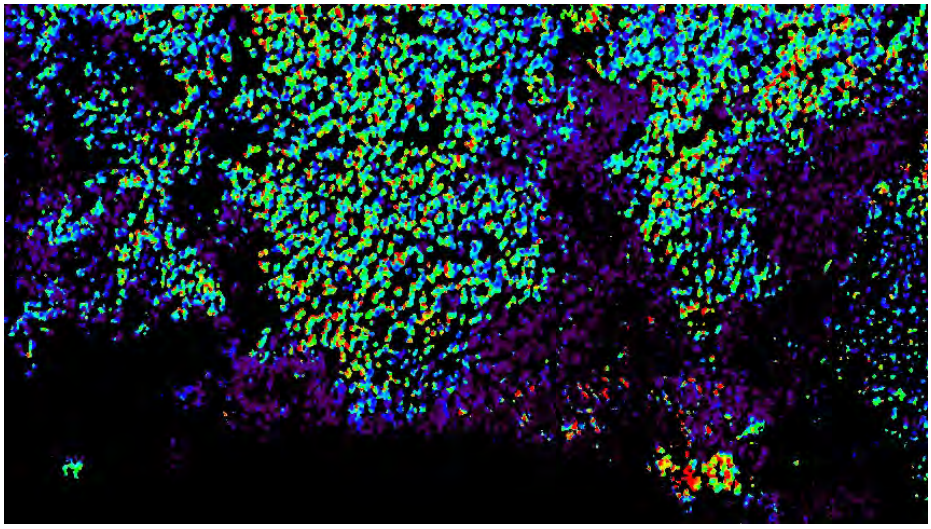


Fig. 4. Map of chlorophyll content.

Used Software

- Discrete Anisotropic Radiative Transfer model (DART, `http://www.cesbio.ups-tlse.fr/us/dart.htm`)
- custom Python scripts

Main results:

- Map of chlorophyll content—Černá Hora, Šumava National park
- verification of method and software

Molecular Dynamics Simulations of Biomolecules

Ivan Barvík

Institute of Physics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 5, 121 16 Prague 2, Czech Republic
ibarvik@karlov.mff.cuni.cz

MetaCentrum computing facilities have been used for the following projects in 2010:

Inhibitors of Hepatitis C virus RNA dependent RNA polymerase (HCV RdRp)

Hepatitis C virus (HCV) is spread among 3% of world population, of which only about 25% are having symptoms. Current treatment with interferon-α (immunomodulation) and ribavirin (nucleotide inhibitor) has limited efficiency and numerous side effects. Only 40% of people exposed to HCV outbreak fully recover while rest is affected by chronic liver problems, with possibility of developing cirrhosis or even liver cancer. The HCV RNA genome is reproduced in host cells by RNA dependent RNA polymerase (RdRp). RdRp has a typical structure of polymerases resembling the right hand (Palm, Thumb, and Fingers domains—see Figure 1). Conserved aspartic acids in the Palm domain bind Mg^{2+} ions. But unlike other polymerases, Thumb and Fingers are bound, and hence not allowed to change conformation. Because of this, RdRp is much more rigid than other polymerases. RdRp is not present in mammalian cells. Therefore, it is a suitable target for inhibition. One possibility of treatment is non-nucleoside inhibition (NNI) using allosteric inhibitors. They prevent HCV RdRp from reaching an active conformation. Several suitable binding sites for NNIs on the enzyme allow combination therapy. Further, nucleotide inhibitors of HCV RdRp have been explored in recent years by many groups.

Phosphoramidate dinucleosides named "GC 3-OH" series, carrying various phosphoramidate linkages have been previously reported as hepatitis C virus (HCV) inhibitors. To enhance the efficacy of these dinucleotides, novel "GC 3-H" series were synthesized as potential chain terminators [1]. Their inhibition potency was strongly increased by the introduction of novel neutral and bis-negatively charged phosphoramidate side chains. Their inhibitory effect on HCV NS5B polymerase was evaluated in vitro and in HCV subgenomic replicon containing Huh-6 cells. As expected, 3-H compounds were more potent than their 3-OH counterparts to inhibit HCV polymerase activity. The most potent inhibitor, a 5' phosphorylated dinucleotide bearing a bis-negatively charged amino side chain, exhibits an IC50 value of 8 μM in vitro and EC50 value of 2.6 μM in the HCV subgenomic replicon system.

A molecular structure model was presented in [1] to propose an interpretation of the gain afforded by the 3-H-cytidine modification. In summary, these

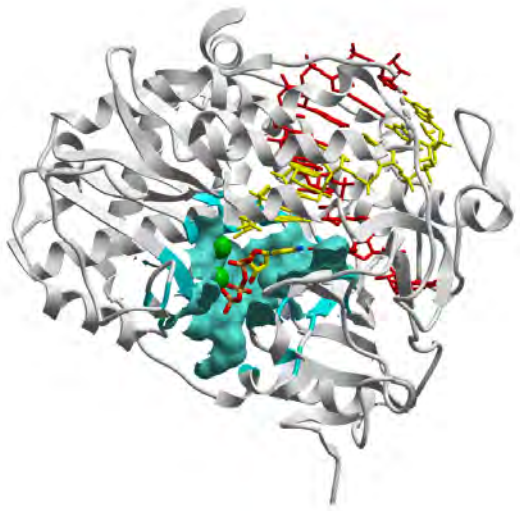


Fig. 1. RNA dependent RNA polymerase in complex with the RNA:RNA duplex and incoming NTP

data explain why GC 3-H bearing a neutral side chain in the Sp absolute configuration, which is not able to interact with basic Arg/Lys residues of the NTPs channel, exhibited remarkably stronger inhibition activity (comparing either to the Sp-GC 3-OH counterpart or both 3-H and 3-OH Rp isomers). Differences in the recognition of the nonbridging oxygen from the phosphoramidate linkages (Sp-GC 3-OH vs Sp-GC 3-H) by either Mg2 or Mg1 ions are rooted in the hydrogen bonds connecting the 3-OH group with the Asp225 residue. As a consequence, Sp-GC 3-OH is repositioned slightly comparing to Sp-GC 3-H leading to strengthening of contacts with Mg1, which offers less sound stabilization in the active site (comparing to Mg2). Further, in the case of GC 3-H series, Asp225 and Ser282 of HCV NS5B were connected by mutual hydrogen bonds and the 2-OH group of GC 3-H series was recognized firmly by Ser282. In contrast, in the case of GC 3-OH series, the 3-OH group was competing with Ser282 and often replaced it in Asp225 binding, destabilizing 2-OH interactions with Ser282. It could contribute further to a better stabilization of the 3-H GC series than the GC 3-OH series in the active site of HCV RdRp and therefore to a better inhibition activity.

Regulation of Bacillus subtilis RNA polymerase (RNAP) transcription initiation

RNA in all cellular organisms is synthesized by a complex molecular machine—the DNA-dependent RNA polymerase (RNAP). The RNAP core is evolutionarily conserved in sequence, structure and function from bacteria to man. The cat-

alytically competent core RNAP consists of $\alpha 2\beta\beta'\omega$ subunits. Promoter-specific initiation of transcription requires an additional subunit, σ, which binds the core RNAP to form the holoenzyme. In bacteria, rapid changes in gene expression can be achieved by affecting the activity of RNA polymerase with small molecule effectors during transcription initiation [2]. An important small molecule effector is the initiating nucleoside triphosphate (iNTP). At some promoters, an increasing iNTP concentration stimulates promoter activity, while a decreasing concentration has the opposite effect. In [2], the promoter region, corresponding to the transcription bubble was identified as key for B. subtilis rRNA promoter regulation via the concentration of the iNTP. Within this region, the conserved −5T (3 bp downstream from the 10 hexamer) was required for this regulation.

In silico modeling with a B. subtilis RNAP homology model suggested [2] that the base of the −5 promoter position of the non-template strand may make contacts with the 428-ERVVRE-433 motif of the β subunit of RNAP (Figure 2). Moreover, the B. subtilis promoter non-template −5 position has the potential to interact with σA region 1.2. However, according to the B. subtilis model, interactions involving the base are less likely to be formed in the open complex but can possibly occur during isomerization, when σA region 1.2 may be present in the immediate vicinity of the non-template strand −5 base. The base of the −5 promoter position of the template strand may contact the 272-EEDD-275 motif of σA.

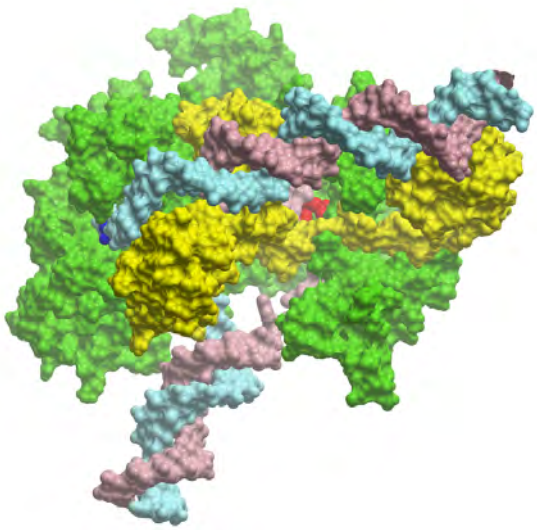


Fig. 2. Model of the promoter DNA:DNA transcription bubble and possible interactions with B. subtilis RNAP β and σA subunits. Green, b; yellow, σA; light pink, DNA template strand; light blue, DNA non-template strand.

Thermostability of Bacillus stearothermophilus Elongation Factor Tu (EFTu)

Elongation factor Tu (EF-Tu) is a eubacterial, monomeric, model GTP-binding protein, required for delivery of individual aminoacyl-tRNAs (aa-tRNAs) to the translating ribosome. Its molecule consists of three covalently linked domains 1-3 (numbering from the N-terminus, see Figure 3). Domain 1, also called the G domain, binds GDP or GTP and contains the catalytic GTPase center with a low intrinsic activity. Domains 2+3 modulate, to various degree and direction, depending on EF-Tu origin, these two activities and thermostability of domain 1. All three domains of EF-Tu in the GTP-bound state are absolutely required to form the binding pocket for aa-tRNA and to deliver it to the A-site of the mRNA-programmed ribosome. The cognate codon-anticodon recognition leads to a highly accelerated activation of the EF-Tu GTPase and a rapid hydrolysis of EF-Tu-bound GTP into GDP and Pi. The GTPase activation signal originating from the codon recognition event is transmitted through the mRNA.ribosome.aa--tRNA.EF-Tu.GTP complex to the GTPase center of EF-Tu by a mechanism involving chain conformational changes in individual components of the complex, including EF-Tu.

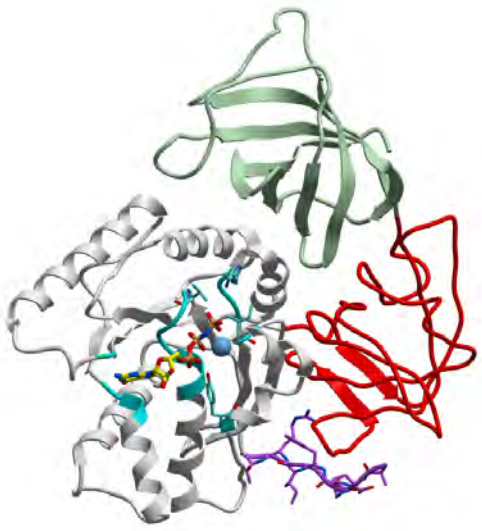


Fig. 3. The N-terminal region (magenta) is crucial for the thermostability of the G-domain (white) of Bacillus stearothermophilus EF-Tu.

The EF-Tu G-domain sets up the basal level of thermostability for the whole protein. In [3], we chose to locate the thermostability determinants distinguishing the thermophilic domain 1 from a mesophilic domain 1. By an ap-

proach of systematically swapping protein regions differing between G-domains from mesophilic Bacillus subtilis and thermophilic B. stearothermophilus, it was demonstrated that a small portion of the protein, the N-terminal 12 amino acid residues, plays a key role in the thermostability of this domain. MD simulations were applied to both G-domains [3]. In the absence of 3D structures for EFTus and G-domains from B. subtilis and B. stearothermophilus, homology models of both G-domains were prepared based on experimentally determined 3D structures of EF-Tus from E. coli and Thermus thermophilus. The MD simulations showed that the B. stearothermophilus C-terminal region may be more rigid and, therefore, more resistant to denaturation by heat than the corresponding region from B. subtilis. This is due to the presence of a proline residue localized in this region (Pro182 in B. stearothermophilus–Ala183 in B. subtilis). Further, MD simulations results suggested that N-terminal regions from B. subtilis and B. stearothermophilus G-domains can adopt conformations leading to stabilizing interactions with the functionally important effector region.

DNA templated self-assembly of boronic oligonucleotides

The Darwinian evolution of biomolecules is considered as a permanent process of mutation, selection and amplification. Several research groups have investigated many chemical systems able to achieve nonenzymatic oligonucleotide ligation leading to reversible backbone linkages which provided a means to repair or transform themselves in response to their environment. In [4], a DNA-templated ligation occurring through the reaction of a boronic acid at the 5'-end of one strand with a cis-diol function on the adjacent strand was examined. The resulting joined duplex differs from natural DNA by replacement of a phosphodiester with a boronate internucleosidic linkage (see Figure 4).

Software Packages

NAMD, AMBER, Gaussian09

Acknowledgments

This work was supported by Grants from the Ministry of Education, Youth and Sports of the Czech Republic (2B06065, MSM 0021620835). The access to the MetaCentrum computing facilities provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" LM2010005 funded by the Ministry of Education, Youth, and Sports of the Czech Republic is highly appreciated.

References

1. Priet S., Zlatev I., **Barvík Jr. I.**, Geerts K., Leyssen P., Neyts J., Dutartre H., Canard B., Vasseur J.J., Morvan F., Alvarez K.:

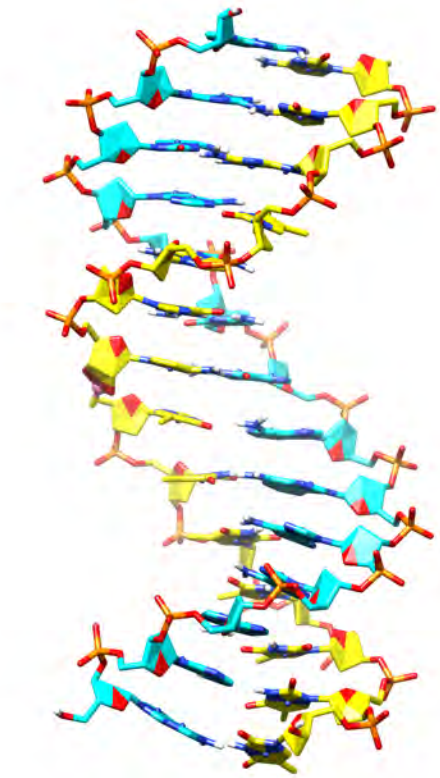


Fig. 4. The DNA:DNA duplex structure with a boronic internucleotide linkage inserted into the center of the yellow chain.

3'-Deoxy phosphoramidate dinucleosides as improved inhibitors of hepatitis C virus subgenomic replicon and NS5B polymerase activity.
Journal of Medicinal Chemistry, 2010, vol. 53, p. 6608-6617.

2. Sojka L., Kouba T., **Barvík Jr. I.**, Šanderová H., Maderová Z., Jonák J., Krásný L.:
Rapid changes in gene expression: DNA determinants of promoter regulation by the concentration of the transcription initiating NTP in Bacillus subtilis.
Nucleic Acids Research, 2011, in press

3. Šanderová H., Tišerová H., **Barvík Jr. I.**, Sojka L., Jonák J., Krásný L.:
The N-terminal region is crucial for the thermostability of the G-domain of Bacillus stearothermophilus EF-Tu.
Biochimica Biophysica Acta, 2010, vol. 1804, p. 147-155.

4. Martin A.R., **Barvík Jr. I.**, Luvino D., Smietana M., Vasseur J.J.:
DNA-Templated Self-Assembly by Dynamic and Programmable Boronic Ester Formation.
Angewandte Chemie, 2011, in press

5. **Barvík I.**, Zíma V.:
Molecular dynamics simulations of RNA polymerases.
NACON VIII 8th International Meeting on Recognition Studies in Nucleic Acids, The Edge, University of Sheffield, UK, 12-16 September 2010, Poster Abstracts, p. 91

6. **Barvík I.**, Zíma V.:
Molecular Dynamics Simulations of Bacterial and Viral RNA Polymerases.
Book of Abstracts: RNA in Motion, 19th Annual Growth Factor and Signal Transduction Symposium, Ames, Iowa, September 9-12, 2010, p. 29

7. Herrera SJE, **Barvík I.**, Štěpánek J.
Spectroscopic Study of Magnesium Effect on Stability of RNA and DNA Triple Strand Polynucleotide
22nd International Conference on Raman Spectroscopy, AUG 08-13, 2010 Boston
Book Series: AIP Conference Proceedings, Volume: 1267, Pages: 613-614

8. Zíma V., **Barvík I.**:
The homology modeling and molecular dynamics simulations of the ion channels from TRP family
8th Annual Ion Channel Retreat, Vancouver BC, Canada, 28.-30.6. 2010

The Molecular Dynamics Study of Complex Protein Systems

Milan Melicherčík, Žofie Sovová, Morteza Khabiri, Natalia Kulik, Vasilina Zayats, and Rüdiger Ettrich

Institute of Nanobiology and Structural Biology of GCRC
Academy of Sciences of the Czech Republic, and University of South Bohemia
Zámek 136, 37333 Nové Hrady
melichercik@nh.cas.cz, sovova@nh.cas.cz, khabiri@nh.cas.cz,
kulik@nh.cas.cz, zayats@nh.cas.cz, ettrich@nh.cas.cz

Molecular dynamics (MD) as a computational method is based on the atoms' interactions. Hereby a classical description is used based on the Newton equation in motion with a consistent system of empirical parameters derived from experiments and quantum calculations (force field). The simulations are able to describe motions and conformational changes within the system on a molecular level (advantage over experimental methods) and it allows simulations of much larger systems when compared to quantum methods. Presently, it is possible to simulate systems containing multimeric systems of proteins, nucleic acids, lipid membranes, etc. (hundreds of thousands of atoms, including the solvent) up to several hundreds of ns on an all-atom scale. Data generated in this way often corresponds to physiologically relevant timescales and system sizes, and therefore allows to describe complex quarternaly changes in the systems and to propose functional hypothesises e.g. about the mechanism of enzyme functionality. Of course, comparison of these results with experimental data is essential for interpretation and validation. Computational resources of MetaCentrum are used mainly for the molecular dynamics simulations in Gromacs, that typically take thousands CPU hours and that are preferrably run in paralell. Much less frequently, Gaussian is used to optimize small molecules and subsequently to calculate partial charges on their atoms.

Project 1: Functionality of Arginine Repressor

Arginine repressor (ArgR) is the master regulator of the arginine regulon in a wide variety of bacteria [1], acting as direct sensor and transcriptional transducer of intracellular L-arginine (L-arg) concentrations to provide feedback control over biosynthesis and catabolism of L-arg. The co-effector L-arg binds to a central hexamerization domain, altering DNA affinity and specificity [2] of peripheral domains. The structural organization of ArgR into N- (ArgRN) and C-terminal (ArgRC) domains, and the functional division of labour between them, are conserved even among distant orthologs that display an unexpected diversity of reported biochemical properties, notably the L-arg dependence of hexamerization and DNA-binding equilibria [3-7]. An allosteric mechanism was inferred by comparison of crystallized intact unliganded apoprotein from the

thermophile *Bacillus stearothermophilus* with its liganded C-terminal domain fragment, which differ by $\sim$ 15 degrees rotation about the trimer-trimer interface that was ascribed to L-arg binding and presumed to be transmitted to the DNA-binding domains [8]. A similar degree of rotation was reported recently between apo- and holoArgRC of *Mycobacterium tuberculosis* [9]. Unlike the *Bacillus* and *Mycobacterium* proteins, crystal structures of the *E. coli* ArgR C-terminal domain with (holoEcArgRC) and without (apoEcArgRC) bound L-arg are essentially identical [10]. Crystalline apo- and holoEcArgRC hexamers are also entirely symmetric, a finding that is seemingly incongruent with the complex thermodynamics of L-arg binding. Isothermal titration calorimetry (ITC) confirms that EcArgR and EcArgRC hexamers bind six equivalents of L-arg, but with a multiphasic binding mechanism in which the first binding event has $\sim$ 100-fold greater affinity than the subsequent five events [11]; thus L-arg binding is itself allosteric.

Molecular dynamics simulations with ArgRC, the hexameric domain that binds L-arginine with negative cooperativity, reveal that conserved arginine and aspartate sidechains in each ligand-binding pocket promote rotational oscillation of apoArgRC trimers by engagement and release of hydrogen-bonded salt bridges. Binding of exogenous L-arginine displaces resident arginine residues and arrests oscillation, shifting the equilibrium quaternary ensemble and promoting motions that enable formation of a cooperative hydrogen-bond network among subunits while maintaining the configurational entropy of the system. A single L-arg ligand is necessary and sufficient to arrest oscillation, and the symmetry of the hexamer is maintained despite the conceptual asymmetry of the partially-liganded state. The results also offer the first opportunity to describe in structural and thermodynamic terms the symmetric relaxed state predicted by the concerted allostery model of Monod, Wyman, and Changeux [12], revealing that this state is achieved by exploiting the dynamics of the assembly and the distributed nature of its cohesive energy. This example reveals that symmetry can be maintained even when binding sites fill sequentially due to negative cooperativity, which was not anticipated by the Monod, Wyman, and Changeux model. The molecular mechanism identified here neither specifies nor requires a pathway for transmission of the allosteric signal through the protein, and it suggests the possibility that binding of free amino acids was an early innovation in the evolution of allostery.

Simulations of intact protein (both N and C domains) were based on apo crystal structures of protein from *Bacillus stearothermophilus* and *Bacillus subtilis*, holo crystal from *Mycobacterium tuberculosis* and model based on known N and C domains from *Escherichia coli*. These simulations revealed rotation in apo state similar to the simulations where only C domains were present. In holo state the systems show no rotation. Unfortunately crystal structured are known only in apo (both *Bacilli*) or holo (*Mycobacterium*) state for each bacteria, never both for any species. Therefore there is problem to direct comparison of same protein in both states. The workaround is removing/adding ligand molecules (6 Arg) from holo-/to apo-state. But simulations started from these structures

didn't reach the plateau phase and there are still modification of the structure; even with simulated time over 100 ns. The next step is also the study of transitions between states differing with number of bounded ligand's molecules (apo state vs. state with 1 bounded Arg; state with 1 bounded Arg vs. state with 2 Arg). We are currently trying to use also metadynamics method to study this changes—both in energetical and structural way.

Bibliography

[1] W. K. Maas, The arginine repressor of *Escherichia coli*, Microbiol. Mol. Biol. Rev. 58 (1994) 631-640.

[2] D. Szwajkajzer, L. Dai, J.W. Fukayama, B. Abramczyk, R. Fairman, J. Carey, Quantitative analysis of DNA binding by the *Escherichia coli* arginine repressor, J. Mol. Biol. 312 (2001) 949-962.

[3] D. Lim, J.D. Oppenheim, T. Eckhardt, W.K. Maas, Nucleotide sequence of the argR gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 6697-6701.

[4] L.G. Czaplewski, A.K. North, M.C. Smith, S. Baumberg, P.G. Stockley, Purification and initial characterization of AhrC: The regulator of arginine metabolism genes in *Bacillus subtilis*, Mol. Microbiol. 6 (1992) 267–275.

[5] C.-D. Lu, J.E. Houghton, A.T. Abdelal, Characterization of the arginine repressor from *Salmonella typhimurium* and its interactions with the carAB operator, J. Mol. Biol. 225 (1992) 11-24.

[6] M. Dion, D. Charlier, H. Wang, D. Gigot, A. Savchenko, J.-N. Hallet, N. Glansdorff, V. Sakanyan, The highly thermostable arginine repressor of *Bacillus stearothermophilus*: Gene cloning and repressor-operator interactions, Mol. Microbiol. 25 (1997) 385-398.

[7] A. Morin, N. Huysveld, F. Braun, D. Dimova, V. Sakanyan, D. Charlier, Hyperthermophilic *Thermotoga* arginine repressor binding to full-length cognate and heterologous arginine operators and to half-site targets, J. Mol. Biol. 332 (2003) 537-553.

[8] J. Ni, V. Sakanyan, D. Charlier, N. Glansdorff, G.D. Van Duyne, Structure of the arginine repressor from *Bacillus stearothermophilus*, Nat. Struct. Biol. 6 (1999) 427-432.

[9] L.T. Cherney, M.M. Cherney, C.R. Garen, G.J. Lu, M.N. James, Structure of the C-domain of the arginine repressor protein from *Mycobacterium tuberculosis*, Acta Cryst. D64 (2008) 950-956.

[10] G.D. Van Duyne, G. Ghosh, W.K. Maas, P.B. Sigler, Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*, J. Mol. Biol. 256 (1996) 377-391.

[11] L. Jin, W.F. Xue, J.W. Fukayama, J. Yetter, M. Pickering, J. Carey, Asymmetric allosteric activation of the symmetric ArgR hexamer, J. Mol. Biol. 346 (2005) 43-56.

[12] J. Monod, J.-P. Changeux, F. Jabcob, Allosteric proteins and cellular contorol systems, J. Mol. Biol. 6 (1963) 306-329.

Project 2: Molecular Dynamics as a Tool in Protein Studies

NKR-P1 Receptors

Natural killer (NK) cells are the third major group of lymphocytes derived in blood morrow. They do not express T-cell or B-cell receptors and mediate natural killing against prototype NK-cell-sensitive targets. NK cells are involved in rejecting tumorous, infected or differently harmed cells by the innate immune system. They distinguish these broken cells from normal cells by interacting with receptors on the cellular surface. This mechanism is described by so-called missing-self hypothesis [1].

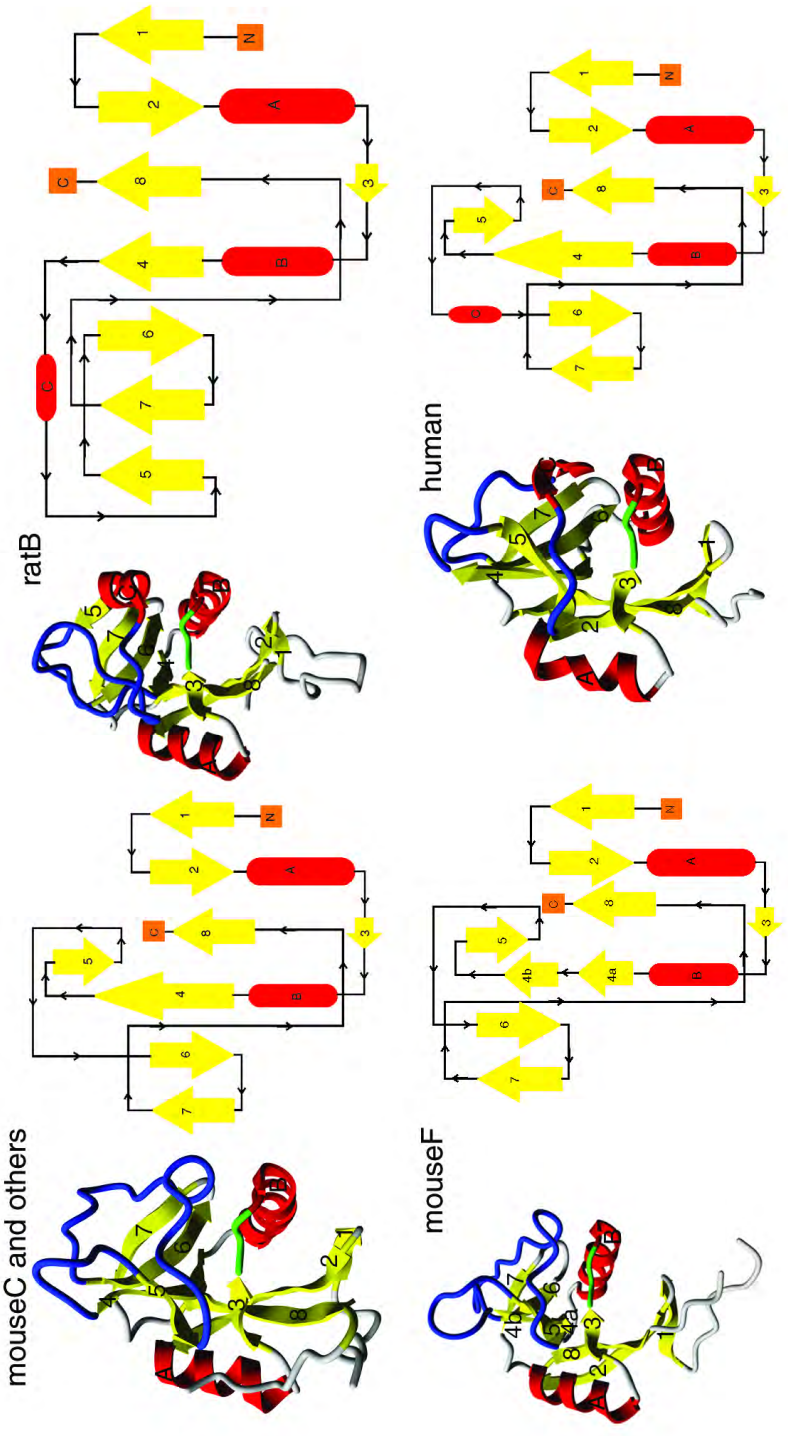
One group of NK cell receptors on cellular surface are the NKR-P1 receptors, that are homodimeric type II transmembrane proteins each containing extracellular C-type lectin-like domains. Functional NKR-P1s can be found in rat and mice as well as in human and chicken. Extensive searches of genome databases [2, 3] revealed, that orthologues of these receptors are present also in most other mammals. However functionality of these have not yet been proofed. Two types of ligands were reported to interact with NKR-P1 receptor: saccharides [4] and Clr proteins [5, 6].

As no crystal structure of NKR-P1 receptor was published till the end of 2010, homology models (in Modeller software [7]) of the extracellular domain were generated for human NKR-P1, mouse NKR-P1A and NKR-P1C and rat NKR-P1A and NKR-P1B. Later, based on phylogenetic analysis performed within the project, models of mouse NKR-P1E and NKR-P1F were built as well. All the models were minimized in Gromacs software [8], some partial charge calculations were done in Gaussian [9] using the computational facility of MetaCentrum.

Four different folds were identified for these receptors, with one of them the prevailing fold. The remaining three folds are represented by only one member each, which may show that these unique folds are important for proper receptors function. Calculated models were validated and their principal correctness confirmed by spectroscopic techniques. For further details, see ref. 10.

PsbP Protein

Oxygenic photosynthesis is an enzymatic process that converts sunlight into chemical energy. It takes place in thylakoid membranes of phototrophic organisms where different protein-pigment catalyzers are embedded. The first one in the cascade—the photosystem II (PSII)—is in higher plants dimeric structure


ratB
human
mouseC and others
mouseF

where each monomer has 16 transmembrane protein subunits and 3 so-called extrinsic proteins that are attached to the lumenal side of the thylakoid membrane. They stabilize and protect oxygen evolving complex splitting water molecule into oxygen ion and 2 electrons.

The reported project was solved together with crystalographers, who had resolved a crystal structure of one of extrinsic proteins of PSII designed as PsbP, but the resultant structure missed two loop region, one of 5 and the second of 20 amino acids [11]. The task was to model these missing regions and examine their dynamic behavior.

This task showed to be a bit tricky as the helix preceeding the larger of the two missing loops got unastable in most attempts to run a molecular dynamics simualtion, loosing its secondary structure in an early stage of the simulation, as probably its stability is influenced by the hydrogen bonding network of the residues missing in the crystal structure and added by loop modeling. This effect then led to destabilization of the whole protein. This was overcome by a step by step procedure in loop modeling, basically applying restraints on all the protein structure but the missing loops until a stable supporting hydrogen bonds network was established in that region. As predicted, both modelled loops in the final equilibrated structure, though stable in its conformation, showed a higher flexibility when compared to average of all amino acids.

Bibliography

[1] Ljunggren HG, Karre K (1990) In search of 'missing self': MHC molecules and NK cell recognition. Immunol. Today 11:237–244 homodimeric type II transmembrane proteins each containing extracellular C-type lectin-like domains

[2] Zelensky AN, Gready JE (2003) Comparative analysis of structural properties of the C-type-lectin-like Domain (CTLD). Proteins 52: 466–477

[3] Hao L., Klein J., Nei M (2006) Heterogenous but conserved natural killer receptor gene complexes in four major orders of mammals. PNAS 103:3192–3197

[4] Bezouška K, Yuen CT, O'Brien J, Childs RA, Chai W, Lawson AM, Drbal K, Fišerová A, Pospíšil M, Fiezi T (1994) Oligosaccharide ligands for NKR-P1 protein activate NK cells and cytotoxicity. Nature 372:150–157

[5] Kveberg L, Dai KZ, Westgaard IH, Daws MR, Fossum S, Naper C, Vaage JT (2009) Two major groups of rat NKR-P1 receptors can be distinguished based on chromosomal localization, phylogenetic analysis and Clr ligand binding. Eur. J. Immunol. 39: 541–551

[6] Tian W, Nunez R, Cheng S, Ding Y, Tumang J, Lyddane C, Roman C, Liou HC (2005) C-type lectin OCILRP2/Clr-g and its ligand NKR-P1f costimulate T cell proliferation and IL-2 production. Cell. Immunol. 234:39–53

[7] Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Science 9:1753–1773

[8] Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. 4:435–447

[9] Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. et al., Gaussian, Inc., Wallingford CT, 2004.

[10] Sovova Z, Kopecky V Jr., Pazderka T, Hofbauerova K, Rozbesky D, Vanek O, Bezouska K, Ettrich R (2010) Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop region invelved in ligand specificity. J. Mol. Model. Doi 10.1007/s00894-010-0837-y

[11] Kopecký V, Kohoutová J, Lapkouski M, Hofbauerová K, Sovová Ž, Kutá-Smatanová I, Revuelta JL, Arellano JB, Ettrich R (2011) Drop coating deposition Raman—a powerful new tool demonstrated on extrinsic PsbP protein of PSII from Spinacea oleracea. Submitted.

Project 3: Computational Study Identifies Mutual Effect of Organic Solvents and Dehalogenases

Based on the growing knowledge of the influence of organic solvents on enzyme structure, function and dynamics and the possibilities offered by protein engineering to improve enzyme function in nonaqueous media, there has been significant interest to know the molecular behaviors of solvents with enzyme strucuture. For this reason to investigate the effect of the organic solvents and protein relative to each other we focus on the topology of the organic solvents around the enzymes surface. We used for our computational study Haloalkane dehalogenases in different organic solvents, to be able to compare the results with experimental data gained in Prof. Damborsky's group at Masaryk University. We simulated several systems using the explicit water model and different percentages of four organic solutions as follow: ACETONE 20%, Formamide 5%, Isopropanol 10%, DMSO 42%. All simulations for these solutions were simulated for at least 20 ns till they will be best homogenous. Three different dehalogenases (DhaA, DhaA57 (mutant), LINB, DBJA) were solvated inside these solutions for 35 ns. Since the protein has low dielectric constant in solution the behavior of the organic solvents near the enzyme surface is similar to the behavior of organic solvents at air (with dielectric constant ~ 1) water solution. The formamide stays flat between polar and non polar part of the enzyme surfaces meanwhile acetone and ispn are mainly stay at nonpolar area of the surface while their methyl group is faced to this regions. The comparison of B-factors in different organic solvent shows the structural stability of the enzymes in all organic solvents. Since DbjA is a dimmer the organic solvents have more additional effect on monomers behavior relative to each other. From physico-chemical properties sight the water and formamide are very close to each other then it is reasonable that the angle between monomer A and monomer B in formamide and water has slightly different while the nature of acetone and ispn cause increasing the angle between these two monomers. In addition, the covariance shows that the

penetration of acetone and ispn between two monomers make the motion of two monomers more stable. Therefore we are able to conclude that the comparison of these properties with the function and activity of these three enzymes with the structural effect of organic solvents on the enzyme surfaces may give a useful approximation of the general effect of organic solvents relative to their physico-chemical properties with considering the properties of proteins surface.

Project 4: Computational Modelling of Catalytic Properties and Modified Substrates of Glycosyl hydrolases

Glycosyl hydrolases (also called glycoside hydrolases or glycosidases) are a large group of enzymes, which catalyze the hydrolysis of the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety.

At the beginning a screening study showed N-acetyl galactosamindase (NAGALase) activity for an enzyme from *A. niger*, and further purification hinted that this activity comes from a mixture of several Galactosidases (GALases) from *A. niger*. In the next step this NAGALase was biochemically characterized, and in a concerted effort with the experimentalists we started to elucidate the exact structure-functional relationship.

N-terminally sequencing, however, made the story much more complicated than we first thought, as it identified the gene that was in the *A. niger* genome assigned as encoding GALase gene variant A, but we clearly found a NAGALase.

Analysis of a series of 3D models generated by us and comparison to solved crystal structures of characterized enzymes revealed the significant difference in the size of the active centers in *aglA* and *aglB* and lead to an explanation of the specificity for hydrolyzed carbohydrates. Substrate docking clearly demonstrated the preference of *aglA* for α-D-N-acetylgalactosamine over galactose due to the active site extension in the vicinity of the substrates N-acetyl group that we call 'N-acetyl recognition loop'. Therefore we can state that the GALase type A gene from *Aspergillus niger* does not encode a GALase, and therefore the gene was wrongly assigned, but encodes *Aspergillus niger* NAGALase that we were able to characterize structurally and functionally.

Additionally, using our earlier reported model of beta-N-hexosaminidase (HEX) from *A. oryzae* including dimerization and glycosylation, we examined the interaction of active site amino acids with natural substrate and modified substrates and inhibitors. Hereby, the importance of the C3-atom fixation in the active site was shown. Deoxydation and dehydration at C-4 carbon on pyranosyl ring lead to inactive conformational of substrate. We generated a homology model of HEX from *P. oxalicum* was done and compared with *A. oryzae*. Generally 3D structures of both enzymes are similar. Observed differences in substrate affinity to N-acetyl modified substrates explained by easier penetration of substrates into the active site of *P. oxalicum* due to the sequence difference in the loop of *P. oxalicum* making it more flexible. Other significant difference in the kinetics between two enzymes is two types of inhibition in *P. oxalicum* by products

(GlcNAc and GalNAc) and substrate pNP-GlcNAc access. Secondary binding place was found. *A. oryzae* has some mutations in this place enabling GlcNAc binding.

The sources of METAcentrum were use for MD calculation by Yasara. Substrate–enzyme complexes of galactoaminidase from *A. niger* and hexosaminidases from *A. oryzae* and *P. oxalicum* and modified carbohydrates were studied.

Project 5: Structure and Functions of Transient Receptor Potential Channel TRPA1

Transient receptor potential (TRP) channels are a large superfamily of nonselective cation channels. TRPA1 is a candidate for mechanically gated transduction channels potentially mediating the sensations of hearing, touch, and some forms of pain. Human TRPA1 is a 127.4 kDa protein comprised of 1119 amino acids. Like other TRPs also TRPA1 has six predicted membrane-spanning domains (S1 to S6) and the pore between S5 and S6. In this work we focus on homology modeling of its for TRPs unusually long N-terminal intracellular region containing 18 predicted ankyrin repeats. Ankyrin repeats have been implicated in protein-protein interactions, provide elasticity and make molecular springs. Also a calcium-binding domain, EF-hand, was indicated at the N-terminus, consisting of 12 residues involved in Ca-dependent activation. Simulations of the dynamic behavior of tree-dimensional all-atom models indicate stability and equilibration, and let us describe structural and functional properties to understand the system. Structural models are build using Modeller, for visual analyzing and energy minimization of the created models Yasara is used, and molecular dynamics simulations are carried out in GROMACS (molecular dynamics simulation package). The general aim is to embed the results of this work later into an all-atom model of the channel the membrane to get a stable tetrameric overall structure of fully functional TRPA1 in its natural environment.

Bibliography

[1] Doerner, J.F., Gisselmann, G., Hatt, H., Wetzel, C.H.: Transient Receptor Potential Channel A1 Is Directly Gated by Calcium Ions, *JBC*, 2007, *282*, 13180–13189.

[2] Benedikt, J., Samad, A., Ettrich, R., Teisinger, J., Vlachova, V.: Essential role for the putative S6 inner pore region in the activation gating of the human TRPA1 channel, *BBA*, 2009, *1793*, 1279-1288.

[3] Michaely, P., Tomchick, D.R., Machius, M., Anderson, R.G.: Crystal structure of a 12 ANK repeat stack from human ankyrinR, *EMBO J.*, 2002, *21*, 6387–6396.

Programs

Gromacs, Yasara, Gaussian 03

Acknowledgments

We gratefully acknowledge support from the Ministry of Education, Youth and Sports of the Czech Republic (MSM6007665808, LC06010), Academy of Sciences of the Czech Republic (AVOZ60870520), Grant Agency of the Czech Republic (203/08/0114, P207/10/1934 and KJB101120805), Ministry of Education of the Czech Republic (MSM 0021620835) and joint Czech–US National Science Foundation International Research Cooperation (ME09016 and NSF INT03-09049). Access to METACentrum supercomputing facilities was provided under the research intent MSM6383917201.

Publications Acknowledging Metacentrum

Rebecca Strawn, Milan Melicherčík, Michael Green, Thomas Stockner, Jannette Carey, Rüdiger Ettrich (2010) Symmetric allosteric mechanismof hexameric Escherichia coli arginine repressor exploits competition between L-arginine ligands and resident arginine residues PLOS Computational Biology 6: 6. e1000801 (open access).

Rebecca Strawn, Thomas Stockner, Milan Melicherčík, Lihua Jin, Wei-Feng Xue, Jannette Carey, Rüdiger Ettrich (2011) CHAPTER FIVE: Synergy of molecular dynamics and isothermal titration calorimetry in studies of allostery. In: Methods in Enzymology, Biothermodynamics, Part D Volume 492 Edited by: Michael L. Johnson, Jo M. Holt, Gary K. Ackers. pp. 151-189 Academic Press - Elsevier Publishing Company, ISBN: 978-0-12-386003-3

Natallia Kulik, Lenka Weignerová, Tomáš Filipi, Petr Pompach, Petr Novák, Hynek Mrázek, Kristýna Slámová, Karel Bezouška, Vladimír Křen, Rüdiger Ettrich (2010) The α-galactosidase type A gene aglA from Aspergillus niger encodes a fully functional α-N-acetylgalactosaminidase Glycobiology 20: 11. 1410-1419.

Kristýna Slámová, Radek Gažák, Pavla Bojarová, Natallia Kulik, Rüdiger Ettrich, Helena Pelantová, Petr Sedmera, Vladimír Křen (2010) 4-Deoxy-substrates for beta-N-acetylhexosaminidases: how to make use of their loose specificity Glycobiology 20: 8. 1002-1009.

Natallia Kulik and Kristýna Slámová (2011) Computational Modelling of Catalytic Properties and Modified Substrates of Fungal β-N-Acetylhexosaminidases Mini-Reviews in Organic Chemistry, scheduled for publication 9/2011

Karel Bezouška, Helena Ryslava, Alzbeta Kalendova, Veronika Doubnerova, Premysl Skocdopol, Anna, Hodkova, Zdenek Kukacka, Kristina Slamova, Pavla Bojarova, Petr Pompach, Vinay Kumar, Natallia Kulik, Rüdiger Ettrich, Vladimir Kren (2011) Enzymatic characterization and molecularmodeling of an evolutionarily interesting fungal beta-N-acetylhexosaminidase FEBS Journal accepted with minor revisions

Žofie Sovová, Vladimír Kopecký Jr, Tomáš Pazderka, Kateřina Hofbauerová, Daniel Rozbeský, Ondřej Vaněk, Karel Bezouška, Rüdiger Ettrich (2011) Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop regioninvolved in ligand specificity Journal of Molecular Modeling, online 2010 Sep 14: DOI: 10.1007/s00894-010-0837-y

Vladimir Kopecky Jr, Jaroslava Kohoutova, Mikalai Lapkouski, Katerina Hofbauerova, Zofie Sovova, Ivana Kuta Smatanova, Jose L Revuelta, Juan B Arellano,Rüdiger Ettrich (2011) Can drop coating deposition Raman spectroscopy distinguish shortcomings or inaccuracies in protein crystals? Case study of PsbP protein of photosystem II from Spinacia oleracea. Submitted.

Research Projects of Loschmidt Laboratories in 2010

Jan Brezovský and Jiří Damborský

Loschmidt Laboratories, Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic
briza@chemi.muni.cz, jiri@chemi.muni.cz

People Involved in the Projects

David Bednář (bednar)
Lada Biedermannová (lada)
Jan Brezovský (briza)
Jiří Damborský (jiri)
Lukáš Daniel (lukecz)
Artur Gora (artgora)
Eva Chovancová (akllupe)
Antonín Pavelka (tonda)

Virtual Screening of Substrate Specificities of Four Enzymes from Haloalkane Dehalogenase Family

Docking-based virtual screening method was applied to study substrate specificity of haloalkane dehalogenases (HLDs) which have great potential for bioremediation or decontamination of broad range of important environmental pollutants in this project. A database containing 41,366 possible substrates of HLDs has been screened against four HLDs—DhaA, DbjA, Dmba and LinB, revealing the most promising potential substrates of HLDs (Figure 1). 102 substrates with better binding affinity and reactivity comparable to the well-known substrates were identified. As validation of methodology, 17 of these compounds were purchased and tested experimentally with DbjA. More than 85% of tested molecules were identified as substrates to DbjA, demonstrating that the virtual screening methodology applied in this project was very successful. Moreover, novel compounds previously not considered as the substrates for HLDs were identified.

Structural Basis of Enantioselectivity in Haloalkane Dehalogenases

The structural basis of enantioselectivity of HLDs and the possibility to its modulation are shown using enantiodiscrimination of two classes of substrates by DbjA as the example (Figure 2). The discrimination of α-bromoesters is due to hydrogen-bonding of the substrate's halogen and carbonyl oxygen with a pair of catalytic halide-stabilizing residues, which is in conformity with observed dominance of enthalpy. On the other hand, the enantioselectivity of β-bromoalkanes

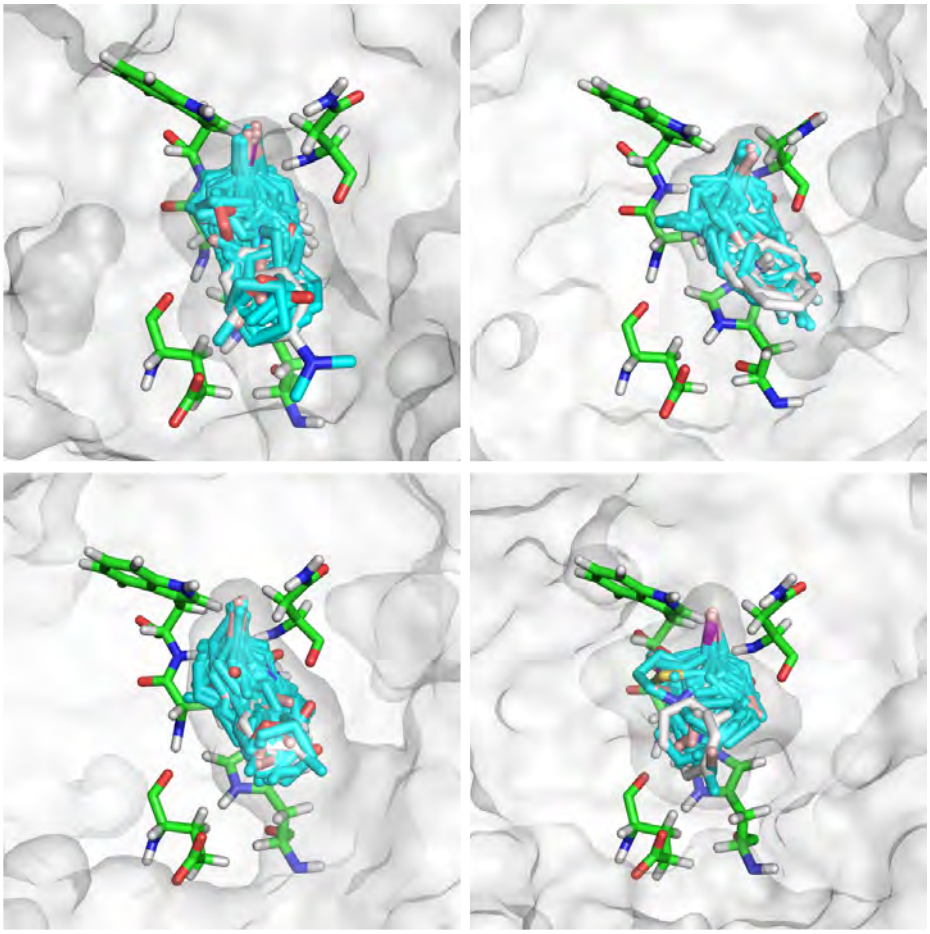


Fig. 1. Top 50 potential substrates of HLDs identified by virtual screening. Cyan sticks represents substrates docked into the active site of DbjA (Upper left), DhaA (Upper right), DmbA (Lower left) and LinB (Lower right). Catalytic residues are shown in green sticks.

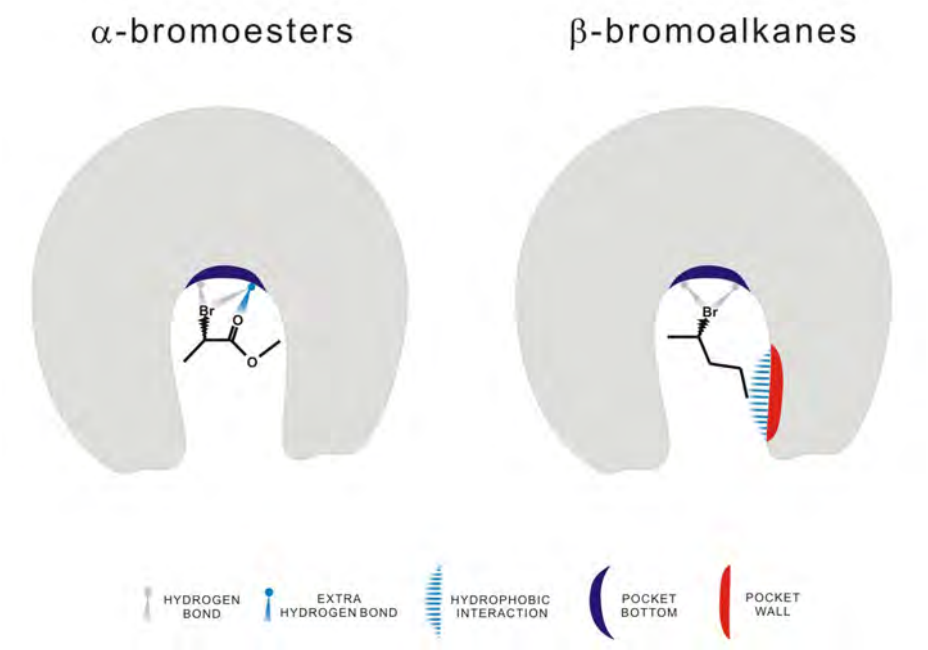


Fig. 2. Two structural bases of enantioselectivity of DbjA. The discrimination of α-bromoesters is due to *hydrogen-bonds* between a pair of catalytic halide-stabilizing residues located at the bottom of the pocket and the substrate's halogen and carbonyl oxygen, while enantiodiscrimination of β-bromoalkanes is due to *hydrophobic interaction* between the alkyl chain and the wall of the pocket in addition to the *hydrogen bonds* formed by their halogens.

is governed by entropy, which is reflected by importance of hydrophobic interactions between the alkyl chain of the substrate and the wall of the active site. Accordingly, the introduction of mutations affecting the structure of DbjA can modify its enantioselectivity towards one group of substrates, without affecting its selectivity towards another group. These findings highlight the complexity of enzymes' active sites, and the diversity of substrate binding modes. These diverse sources of enantioselective recognition within a single active site need to be considered in the rational design of enantioselective biocatalysts.

Financing Sources

Ministry of Education, Youth and Sports of the Czech Republic (LC06010, MSM0021622412 and MSM0021622413)
Grant Agency of the Czech Republic (203/08/0114 and 201/07/0927)
Grant Agency of the Czech Academy of Sciences (IAA401630901)

European Developmental Fund (CZ.1.05/2.1.00/01.0001 and CZ.1.05/1.1.00/02.0123)

Used Programs and Applications

AMBER 7 – 10
AUTODOCK 3, 4
BLAST 2.2.20
CD-HIT 07.01.31
CHARMM C33
FOLDX 3.0 beta4
GAUSSIAN 94, 03
MUSCLE 3.6
PYMOL 0.99 – 1.1
RATE4SITE 2.01
VMD 1.8.5 – 1.8.7

Publication dedicated to MetaCentrum:

Prokop, Z., Sato, Y., Brezovský, J., Mozga, T., Chaloupková, R., Koudelaková, T., Jeřábek, P., Štěpánková, V., Natsume, R., van Leeuwen, J.G.E., Janssen, D.B., Florián, J., Nagata, Y., Senda, T., Damborský, J., 2010: Enantioselectivity of Haloalkane Dehalogenases and its Modulation by Surface Loop Engineering. *Angewandte Chemie-International Edition* 49: 6111-6115. **IF = 11.829**

Patent application prepared using MetaCentrum resources:

Prokop, Z., Koudeláková, T., Štěpánková, V., Chaloupková, R., Chovancová, E., Gora, A.W., Brezovský, J., Damborský, J. Improving Stability of Proteins at Higher Temperatures and in Presence of Organic Solvents by Engineering of Tunnels and Channels. Masaryk University, Brno, Czech Republic, 2011.

Software developed using MetaCentrum resources:

HOTSPOT WIZARD 1.0, `http://loschmidt.chemi.muni.cz/hotspotwizard`, tool for prediction of protein hot spots

CAVER 2.0, `http://www.caver.cz`, tool for calculation of tunnels and channels in biomolecules

Presentations and posters prepared using MetaCentrum resources:

Damborský, J., Koudelaková, T., Štěpánková, V.: Structure and Activity of Haloalkane Dehalogenases in Organic Co-solvents, Gordon Research Conference, July 11-16, 2010, Smithfield, USA (presented by JD).

Štěpánková, V., Khabiri, M., Minofar, B., Prokop, Z., Ettrich, R., Damborský, J.: Effect of Organic Solvents on Structure and Function of Haloalkane Dehalogenases, 8th Discussion in Structural Molecular Biology and Bioinformatics, March 18-20, 2010, Nove Hrady, Czech Republic (presented by VŠ).

Chovancová, E., Pavelka, A., Damborský, J.: Hotspot wizard: nástroj pro inženýrství enzymů, Sigma-Aldrich, May 25-28, 2010, Milovy, Czech Republic (presented by ECH).

Prokop, Z., Fořtová, A., Chaloupková, R., Pavlová, M., Chovancová, E., Brezovský, J., Klvaňa, M., Koudeláková, T., Mozga, T., Pavelka, A., Dvořák, P., Straková, M., Damborský, J.: Engineering of Enzymes by Modification of Access Tunnels, Novel Enzymes, April 14 - 17 2010, Exeter, UK (presented by ZP).

Dvořák, P., Pavlová, M., Klvaňa, M., Brezovský, J., Chaloupková, R., Prokop, Z. and Damborský, J.: Increasing Activity of Bacterial Biocatalyst with Toxic Anthropogenic Substrate using Methods of Protein Engineering. 25th International Symposium on Microscale Bioseparations, March 21-25, 2010, Praha, Czech Republic (presented by PD).

Štěpánková, V., Chaloupková, R., Brezovský, J., Khabiri, M., Ettrich, R., Damborský, J.: How do Organic Solvents Affect Structure and Function of Haloalkane Dehalogenases? XIV. Setkání biochemiků a molekulárních biologů, 20.-21. dubna 2010, Brno, Czech Republic (presented by VŠ).

Biedermannová, L., Damborský, J., Wade R.: Role of Tunnels in Enzymes. December 7-9, 2009, Methods of Molecular Simulation, Heidelberg, Germany (presented by LB).

Koudeláková, T., Chaloupková, R., Zimmer, Ch., Khabiri, M., Ettrich, R., Bornscheuer, U.T., Damborský,J.: Improvement of Haloalkane Dehalogenase Stability by Engineering of its Access Tunnels, Trends in Enzymology, June 13-17, Ascona, Switzerland (presented by TK).

Biedermannová, L., Prokop, Z., Chaloupková, R., Dvořák, P., Damborský, J., Wade, R.: Investigation of Ligand Passage in Proteins Investigation of Ligand Passage in Proteins, Computational Chemistry, August 29 - September 3, Les Diablerets, Switzerland (presented by LB).

Damborský, J., Prokop, Z., Chaloupková, R., Chovancová, E., Koudeláková, T., Brezovský, J., Štěpánková, V., Pavelka, A., Dvořák, P.: Engineering of Enzymes by Modification of Access Tunnels, GEM 2010, May 8-9, 2010, Brno, Czech Republic (presented by JD).

Prokop, Z., Sato, Y., Brezovský, J., Mozga, T., Chaloupková, R., Koudeláková T., Jeřábek, P., Štěpánková, V., Natsume, R., van Leeuwen, J.E.V., Janssen, D.B., Florian, J., Nagata, Y., Senda, T., Damborský, J.: Enantioselectivity of Haloalkane Dehalogenases and its Modulation by Surface Loop Engineering, Trends in Enzymology, June 13-17, Ascona, Switzerland (presented by ZP).

Quantum Chemical Study of Soluble Diphenyl-diketo-pyrrolopyrrole Derivatives

Petr Toman[1], Martin Weiter[2], and Martin Vala[2]

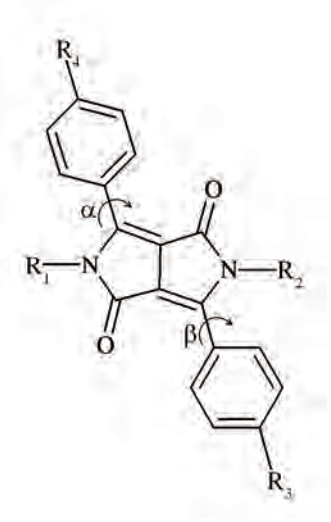
[1] Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic, Heyrovský Sq. 2, 162 06 Praha 6, Czech Republic
[2] Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 612 00 Brno, Czech Republic
toman@imc.cas.cz, weiter@fch.vutbr.cz, vala@fch.vutbr.cz

1 Introduction

Diphenyl-diketo-pyrrolopyrrole was for the first time synthesized by Farnum et al. in 1974 [1]. Its derivatives, called DPPs (see Figure 1), are high performance pigments with a wide color range from yellow-green to bluish red, that are used in paints, plastics, color inks, and other applications [2,3]. These compounds exhibit excellent photostability, resistance to chemicals, heat, and weather together with a high quantum yield of fluorescence. Thus, they are potential materials for luminescent devices [4]. However, DPPs are insoluble in water and common organic solvents due to the presence of strong hydrogen bonds in the solid state [5], and consequently common solution-based techniques (spin-coating, drop-casting, inject printing, etc.) cannot be used during the device preparation. In order to overcome this drawback, modified solubility can be achieved either through alkylation of the NH group or breaking of the conformation planarity [4]. In this paper, molecular conformations, absorption spectra, and other molecular properties of the newly synthesized derivatives are presented.

2 Methods

The optimization of the molecular conformations of the studied DPPs was carried out by means of employing both the Hartree-Fock (HF) method and the B3LYP method, which is based on the Becke three parameter functional with the non-local Lee-Yang-Parr correlation functional [6,7]. Medium size basis set 6–31G(d) was used. The B3LYP method belongs to the hybrid HF / density functional theory (DFT) methods that usually provide reliable description of different ground state molecular conformations and are used for a major part of the present-day calculations of the electronic properties. The B3LYP method was successfully used for the conformation study of many different π–conjugated systems [8–10]. Although the B3LYP method requires more computational time than the HF method, its overall computational requirements are much lower than demands of other "correlated" methods. An important difference of the results obtained by means of using HF and B3LYP methods is usually that



	$-R_1$	$-R_2$	$-R_3$	$-R_4$
DPP1	–H	–H	–H	–H
DPP2	$-C_4H_9$	–H	–H	–H
DPP3	$-C_4H_9$	$-C_4H_9$	–H	–H
DPP4	–H	–H	$-N(CH_3)_2$	–H
DPP5	–H	–H	$-N(CH_3)_2$	$-N(CH_3)_2$
DPP6	–H	–H	1-piperidinyl	1-piperidinyl
DPP7	$-C_4H_9$	–H	1-piperidinyl	1-piperidinyl
DPP8	$-C_4H_9$	$-C_4H_9$	1-piperidinyl	1-piperidinyl
DPP9	$-C_4H_9$	–H	–Cl	–Cl
DPP10	$-C_4H_9$	$-C_4H_9$	–Cl	–Cl

Fig. 1. Chemical structure and labeling of the studied diphenyl-diketo-pyrrolopyrrole derivatives and the definition of the phenyl torsion angles α and β.

the former method underestimates and the latter one slightly overestimates the electron delocalization and consequently the degree of conjugation in molecules possessing extended conjugation. The B3LYP conformations are usually close to the molecular conformations either calculated using more precise methods or obtained experimentally from structural analysis, see e.g. [11–13].

Absorption spectra of the studied derivatives were calculated using the time-dependent B3LYP (TD–B3LYP) method at the optimized B3LYP geometry. Time-dependent density functional methods recently became an effective and rather accurate tool for single point calculations of electronic excitations in various, namely conjugated, molecular systems [14–16]. However, this method is not suitable for the excited state conformation optimization necessary for luminescence spectra calculations. For this reason, relaxed (exciton) conformations of the S_1 state were optimized by means of ab initio configuration interaction method with single-excitation (CIS) method. The exciton conformations were subsequently used for the luminescence spectra calculations using TD–B3LYP method. In order to keep the same level of calculation to be able to determine the Stokes shift $\Delta E_{\mathrm{Stokes}}$ and deformation energy E_{def} of the relaxed exciton state, calculation of the first absorption peak S_1 was also performed at the optimized HF geometry.

3 Results

Figure 2 shows equilibrium molecular conformations optimized by B3LYP/6-31G(d) method. The calculated molecular parameters of the studied derivatives are listed in Table 1. The absorption spectra are shown in Figure 3. Regardless of the method, it is obvious that the calculated phenyl torsion angles α and β can be divided into two groups: angles less than ca. 17°and angles greater than ca. 27°. The differences in absolute values obtained using HF and B3LYP methods, respectively, can be explained by the different degree of the electron delocalization predicted by these methods. The results show that the N-alkylated derivatives possess significantly rotated phenyl groups of the central DPP unit in the position adjacent to the alkyl, while the basic **DPP1** molecule is completely planar. Phenyl group rotation strongly decreases the overlap between π-orbitals of the central DPP unit and the respective phenyl. As a result, the frontier orbital (HOMO and LUMO) energetic levels are shifted and consequently, absorption and luminescence spectra are modified. The lowest absorption peak energies $E_{S0\rightarrow S1}$ of the N-alkylated derivatives **DPP2** and **DPP3** exhibit a considerable hypsochromic shift (in comparison with **DPP1**) significantly correlated with the phenyl torsion angles α and β. The same effect of the N-alkylation was found also for derivatives substituted by donors or acceptors. Simultaneously, the oscillator strengths of the absorption peaks of the molecules **DPP4**–**DPP8** are notably higher than that of the unsubstituted **DPP1** molecule. On the contrary, the absorption peaks of the molecules **DPP2**, **DPP3**, and **DPP10** are reduced. These findings are in a good agreement with the experimentally measured molar absorption coefficients. The calculated results further show that Stokes shifts $\Delta E_{\mathrm{Stokes}}$ and deformation energies E_{def} of the excited state of the N-alkylated derivatives are considerably increased. The calculated luminescence peak E_{lum} depends only slightly on the phenyl torsion angles α and β. The substitution of phenyls by donor or acceptor groups has almost no influence on the phenyl torsion angles and the molecular conformation of the central DPP unit. However, it leads to the bathochromic shift of the absorption and luminescence peaks due to the increased effective extent of the conjugation.

4 Conclusions

Newly synthesized diphenyl-diketo-pyrrolopyrroles possessing electron-donating or withdrawing groups were characterized by means of quantum chemical methods. Some of these derivatives were N-alkylated on the central unit in order to increase their solubility. It was found that the most important parameter governing absorption and luminescence is the effective extent of the conjugation, which is deeply wedded with the planarity of the molecular conformation. Unlike N-alkylation leading to significant phenyl ring turns, substitution of either electron-donating or withdrawing groups breaks the molecule planarity only slightly. The main effect of the substitution of electron-donating groups is a hyperchromic and bathochromic shift in absorption spectra. This finding suggests

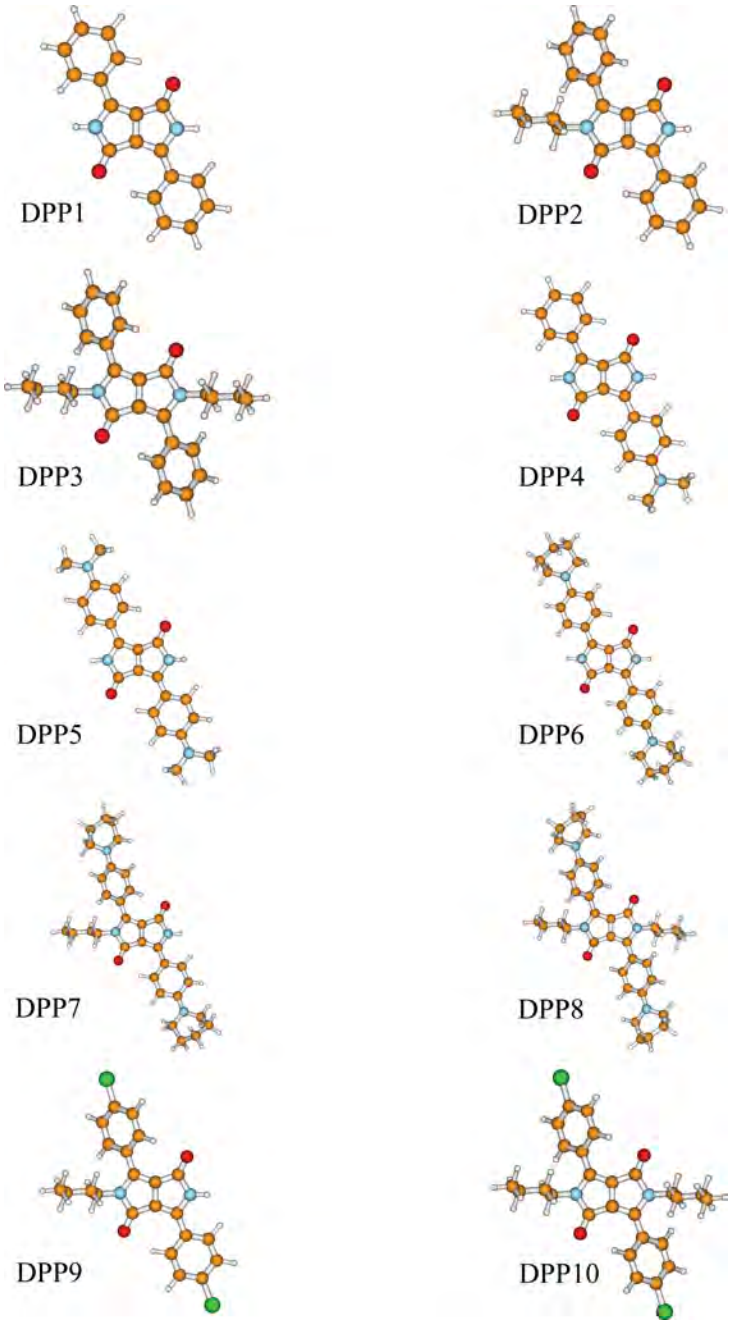


Fig. 2. Molecular conformations of DPPs optimized by the B3LYP/6–31G(d) method.

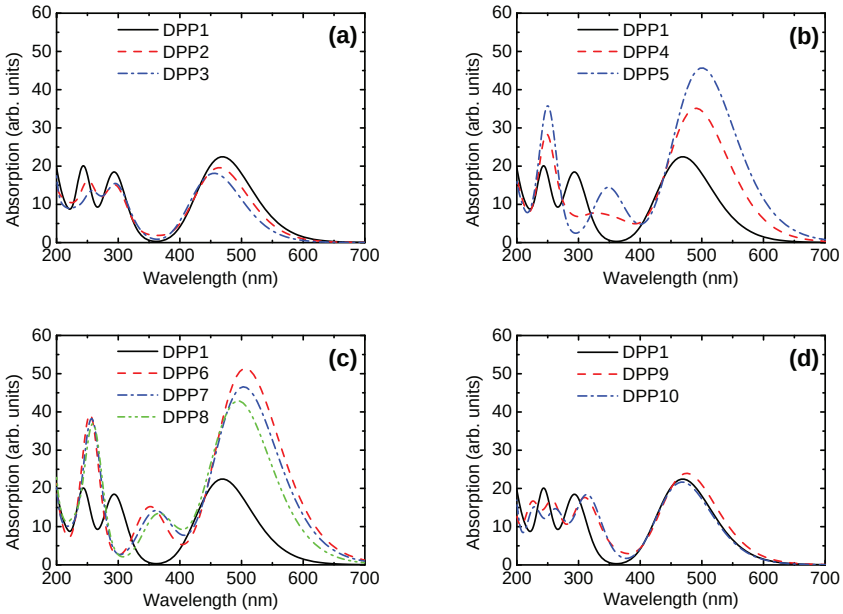


Fig. 3. Absorption spectra calculated by the TD–B3LYP/6–31G(d) // B3LYP/6–31G(d) method: The influence of *N*-alkylation (a) and substitution of various donor (b, c) and acceptor (d) groups.

	B3LYP/6–31G(d)		HF/6–31G(d)		$E_{S0\rightarrow S1}$	$f_{S0\rightarrow S1}$	E_{lum}	f_{lum}	ΔE_{Stokes}	E_{def}
	α (°)	β (°)	α (°)	β (°)	(eV)		(eV)		(eV)	(eV)
DPP1	0	0	0	0	2.84	0.49	2.43	0.53	0.41	0.34
DPP2	36	7	47	17	2.93	0.40	2.42	0.48	0.51	0.44
DPP3	36	36	46	46	3.01	0.37	2.44	0.44	0.57	0.48
DPP4	0	0	0	0	2.69	0.79	2.36	0.79	0.33	0.31
DPP5	0	0	0	0	2.64	1.03	2.32	1.02	0.32	0.29
DPP6	3	3	10	10	2.62	1.13	2.29	1.13	0.33	0.33
DPP7	27	6	43	13	2.71	0.95	2.30	1.02	0.42	0.39
DPP8	30	30	42	42	2.80	0.84	2.32	0.95	0.48	0.46
DPP9	34	7	46	16	2.87	0.49	2.37	0.58	0.51	0.44
DPP10	34	34	45	45	2.95	0.43	2.38	0.53	0.57	0.48

Table 1. Phenyl torsion angles α and β calculated by means of the HF and B3LYP methods. Lowest absorption energy $E_{S0\rightarrow S1}$, oscillator strength $f_{S0\rightarrow S1}$ of this transition, 1st luminescence peak E_{lum}, its oscillator strength f_{lum}, Stokes shift ΔE_{Stokes}, and deformation energy E_{def} of the relaxed excited state. Absorption transitions were calculated at the HF optimized geometry, luminescence peaks at the CIS optimized geometry of the relaxed excited state.

electron-accepting character of the central unit. However, it should be noted that N-alkylation reduces this effect.

References

1. D.G. Farnum, G. Mehta, G.G.I. Moore, F.P. Siegal, *Tetrahedron Lett.* **15(29)**, 2549 (1974).
2. Z.M. Hao, A. Iqbal, *Chem. Soc. Rev.* **26**, 203 (1997).
3. A. Iqbal, M. Jost, R. Kirchmayr, J. Pfenninger, A. Rochat, O. Wallquist, *Bull. Soc. Chim. Belg.* **97**, 615 (1988).
4. T. Potrawa, H. Langhals, *Chem. Ber.* **120(7)**, 1075 (1987).
5. A. Iqbal, L. Cassar, A.C. Rochat, J. Pfenninger, O.Wallquist, *J. Coat. Technol.* **60**, 37 (1988).
6. A.D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
7. C.T. Lee, W.T. Yang, R.G. Parr, *Phys. Rev. B* **37**, 785 (1988).
8. P. Toman, S. Nešpůrek, K. Yakushi, *J. Porphyr. Phthalocya.* **6**, 556 (2002).
9. J.D. Guo, Y. Luo, F. Himo, *Chem. Phys. Lett.* **366**, 73 (2002).
10. I.S. Ignatyev, T. Sundius, *Chem. Phys. Lett.* **326**, 101 (2000).
11. P. Toman, S. Nešpůrek, J.W. Jang, C.E. Lee, *Int. J. Quantum Chem.* **101(6)**, 746 (2005).
12. C.H. Choi, M. Kertesz, *Chem. Commun.* **(22)**, 2199 (1997).
13. C.H. Choi, M. Kertesz, *J. Chem. Phys.* **108(16)**, 6681 (1998).
14. M.E. Casida In Chong D.P. (ed.), Recent Advances in Density-Functional Methods, Part I, World Scientific, Singapore (1995), pp. 155–192.
15. E.K.U. Gross, J.F. Dobson, M. Petersilka In Nalewajski R.F. (ed.), Topics in Current Chemistry, Vol. 181, Density Functional Theory II, Springer, Berlin/Heidelberg (1996), pp. 81–172.
16. M. Takahashi, M. Kira, K. Sakamoto, T. Müller, Y. Apeloig, *J. Comput. Chem.* **22(13)**, 1536 (2001).

Used programs and applications

Gaussian 09

Acknowledgement

This work was supported by the Czech Science Foundation (Project No. P205/10/2280), by the Ministry of Education, Youth, and Sports of the Czech Republic, and by the Polish Ministry of Science and Higher Education (Czech–Polish cooperation project No. MEB051010 – 8194/2010).

The access to the MetaCentrum computing facilities provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" LM2010005 funded by the Ministry of Education, Youth, and Sports of the Czech Republic is highly appreciated.

List of publications dedicated to MetaCentrum

1. M. Vala, J. Vyňuchal, P. Toman, M. Weiter, S. Luňák Jr.: ,,Novel, soluble diphenyl-diketo-pyrrolopyrroles: Experimental and theoretical study", *Dyes and Pigments* **84**, 176–182 (2010).
2. S. Ehala, E. Marklík, P. Toman, V. Kašička: ,,ACE applied to the quantitative characterization of benzo-18-crown-6-ether binding with alkali metal ions in a methanol-water solvent system", *Electrophoresis* **31**, 702–708 (2010).
3. J. Kříž, P. Toman, E. Makrlík, J. Budka, R. Shukla, R. Rathore: ,,Cooperative Interaction of Hydronium Ion with an Ethereally Fenced Hexaarylbenzene-Based Receptor: An NMR and Theoretical Study", *J. Phys. Chem. A* **114**, 5327–5334 (2010).
4. P. Toman, E. Makrlík, P. Vaňura: ,,A combined experimental and theoretical study on the complexation of the ammonium ion with benzo-18-crown-6", *Monatsh. Chem.* **141**, 301–304 (2010).
5. P. Toman, E. Makrlík, P. Vaňura, V. Kašička, R. Rathore: ,,A combined extraction and DFT study on the complexation of H3O+ with a hexaarylbenzene-based receptor", *Monatsh. Chem.* **141**, 737–741 (2010).
6. P. Toman, E. Makrlík, P. Vaňura, V. Kašička, R. Rathore: ,,Theoretical study on the complexation of the sodium cation with a hexaarylbenzene-based receptor", *Monatsh. Chem.* **141**, 1309–1311 (2010).
7. E. Makrlík, P. Toman, P. Vaňura, P. Selucký, R. Rathore: ,,Experimental and DFT study on the complexation of NH_4^+ with a hexaarylbenzene-based receptor", *J. Mol. Struct.* **977(1-3)**, 254–257 (2010).
8. E. Makrlík, P. Toman, P. Vaňura, R. Rathore: ,,Extraction and DFT study on the complexation of K^+ with a hexaarylbenzene - based polyaromatic receptor", *Acta Chim. Slov.* **57**, 948–952 (2010).

List of conference presentations prepared using MetaCentrum resources

1. 18^{th} International Conference on Composites/Nano Engineering: P. Toman, S. Nešpůrek, W. Bartkowiak: *Influence of polar additives on the hole mobility in poly[(phenylene)vinylene] composites*, 4.–10. 7. 2010, Anchorage, USA.
2. 10^{th} Austrian Polymer Meeting and 2^{nd} Joint Austrian-Slovenian Polymer Meeting: P. Toman, M. Menšík, W. Bartkowiak: *Modeling of the charge carrier transport in disordered polymer materials*, 8.–10. 9. 2010, Leoben, Austria.

Application of Interval Arithmetic in Computer Assisted Proof

Jiří Benedikt *

Západočeská univerzita v Plzni

Abstract. It is proved in [3] that the second eigenfunction of the p-Laplacian, $p > 1$, on the disc is not radial. The proof is divided into three parts. As for the cases $1 < p < 1.01$ and $p > 226$, the result was proved analytically (using, e.g., asymptotical analysis). In this paper, we are concerned with the case $1.01 \leq p \leq 226$ where the proof relies on the aid of computers.

1 Introduction

Let $D \subset \mathbb{R}^2$ be the open unit disc centered at the origin. We consider the following eigenvalue problem

$$\begin{cases} -\Delta_p u = \lambda |u|^{p-2} u & \text{in } D, \\ u = 0 & \text{on } \partial D, \end{cases} \tag{1}$$

where $\Delta_p u \stackrel{\text{def}}{=} \operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right)$ is the p-Laplacian, $p > 1$, and λ is the spectral parameter. It is a well-known fact that the principal eigenfunction of (1) (corresponding to the least eigenvalue λ_1 of (1)) is a radial function which does not change the sign in D (see, e.g., Kawohl, Fridman [6]) and it is unique up to a multiple by a nonzero real number. It is also well-known that there is the second eigenvalue of (1), $\lambda_2 > \lambda_1$. There are no eigenvalues of (1) in (λ_1, λ_2), and an eigenfunction associated with λ_2 changes the sign exactly once in D (see, e.g., Anane, Tsouli [2]).

Theorem 1 ([3]). *An eigenfunction associated with λ_2 is not radial for all $p \in (1, +\infty)$.*

Let us assume $1.01 \leq p \leq 226$. We are interested in the role of computers in the proof of Theorem 1.

* The author have been supported by the Research Plan MSM 4977751301 of the Ministry of Education, Youth and Sports of the Czech Republic, and grant no. ME 10093.

2 Preliminaries

In the following, we denote $\varphi_p(s) \stackrel{\text{def}}{=} |s|^{p-2}s$, $s \neq 0$, $\varphi_p(0) \stackrel{\text{def}}{=} 0$, and $p' \stackrel{\text{def}}{=} \frac{p}{p-1}$. Let us consider the initial value problem

$$\begin{cases} -\left(r\varphi_p(u')\right)' = r\varphi_p(u) & \text{in } (0,+\infty), \\ u(0) = 1, \quad u'(0) = 0. \end{cases} \tag{2}$$

According to Del Pino, Manásevich [4, Lemma 5.2], (2) has a unique solution defined on $[0,+\infty)$ which we denote by $J_{0,p} = J_{0,p}(r)$. Moreover, [4, Lemma 5.3] implies that $J_{0,p}$ is oscillatory with zeros $0 < \nu_1(p) < \nu_2(p) < \cdots \to +\infty$, and [4, Lemma 5.1] claims that these zeros are simple.

For $p = 2$, the equation in (2) can be written as

$$u'' + \frac{1}{r}u' + u = 0, \quad r \neq 0,$$

and so the solution of (2) coincides with the Bessel function $J_0 = J_0(r)$, see Abramowitz, Stegun [1]. We have

$$\nu_1(2) = 2.4048\cdots, \quad \nu_2(2) = 5.5201\cdots, \quad \ldots$$

For $p \neq 2$, $p \in (1,+\infty)$, the solution $J_{0,p} = J_{0,p}(r)$ of (2) can be thus regarded as a generalization of the Bessel function $J_0 \equiv J_{0,2}$.

It is proved in [3, Sec. 3] that the statement of Theorem 1 holds true provided

$$2\nu_1 < \nu_2. \tag{3}$$

The proof of (3) is based on guaranteed bounds for ν_1 and ν_2 that are obtained by self-validated numerical integration of the initial value problem (2) in interval arithmetic.

3 Interval arithmetic

Lemma 1. *The condition* (3) *holds for all* p *satisfying* $1.01 \leq p \leq 226$.

The proof is based on self-validated numerical computation. Since the methods of self-validated numerical computation are not widely known, we explain some basic principles of this approach in order to convince the reader that our computer assisted proof is correct in "true" mathematical sense. Further details about the interval arithmetic and self-validated computation can be found, e.g., in Moore, Kearfott, Cloud [7].

The first thing one should bear in mind while doing a computer assisted proof is that any digital computer has only a finite number of states. Thus any system of numbers representable by a digital computer is finite and all computations related to continuous systems are inherently approximate. To decide about the validity of the numerical result, one has to keep track of the error propagation in

the course of computation. The interval arithmetic keeps this track and provides rigorous *a posteriori* bound of the error at the end of the computation. Below we are more precise about this issue.

The most common system for representing the real numbers in modern computers (finite state machines) is *double-precision floating-point* system defined by the IEEE 754 standard, see [8]. This system consists of the real numbers that can be written as

$$\pm(1 + m_1 2^{-1} + m_2 2^{-2} + \cdots + m_{52} 2^{-52}) 2^e, \tag{4}$$

where $m_i \in \{0, 1\}$ for $i = 1, 2, \ldots, 52$, $e \in \mathbb{Z}$, $-1022 \leq e \leq 1023$. Moreover, it contains 0, and also includes the symbols $-\infty$, $+\infty$ and `NaN` (not a number). Let $\mathbb{F}$ denote all the real numbers representable in this system. Since $\mathbb{F}$ is a finite set, not every real number is in this set ($\mathbb{F} \subsetneq \mathbb{R}$) and, moreover, the set $\mathbb{F}$ is not closed with respect to the four basic arithmetic operations $+$, $-$, $\cdot$, $/$, as they are defined over $\mathbb{R}$. For instance, it can be seen from (4) that the numbers 1 and 2^{-53} both belong to $\mathbb{F}$, while $1 + 2^{-53}$ does not. The expression $1 + 2^{-53}$ evaluates to 1 on a computer conforming *double-precision floating-point* IEEE 754 norm (in the rounding-to-nearest mode, see IEEE [8] or Goldberg [5]). This means that performing the four basic arithmetic operations on operands from $\mathbb{F}$ produces rounding errors if the result is not in $\mathbb{F}$. Unfortunately, due to time efficiency reasons, the IEEE 754 *lacks the possibility of tracking errors a posteriori*. This *problem is solved by* the so called *interval arithmetic* on $\mathbb{F}$.

In the interval arithmetic, each real number $x \in \mathbb{R}$ is represented by an interval $X = [\underline{X}, \overline{X}] \subset \mathbb{R} \cup \{-\infty, +\infty\}$ where $\underline{X} \in \mathbb{F} \cup \{-\infty\}$, $\overline{X} \in \mathbb{F} \cup \{+\infty\}$, $\underline{X} \leq \overline{X}$ and $x \in [\underline{X}, \overline{X}]$. Let

$$\mathbb{I}_{\mathbb{F}} \stackrel{\text{def}}{=} \left\{ [\underline{X}, \overline{X}] : \underline{X} \in \mathbb{F} \cup \{-\infty\},\ \overline{X} \in \mathbb{F} \cup \{+\infty\} \text{ and } \underline{X} \leq \overline{X} \right\}.$$

We say that the operations $\oplus, \ominus, \odot, \oslash \colon \mathbb{I}_{\mathbb{F}}^2 \to \mathbb{I}_{\mathbb{F}}$ are *interval extensions* (see, e.g., Moore [7]) of the operations $+$, $-$, $\cdot$, $/$ if

$$\forall X, Y \in \mathbb{I}_{\mathbb{F}}\ \forall x, y \in \mathbb{R} : x \in X, y \in Y \Longrightarrow$$
$$x + y \in X \oplus Y, \quad x - y \in X \ominus Y, \quad xy \in X \odot\ Y, \quad 0 \notin Y \Rightarrow \frac{x}{y} \in X \oslash Y.$$

Note that the operation $X \oslash Y$ is not defined if $0 \in Y$. In other words, $\oplus$ is an interval extension of $+$ when for any $X, Y \in \mathbb{I}_{\mathbb{F}}$,

$$[\underline{X} + \underline{Y}, \overline{X} + \overline{Y}] \subset X \oplus Y$$

holds, and similarly for the other three operations. Definitions and implementations of these operations may differ from system to system. In the system *Mathematica*®, version 7.01, where our computations were carried out, $[1, 1] \oplus [2^{-53}, 2^{-53}]$ evaluates to $[1 - 2^{-52}, 1 + 3 \cdot 2^{-52}]$ (these numbers must be entered into *Mathematica*® as machine numbers, i.e. 1.0 and 2.0^{-53}, to reproduce this result). Although this interval is obviously not optimal, we have $1 + 2^{-53} \in$

$[1-2^{-52}, 1+3\cdot 2^{-52}]$. This is the crucial difference comparing to the "standard" computer arithmetic which, as it was mentioned above, yields an approximate result 1 with no infomation about the relevance to the exact result.

The ordering relation $<$ extends naturally to intervals by

$$X \lessgtr Y \iff^{\text{def}} \forall x, y \in \mathbb{R} : x \in X, y \in Y \Longrightarrow x < y$$

which means that $X \lessgtr Y \Leftrightarrow \overline{X} < \underline{Y}$. Thus the interval arithmetic is a suitable tool for comparing quantities that are numerically computed in it. On the other hand, the equivalence relation $=$ does not extend so nicely to intervals. The natural extension of $=$ to intervals is the following

$$X \eqgtr Y \iff^{\text{def}} \forall x, y \in \mathbb{R} : x \in X, y \in Y \Longrightarrow x = y.$$

But this is true if and only if $\underline{X} = \overline{X} = \underline{Y} = \overline{Y}$. Since practical computations usually result in a nondegenerate interval, the interval arithmetic is not used to show that two computed real quantities x, y represented by intervals X, Y are equal.

In the sequel, we need also the operation of a convex hull of two intervals, i.e., the mapping $\mathrm{Hull}\colon \mathbb{I}_{\mathbb{F}}^2 \to \mathbb{I}_{\mathbb{F}}$,

$$\mathrm{Hull}\left([\underline{X}, \overline{X}], [\underline{Y}, \overline{Y}]\right) \stackrel{\text{def}}{=} \left[\min\{\underline{X}, \underline{Y}\}, \max\{\overline{X}, \overline{Y}\}\right].$$

4 Algorithm

The problem (2) is equivalent to the initial value problem for a system of two first-order equations

$$\begin{cases} u' = \varphi_{p'}\left(\dfrac{v}{r}\right), & u(0) = 1, \\ v' = -r\varphi_p(u), & v(0) = 0, \end{cases} \tag{5}$$

$r > 0$, where $v(r) \stackrel{\text{def}}{=} r\varphi_p(u'(r))$, $r \geq 0$.

Let $p \in P \in \mathbb{I}_{\mathbb{F}}$, $P \lessgtr [1.005, 1.005]$. We study (5) by means of a discrete interval-valued dynamical system (see Definition 1) of the form

$$\begin{cases} U_{i+1} = F_1(i, U_i, V_i, P), & U_0 = [1, 1], \\ V_{i+1} = F_2(i, U_i, V_i, P), & V_0 = [0, 0], \end{cases} \tag{6}$$

$i = 0, 1, 2, \ldots$, where $F_1, F_2\colon \mathbb{N} \cup \{0\} \times \mathbb{I}_{\mathbb{F}}^3 \to \mathbb{I}_{\mathbb{F}}$.

In [3, Appendix] an implementation of F_1 and F_2 in *Mathematica*® is presented, for which it is proved that the solution of (6) satisfies the following inclusion property (see [3, Prop. 15]).

Definition 1. *We say that the solution U_i, V_i, $i = 0, 1, 2, \ldots$, of (6) satisfies the* inclusion property *if for any $p \in P$, $i = 0, 1, 2, \ldots$, and $r \in \left[\frac{i}{10000}, \frac{i+1}{10000}\right]$, we have $u(r) \in \mathrm{Hull}(U_i, U_{i+1})$ and $v(r) \in \mathrm{Hull}(V_i, V_{i+1})$ where (u, v) is the solution to (5).*

With [3, Prop. 15] in hands, we can now estimate ν_1 and ν_2 from (3).

Proof of Lemma 1. The proof is computer assisted. We implemented Algorithm 1 in *Mathematica*® (see below), using the implementation of F_1 and F_2 described in [3, Appendix].

Algorithm 1 Algorithm to decide $2\nu_1 < \nu_2$

Input: $P \in \mathbb{I}_{\mathbb{F}}$, $P \ni [1.005, 1.005]$
Output: If `true`, then $2\nu_1 < \nu_2$ is proved for all $p \in P$.

```
U_0 ← [1,1]
V_0 ← [0,0]
i ← 0
while U_i ∋ [0,0] do
   U_{i+1} ← F_1(i, U_i, V_i, P)
   V_{i+1} ← F_2(i, U_i, V_i, P)
   i ← i + 1
end while
m ← i − 1
if not V_{i−1} ∋ [0,0] or not V_i ∋ [0,0] then
   return false
end if
while not U_i ∋ [0,0] do
   U_{i+1} ← F_1(i, U_i, V_i, P)
   V_{i+1} ← F_2(i, U_i, V_i, P)
   if not V_{i+1} ∋ [0,0] or not overline(U_{i+1}) < overline(U_i) then
      return false
   end if
   i ← i + 1
end while
n ← i
while i < 2n do
   U_{i+1} ← F_1(i, U_i, V_i, P)
   V_{i+1} ← F_2(i, U_i, V_i, P)
   if not U_{i+1} ∋ [0,0] then
      return false
   end if
   i ← i + 1
end while
return true
```

Since the solution of (5) (and thus also both ν_1 and ν_2) is sensitive for variations of $p > 1$, we cover the interval $[1.01, 226]$ by a number of subintervals $P_k \stackrel{\text{def}}{=} [1.0099 + 0.0001k, 1.01 + 0.0001k]$, $k \in \{1, 2, \ldots, 2249900\}$. Then we start our *Mathematica*® implementation of Algorithm 1 with P_k as the input P for each $k \in \{1, 2, \ldots, 2249900\}$. Note that P_k does not have the endpoints in $\mathbb{F}$ ($P_k \notin \mathbb{I}_{\mathbb{F}}$), but *Mathematica*® does automatically a built-in conversion of the text input P_k such that the input $P \in \mathbb{I}_{\mathbb{F}}$ into the implemented Algorithm 1

satisfies $P \in \mathbb{I}_{\mathbb{F}}$ and $P_k \subset P$. For instance, *Mathematica*® converts P_1 to

$$P = [4548635623644200, 4549085983606939] \cdot 2^{-52}$$

and P_2 to

$$P = [4549085983606937, 4549536343569676] \cdot 2^{-52}.$$

Clearly, $P \circledS [1.005, 1.005]$ for all $k \in \{1, 2, \ldots, 2249900\}$. Hence the computer actually proves $2\nu_1 < \nu_2$ for p from a larger interval P, containing P_k. The result of the computation was `true` for all $k \in \{1, 2, \ldots, 2249900\}$. Below we prove that the result `true` from Algorithm 1 guarantees $2\nu_1 < \nu_2$ for all $p \in P_k$. Consequently, (3) holds for any

$$p \in \bigcup_{k=1}^{2249900} P_k = [1.01, 226].$$

Let $k \in \{1, 2, \ldots, 2249900\}$ be fixed. As it was mentioned, with $P \supset P_k$, the result of Algorithm 1 is `true`, i.e., it reaches the last line. The fact that the first `while` loop finished implies $U_i \circledS [0, 0]$ for all $i = 0, 1, \ldots, m$. Since U_i, V_i is the solution of (6) satisfying the inclusion property, see [3, Prop. 15], it means that for any $p \in P_k$, the solution (u, v) of (5) satisfies $u(r) > 0$ for all $r \in \left[0, \frac{m}{10000}\right]$. Notice that the loop cannot be infinite since otherwise the inclusion property would imply $u(r) > 0$ for all $r > 0$. It would contradict the fact that $u \equiv J_{0,p}$ is oscillatory.

Further, the fact that the second `while` loop finished guarantees $U_n \circledS [0, 0]$ and $V_i \circledS [0, 0]$ for all $i = m, m+1, \ldots, n$. Thus, $u\left(\frac{n}{10000}\right) < 0$, $v(r) < 0$ (and hence also $u'(r) < 0$) for all $r \in \left[\frac{m}{10000}, \frac{n}{10000}\right]$. The continuity of u then yields that there is exactly one zero of u in $\left(\frac{m}{10000}, \frac{n}{10000}\right)$. Hence $\nu_1 < \frac{n}{10000}$. Notice that neither this loop can be infinite since $\overline{U_{i+1}} < \overline{U_i}$ at each step and $\mathbb{F}$ is a finite set. Thus either $\overline{U_i} < 0$ or $\overline{U_{i+1}} \geq \overline{U_i}$ must occur after a finite number of steps.

Finally, the fact that the third `while` loop finished implies $U_i \circledS [0, 0]$ for all $i = n, n+1, \ldots, 2n$, and so $u(r) < 0$ for all $r \in \left[\frac{n}{10000}, \frac{n}{5000}\right]$. Consequently,

$$\nu_2 > \frac{n}{5000} > 2\nu_1,$$

and the proof is finished. ■

Remark 1. It is important to emphasize that the proof is based on the numerical computation in the interval arithmetic using *a posteriori* rigorous error bounds. Our proof *does not* follow from any *a priori* analysis of the algorithm. Such *a priori* analysis of the algorithm for the discrete system (6) would probably be much more difficult than an analysis of the original initial value problem (5) itself.

References

1. Abramowitz, M., Stegun, I. A., Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, 9$^{\text{th}}$ printing, New York; Dover 1972, pp. 370–374.
2. Anane, A., Tsouli, N., On the second eigenvalue of the p-Laplacian, in: A. Benkirane and J.-P. Gossez (Eds.), Nonlinear Partial Differential Equations (From a conference in Fes. Morocco, 1994), Pitman Research Notes in Math., Vol. 343, pp. 1–9, Longman, Harlow, 1996.
3. Benedikt, J., Drábek, P. and Girg, P., The second eigenfunction of the p-Laplacian on the disc is not radial, to appear.
4. Del Pino, M., Manásevich, R., Global bifurcation from the eigenvalues of the p-Laplacian, J. Differential Equations 92, no. 2 (1991), 226–251.
5. Goldberg, D., What every computer scientist should know about floating-point arithmetic. ACM Comput. Surv. 23, no. 1 (1991), 5–48.
6. Kawohl, B., Fridman, V., Isoperimetric estimates for the first eigenvalue of the p-Laplace operator and the Cheeger constant, Comment. Math. Univ. Carolinæ 44 (2003), 659–667.
7. Moore, R. E., Kearfott, R. B., Cloud, M. J., *Introduction to interval analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2009.
8. IEEE standard for floating-point arithmetic. IEEE Std 754-2008 (2008) 1–58. doi: 10.1109/IEEESTD.2008.4610935
9. Wolfram *Mathematica*® Documentation Center, URL: http://reference.wolfram.com/mathematica/guide/IntervalArithmetic.html.

Utilization of MetaCentrum Computational Resources by the National Centre for Biomolecular Research

Petr Kulhánek, Alexej Kulaš, Arnošt Mládek, Barbora Benešová, Crina-Maria Ionescu, Jakub Štěpán, Jan Alán, Jan Novotný, Jan Ryška, Jan Slavík, Jan Vícha, Jana Přecechtělová, Jiří Fukal, Jiří Klusák, Jiří Wiesner, Josef Pasulka, Judit Šponer, Leona Šerá, Lucie Novosadová, Lukáš Pravda, Martin Novák, Martin Babinský, Martin Filip, Matej Pipiška, Matúš Durec, Michaela Wimmerová, Michal Ďurech, Pavel Janoš, Pavel Kadeřávek, Petr Novák, Radek Marek, Radek Matuška, Radka Svobodová Vařeková, Richard Štefl, Stanislav Geidl, Stanislav Kozmon, Stanislav Standara, Sushil Mishra, Tomáš Bouchal, Tomasz Pawlak, Veronika Papoušková, Zdeněk Kříž, Zora Střelcová, and Jaroslav Koča

National Centre for Biomolecular Research, Faculty of Science,
Masaryk University, Kotlářská 2,
611 37 Brno, Czech Republic
{kulhanek, jkoca}@chemi.muni.cz
http://ncbr.muni.cz

1 Introduction

The National Centre for Biomolecular Research (NCBR) is an independent institute of the Faculty of Science, Masaryk University, Brno. The main activities of NCBR involve research in specific areas of chemistry and biology, as well as teaching, especially at advanced and Ph.D. levels. The research activities of NCBR include projects in the following areas: Computational chemistry and molecular modeling, Computational studies on nucleic acids, NMR spectroscopy, Glycobiochemistry, RNA/Protein interactions, RNA processing and degradation, DNA repair, and Nanobiotechnology. The research is balanced between experimental and theoretical studies.

As experimental research requires good experimental facilities, theoretical research requires strong computational resources, so that the results obtained are precise enough to complement experimental observations. Such resources are available in the MetaCentrum supercomputing centre, and this article describes how they are utilized by computational chemists from NCBR in order to predict the behavior and properties of a wide range of molecular systems. Briefly, we also outline the achievements in our own software development projects, that either accelerate calculations or facilitate their execution within the supercomputing environment. Last but not least, we give an overview of the hardware contributed by NCBR to the common MetaCentrum computational infrastructure.

2 Computational Chemistry

The research effort in the computational chemistry branch is focused on several areas. The first research field represents some applications of quantum chemistry on small systems. These studies include searching for chemical reaction pathways and especially transition states, predicting the properties of supramolecules, and calculating NMR parameters. The second area comprises studies of the conformational behavior and dynamics of larger molecules with biological importance. The most widely used tools in this field are molecular dynamics and docking. Results obtained by theoretical models are validated by available experimental data – especially against NMR and X-ray data. The tuned models and methods are then able to provide information that is not available by conventional experimental tools, and thus can be helpful for experimentalists. In the following subsections we give a short overview of a selection of ongoing projects.

2.1 Molecular Dynamics Studies

Amyloids. Amyloid beta peptides (Abeta) and oxysterols (especially 24-S-hydroxycholesterol known as cerebrosterol)(Fig. 1) appear to play a role in the pathogenesis of Alzheimer disease. The experimental results demonstrated that non-aggregated or aggregated Abeta fragments could interact with racemic 24-hydroxycholesterol. However, it is not clear if the binding occurs exclusively between natural L-isomers of Abeta and natural 24-S-hydroxycholesterol, or if it is rather of nonspecific nature. The 50 ns long MD simulations of Abeta fragments, consisting of either human amino acids sequence or rat sequence in different ionic strength, were performed. The stability of secondary structure elements was analysed. The results showed that the change of ionic strength destabilizes the secondary structure element in the hydrophobic core of Abeta fragments. The results also indicate different behavior of human and rat Abeta amyloids. The human Abeta exhibits higher instability in secondary structure elements in comparison to the rat Abeta. The binding energy of hydroxycholesterol to the Abeta fragment was also evaluated by MM/GB(PB)SA analyses.

MAGE1 protein. The SMC5-6 protein complex is involved in the cellular response to DNA damage. It is composed of 6 to 8 polypeptides of which Nse1, Nse3 and Nse4 form a tight sub-complex. MAGE1 - the mammalian ortholog of Nse3, is the member of the MAGE (melanoma – associated antigen) protein family, and Nse4 is related to the EID (E1A-like inhibitor of differentiation) family of transcriptional repressors. The MD simulations of 20 different complexes of the MAGE1 protein with a 22 amino acid long EID peptide prepared by docking were performed to evaluate their stability. The MM/PB(GB)SA analyses were used for evaluating interaction energies.

Lectins. Gram-negative bacteria such as *Burkholderia*, *Pseudomonas* and *Ralstonia* are capable to live in different environments, not only in nature but also

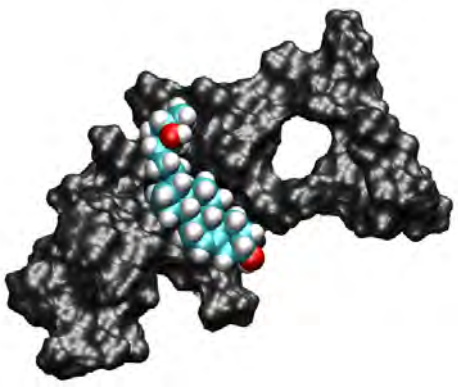


Fig. 1. Complex of Amyloid beta fragment with cerebrosterol.

in industrial components such as oil, diesel, jet fuel and shampoo. Many of these bacteria facilitate highly beneficial processes such as the breakdown of aromatic pollutants or the enhancement of plant growth. The downside of all these bacterial species is that they are also capable of causing opportunistic infections, especially in people with cystic fibrosis (CF). One of the most common infectious agents in CF patients is *Burkholderia cenocepacia*. Plant pathogens, like animal ones, use protein-carbohydrate interactions in their strategy of host recognition, attachment, and invasion. The bacterium *Ralstonia solanacearum*, which is spread worldwide and causes lethal wilt in many agricultural crops, was shown to produce a potent L-fucose-binding lectin, the *R. solanacearum* lectin, a small protein of 90 amino acids with a tandem repeat in its amino acid sequence. The molecular dynamics studies of the lectins BCLA and RSL were used in the study of the conformational behavior of these proteins. The essential dynamics method was used for evaluating cooperative movement in the proteins (Fig. 2). The study is now extended by the anisotropic thermal diffusion method, which is able to simulate the transfer of energy through the protein. The results showed the possibility of energy transfer between the different binding sites in the oligomeric lectins.

The delta subunit. RNA polymerases from gram-positive bacteria differ from RNA polymerases from gram-negative bacteria in the presence of additional subunits interacting with the core enzyme. One of these proteins is the delta subunit. The delta subunit was found to be associated with RNA polymerase of *Bacillus subtilis* almost 30 years ago, but its structure and function were unknown for a long time. Recent reports indicating the importance of the delta subunit for the virulence of *Staphylococcus aureus* and *Streptococcus agalactiae* make the delta subunit interesting also from the medical point of view. We have successfully determined and published the structure of its N-terminal domain, which is rigid and forms a standard fold belonging to the SCOP family of RNA/DNA bind-

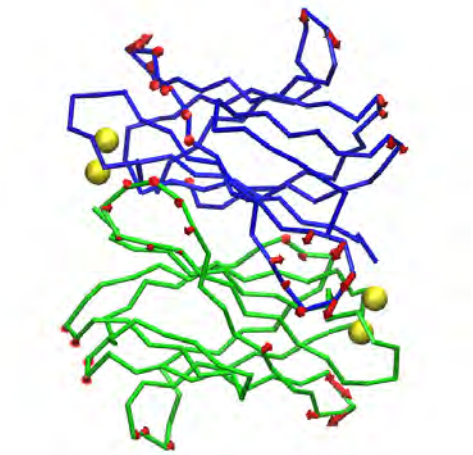


Fig. 2. Essential movements (red arrows) in BCLA.

ing proteins. On the other hand, the C-terminal domain of the delta protein is unstructured, flexible and highly negatively charged. Therefore, computational methods routinely used for NMR structure calculation are not suitable and we are trying to describe a behavior of the molecule by molecular dynamics simulation (Fig. 3).

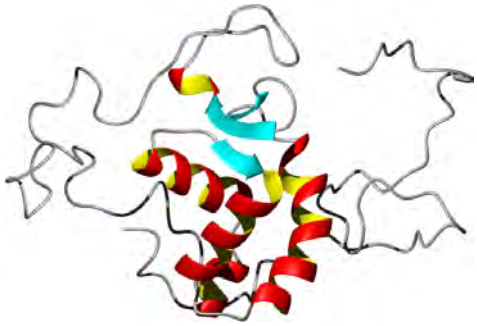


Fig. 3. Studied delta subunit

DNA/alkaloid interaction. Our current calculations are focused on protoberberine alkaloids interacting with DNA as groove binders, which was pre-

dicted by results of NMR experiments. We would like to elucidate the selectivity of this binding with respect to oligonucleotide sequence and the influence on receptor topology. Therefore MD simulations of the 15mer nonpalindromic ds-oligonucleotide with alkaloid palmatine docked inside the whole minor groove are being carried out. We have prepared 28 models of the palmatine-DNA complex, and in each such model one of 7 binding sites is occupied by one of 4 possible ligand orientations. Our aim is to run about 40ns of MD in explicit solvent and evaluate the residential time of the ligand in the individual binding sites, to try to describe the correlations between ligand movement and perturbation of DNA conformation, and to determine the most favorable ligand orientation from the MM point of view.

2.2 Quantum Chemical Calculations

The study of DNA sugar-phosphate backbone energetics. We have investigated the electronic structure and gas-phase energetics of the DNA sugar-phosphate backbone via advanced QM methods. The analysis has been carried out on biologically relevant backbone conformations composed of 11 canonical BI-DNA structures, 8 pathological structures with alfa/gamma torsion angles in the g+/t region, and 3 real noncanonical gamma-trans structures occurring in the loop region of guanine quadruplex DNA. The influence of backbone conformation on the intrinsic energetics was primarily studied using a model system consisting of two sugar moieties linked together via a phosphodiester bond (SPSOM model). To get the conformation of the studied system fully under control, for each calculation we froze the majority of the dihedral angles to their target values. CCSD(T) energies extrapolated to the complete basis set were utilized as reference values. However, the calculations show that inclusion of higher-order electron correlation effects for this system is not crucial, and complete basis set second-order perturbation calculations are sufficiently accurate. The reference QM data are used to assess the performance of 10 contemporary density functionals with the best performance delivered by the PBE-D/TZVPP combination along with Grimme's dispersion correction, and by the TPSS-D/6-311++G(3df,3pd) augmented by Jurečka's dispersion term. In addition, we compare our QM calculations to the molecular mechanics (MM) model based on the Cornell et al. force field. Finally, four additional model systems of different sizes were assessed by comparing their energetics to that of the SPSOM system. We have found that the energetics of the smaller MOSPM model, consisting of a sugar moiety linked to a phosphate group and capped with methyl and methoxy group on the 5'- and 3'-ends, respectively, is fairly similar to that of SPSOM, while the role of the undesired intramolecular interactions is diminished.

The study of dispersion interactions in biomolecular systems. Molecular recognition plays a crucial role in many biological processes, such as bacteria-host identification. Some of these recognition processes are performed by proteins called lectins, which are able to bind saccharides in very specific ways.

As examples, let us mention the world-wide spread *Ralstonia solanacearum* or *Pseudomonas aeruginosa*. Both of the mentioned bacteria interact through their lectins (RSL in *Ralstonia solanacearum* (Fig. 4) and PA-IIL in *Pseudomonas aeruginosa*) predominantly with an L-fucose moiety. Most recent studies show that the key role in the lectin-saccharide complex may be played by dispersive interactions, such as the CH-π interactions between aromatic amino-acids and non-polar faces of saccharides. These interactions were studied by high-level ab initio quantum chemistry and DFT methods. Three saccharides (L-fucose, D-mannose and D-glucose) and benzene (as representative of the aromatic moiety in amino-acids) were chosen for the following sequence of calculations: Semiempirical DFTB+ scan of both faces of the saccharide to unravel the most stable complexes with benzene. These complexes were re-optimized at the DFT-D level (DFT with empirical dispersion corrections by Grimme et al., which seem to be most suitable for this calculation), and basis-set superposition error (BSSE) corrected interaction energies were evaluated. DFT-D interaction energies were compared to CBS (complete basis set) extrapolation of MP2 and CCSD(T) interaction energies of the benzene-saccharide complexes. DFT-D results highly correlate with CCSD(T) interaction energies. Both methods also identified some structural principles of the CH-π interactions involved.

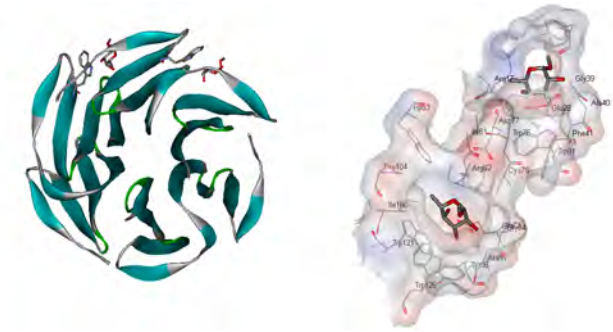


Fig. 4. RSL (left) and two disctinct binding pockets (right).

Photoremovable protecting groups. Photoremovable protecting groups are chromophoric moieties which are released from species possessing desirable physical, chemical, or biological qualities upon irradiation. They have been successfully utilized in many applications in organic synthesis, solid-phase synthesis, biology, or surface sciences. An obvious advantage of photochemical activation is the ability to precisely control the process in time and space, and the "reagentless" character of light.

A new concept of a photoremovable chiral auxiliary is introduced on the model of the chiral benzoin chromophore. This moiety can control asymmetric

formation of a Diels-Alder adduct, and then be removed in a subsequent photochemical step in high chemical yields. Selective formation of the products up to 96% ee was observed in the presence of a Lewis acid catalyst. The experimental results were supported by DFT quantum chemistry calculations. The possible orientations of $SnCl_4$ with respect to the benzoin moiety were studied (Fig. 5). The transition states of chemical reactions with cyclopentadiene were also obtained. The calculated energy barriers were used for estimating diastereomeric and enantiomeric excesses.

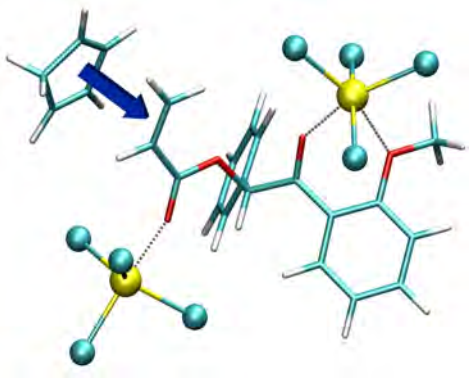


Fig. 5. The direction of cyclopentadiene attack to the benzoine moiety catalyzed by two molecules of $SnCl_4$.

Bambus[6]urils. Bambus[6]uril (BU6) is a derivative composed of glycoluril units, which are connected together with methyl bridges (Fig. 6). The structure of this macromolecule is similar to that of cucurbiturils. The work is focused on the study of the interaction of BU6 with different types of anions. Two approaches were used to study this kind of interaction: a) molecular dynamics and b) quantum mechanics. Anion binding energy was evaluated using DFT calculations. The BP86 functional and TZVPP basis set were used. The interaction energies were further corrected for the BSSE error. The dispersion corrections were also incorporated. The final binding energies predict the same binding affinity order for all hallide anions as was observed experimentally.

2.3 NMR Properties Calculations

Chemical shifts in nucleic acids. Chemical shifts in nucleic acids display large flexibility. Consequently, quantum chemical studies of these systems often require an inclusion of molecular dynamics. This makes such studies extremely demanding. We have employed a combination of an MD simulation with density

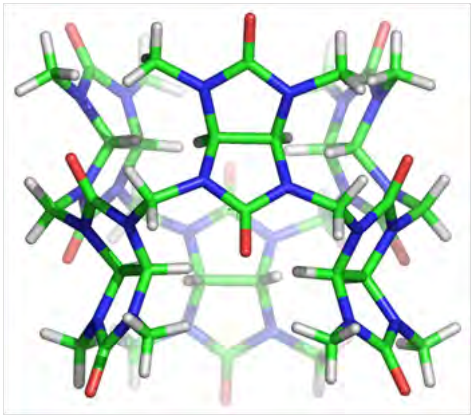


Fig. 6. The structure of bambus[6]uril.

functional (DFT) calculations to examine the dependence of phosphorus chemical shifts (δP) in a B-DNA on backbone torsion angles. Via a comprehensive analysis of the MD/DFT data, we have established a hierarchy of effects influencing δP. We have also obtained simple functions that describe the torsion angle dependences of δP and that can be used as alternatives to the so-called Karplus equations, typically used in NMR to determine torsion angles. Moreover, an elaborate averaging of the calculated chemical shifts provided chemical shift maps revealing the joint effects of α and ζ. The combined MD/DFT calculations thus enabled an unprecedented insight into the torsion angle-chemical shift relationships useful for nucleic acid structure determination (Fig. 7).

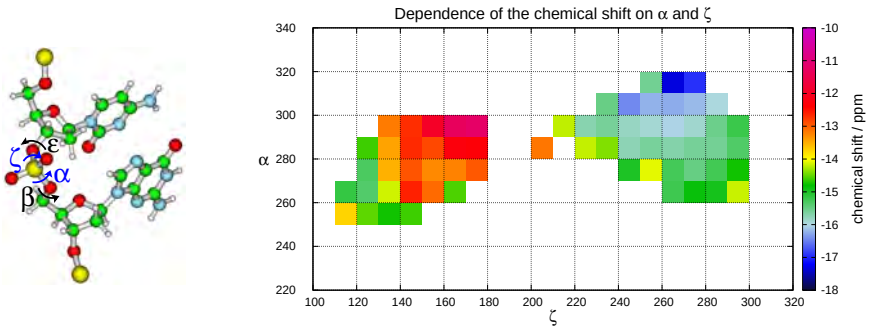


Fig. 7. Dependence of the ^{31}P chemical shift on torsion angles α and ζ in a nucleic acid backbone as obtained from combined molecular dynamics and density functional calculations.

Chemical shifts in 6-halopurines. A prototypical study of NMR chemical shifts in biologically relevant heteroaromatic compounds containing a heavy halogen atom is presented for two isomers of halogen-substituted purines (Fig. 8). A range of purine derivatives modified in position 6 of the basic purine skeleton exhibit a variety of biological activities. Experimental results are complemented by quantum-chemical calculations that provide understanding of the trends in the chemical shifts for the studied compounds and which show how different physical effects influence the NMR parameters. Chemical shifts for isolated molecules are calculated using DFT, the role of solvent effects is studied using PCM models, and relativistic corrections are calculated using the leading-order Breit-Pauli perturbation theory. Calculated values are compared with the experimental data and the effects of structure, solvent and relativity are discussed. Overall, we observe a good agreement of theory and experiment. We find out that relativistic effects cannot be neglected even in the chlorine species when aiming at high precision and a good agreement with the experimental data. Relativity plays a crucial role in the bromine and iodine species. Solvent effects are of smaller importance for ^{13}C shifts but are shown to be substantial for particular ^{15}N shifts. This work represents the first full application of the Breit-Pauli perturbation theory to an organic chemistry problem.

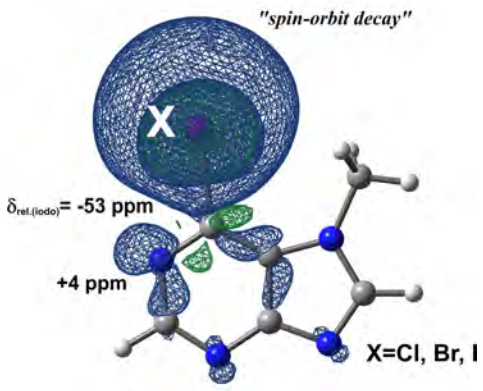


Fig. 8. Chemical shifts in 6-halopurines.

^{129}Xe as NMR probe. ^{129}Xe atom is an excellent NMR probe. Due to its large and polarizable electron cloud, the physico-chemical surroundings of the ^{129}Xe atom are sensitively reflected in the magnetic shielding at the ^{129}Xe nucleus. The size and chemical inertness allows xenon to enter different materials without reacting chemically. Thus the ^{129}Xe NMR spectroscopy provides a unique tool for studying different materials and their properties (orientational and translational order in liquid crystals, solvent dynamics in liquids, structure of porous solids,

polymers, and clathrates). Magnetic resonance imaging based on ^{129}Xe NMR has been recently employed in studies of biological materials using biosensors in which the ^{129}Xe atom is confined in a molecular cage. To our best knowledge, even the problem of ^{129}Xe chemical shift in benzene has not been tackled theoretically yet. The isotropic ^{129}Xe NMR chemical shift of atomic Xe dissolved in liquid benzene was simulated by combining classical molecular dynamics and quantum-chemical calculations (using nonrelativistic DFT as well as relativistic Breit-Pauli perturbation corrections) of ^{129}Xe chemical shielding. The relativistic effects explain about 8% of the total ^{129}Xe chemical shift. Present theoretical results are in a very good agreement with the experimental data. We believe that this approach can be extended to the practical simulations of the ^{129}Xe chemical shift in isotropic liquids, and more complex xenon guest-host systems.

Chemical shifts in complexes containing Pt(IV). Since the discovery of cis-platin in the 1960s, great efforts were spent on the preparation of various transition metal complexes with potential antitumor activity. The main goal of these investigations was to suppress the negative effects of cis-platin (e.g., neurotoxicity, nephrotoxicity, nausea). In our approach, we are using Pt in the oxidation state IV instead of Pt(II) for complexation. Complexes of Pt(IV) showed similar anticancer properties as those with Pt(II), but with significantly reduced side effects. We prepared new complexes of Pt(IV) with cytokinins (Fig. 9) and characterized them via NMR. We supplemented the structural conclusion from the NMR analysis by density-functional theory calculations of structural parameters and nuclear shielding using the Gaussian 03 and 09 packages. Geometries were optimized using the B3LYP/6-31G* level of theory, shielding calculations were performed using GIAO with the IGLO III basis set and effective core potential basis sets (ECP60MDF) for platinum respectively. We obtained very good agreement with experimental data, and solved the structures unequivocally in both solid state and solvent.

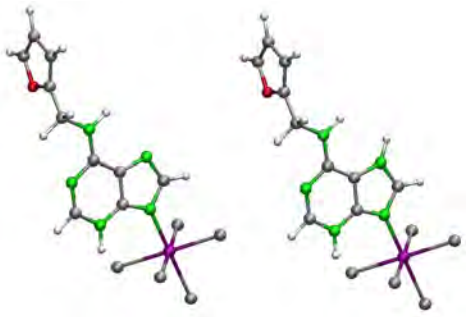


Fig. 9. Studied complexes containing Pt(IV).

2.4 Free Energy Calculations

Acetylcholinesterase. Currently, the main focus of our acetylcholinesterase research was on pK_a calculations for selected ionisable residues. The previous molecular dynamics study revealed unexpected conformational behaviour upon the change of the protonation state of glutamic acids in the active site. With increasing active site charge, the structure grew more flexible and at the maximal charge value, a radical opening of the omega loop (Cys69 - Cys96 in the pdb structure 1N5M) appeared. These results led us to the conclusion that there may be certain active site protonation states preferred, while others may be unfavourable. pK_a shift calculation, a specific thermodynamic integration set-up, is a way to compute pK_a by calculating relative free energy differences by the means of a thermodynamic cycle.

The pK_a calculations were extremely time demanding, since it was necessary to evaluate all possible one-proton transformations for the active site glutamic acid residues, namely Glu202, Glu334 and Glu450 (Fig. 10). As all three residues are close together, a strong two-way influence was expected and confirmed later when the calculations were finished. The resulting pK_a values show that acetylcholinesterase prefers at leaast one Glu in a protonated state. There is much uncertainty about the presence of the second one, but it is probable considering also the stabilising effect of decreasing active site charge on the overall conformational flexibility.

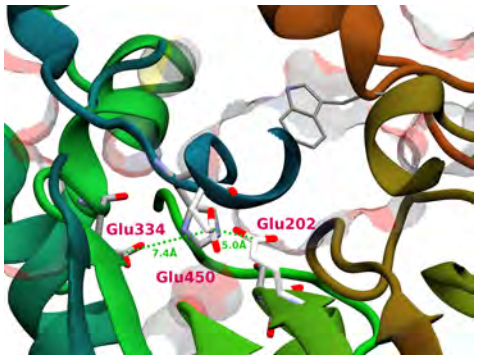


Fig. 10. Active site of acetylcholinesterase.

DNA/alkaloid interaction. Recently we have studied the interactions of alkaloids with fragments of DNA using NMR spectroscopy and molecular modeling methods. Our results showed the possibility of intercalation of the planar alkaloid core between base pairs of the double stranded DNA fragment (Fig. 11). The calculated interaction energies based on MD simulations have not reflected the energy of conformational changes of DNA during intercalation. The umbrella

sampling method was chosen for the free energy calculation of conformational changes of DNA fragments.

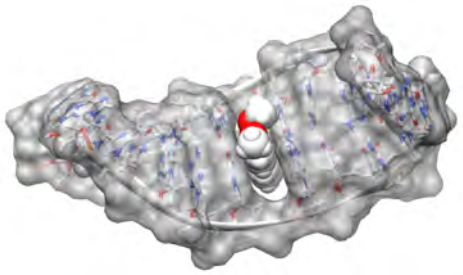


Fig. 11. Alkaloid intercalated into the DNA molecule fragment.

Reaction mechanism of the MutH enzyme. MutH is an integral part of the DNA repair mechanism, which is responsible for removing mismatches introduced in the DNA replication process in certain bacteria. The main role of MutH is to cleave the daughter chain of DNA containing the introduced mismatch. Our study is focused on the understanding of the reaction mechanism of MutH. Firstly, the MutH reactivity was studied employing semiempirical methods (Fig. 12). The free energy along several tested reaction coordinates was calculated using the Adaptive Biasing Method. Unfortunately, the obtained results were very poor and they did not correlate well with known exeprimental data. Thus we have stepped further, to utilizing ab initio dynamics via the Car-Parrinello approach.

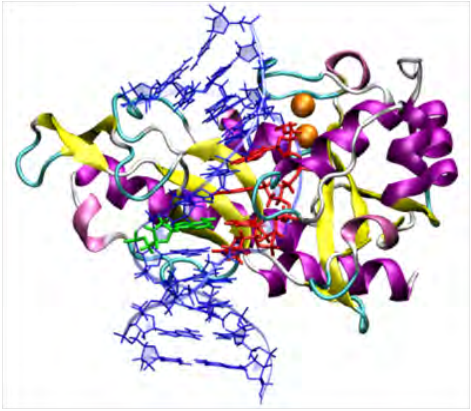


Fig. 12. The structure of the MutH/DNA complex.

2.5 Chemoinformatics and Bioinformatics

Partial atomic charge calculations. Knowledge of atomic charges is necessary for the prediction of molecular properties and chemical behavior, either in the context of Quantitative Structure-Property Relationship (QSPR) analyses, or within molecular simulations of various kinds. The Electronegativity Equalization Method (EEM) is an empirical technique aimed at the fast and fairly accurate calculation of atomic charges. Our aim is to take advantage of EEM in the context of biomolecular studies. We attempted to parameterize EEM for proteins, and for this purpose we used MetaCentrum resources. The main activity was charge calculation for protein fragments using the program Gaussian. We obtained a few sets of well performing parameters (Fig. 13).

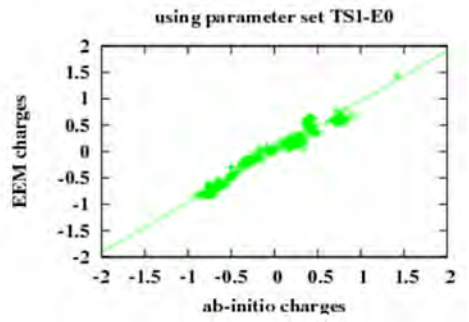


Fig. 13. Correlation between the EEM predicted charges and ab initio charges: R^2=0.95, RMSE=0.106e, D_{avg}=0.086e.

Predicting pK_a values of substituted phenols. The acid dissociation (ionisation) constant pK_a is one of the fundamental properties of organic molecules. We have evaluated different computational strategies and models to predict the pK_a values of substituted phenols using partial atomic charges. Partial atomic charges for 143 phenol molecules were calculated using 75 combinations comprising of five theory levels (MP2, HF, B3LYP, BLYP BP86), three basis sets (6-31G*, 6-311G, STO-3G) and three population analyses (MPA, NPA, Hirshfeld, ESP, Lowdin). The correlations between pK_a and various atomic charge descriptors were examined and the best descriptors were selected for preparing the QSPR models. The pK_as predicted by most of the models correlate strongly with experimental pK_a values. For example, the QSPR model employing HF/Mulliken/6-31G* charges exhibits a squared correlation coefficient of 0.968, a root mean square error of 0.423 and an average pK_a difference of 0.333. The best results were obtained for the MP2 and HF levels of theory. The most suitable basis set was found to be 6-31G*. The Mulliken, Natural and Lowdin population analyses provide accurate models for all tested theory levels and basis

sets. The results provided by the Hirshfeld population analysis were also acceptable, but the QSPR models based on ESP charges show only weak correlations.

Docking and virtual screening. Finding inhibitors of the lectin PA-IIL (Fig. 14) is very important, as inhibition of this lectin may be used in the treatment of specific bacterial infections. We have performed a high throughput virtual screening using about 800 000 ligands (potential inhibitors). The binding of every individual ligand was assesed by scoring functions provided by UCSF Dock and Autodock Vina.

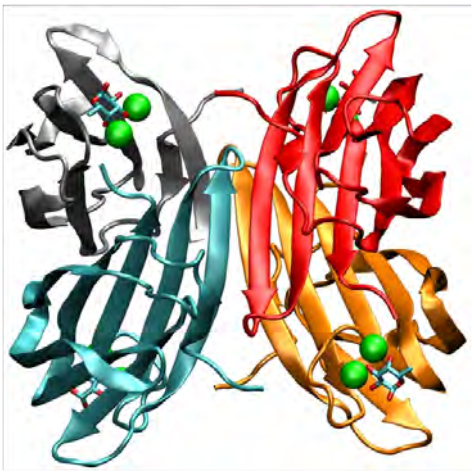


Fig. 14. Lectin PA-IIL, used in the virtual screening project.

3 Software

The computational chemistry calculations described in previous chapter were done using the following pieces of software. Gaussian, Turbomole, and Molpro are used in quantum chemical studies. The benefit of MetaCentrum lays in their possible parallel executions, which significantly decrease the time required for these very time consuming calculations. Molecular dynamics studies are solely performed employing the AMBER software package. Docking and related studies are performed using Dock, Autodock, and Autdock Vina. Besides these commonly used software packages, several applications have been developed at NCBR. They simplify job preparation, management and analysis of results.

3.1 Infinity

Infinity provides uniform and simple tools for an easy job submission and management in various computer environments such as computational clusters or grid environments. Infinity is targeted for users using command line interface, which provides a greater extent of flexibility than web based interfaces. This is especially important for fast and innovative scientific research requiring the usage of computational resources. Not only job management is important for an efficient utilization of computational resources. Thus Infinity provides the inovative Advanced Module System, which simplifies the selection of software according to the available architecture and parallel environments.

3.2 TRITON

TRITON is a graphical tool for modeling protein structures and their properties (Fig. 15). It offers the preparation of input files and visualization and analysis of output data. TRITON can be used for modeling protein mutants and, subsequently, their properties. It also facilitates the modeling of chemical reactions on semi-empirical levels with various tools assessing mutant activities. Another feature of TRITON is the interface to AutoDock, which simplifies the docking procedure and analysis of results. It offers the visualization of affinity maps and calculation of electrostatic potential interactions between the ligand and individual residues in the binding site.

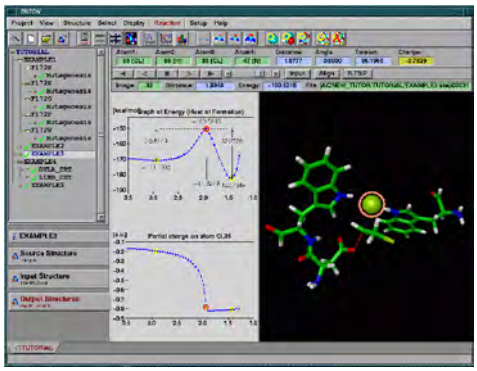


Fig. 15. TRITON screenshot.

3.3 PMFLib

PMFLib is a set of libraries and programs supporting the calculation of free energies by the means of potential of mean force methods in the framework of molecular dynamics simulations. Currently implemented methods are as follows:

Adaptive Biasing Force, Blue Moon, Metadynamics, and Umbrella Sampling. Due to sampling problems, free energy calculations are time consuming. Thus we implemented the Multiple Walkers extension to the Adaptive Biasing Force method and Metadynamics. This extension employs N independent molecular dynamic simulations, so-called walkers, that feel the estimated PMF accumulated from all walkers. The interaction among walkers and estimated PMF was achieved by simple client-server architecture and successfully tested in the heterogeneous MetaCentrum and VOCE grid environments.

3.4 CICADA

Knowledge of conformational space is essential in the assessment of the dynamical behaviour of biomolecules, which is important, for example, in drug design and folding studies. Unfortunately, systematic exploration of conformational space is impossible due to its high complexity. To circumvent this problem, we have developed the program CICADA, that tries to rationalize the search in such a space. The program CICADA is composed of two main components: a) control unit and b) explorers (Fig. 16). The control unit collects information about already searched conformational space. It optimizes the interconversion pathways found among conformers, and rationalizes the strategy of further space exploration. It also manages the explorers, which perform a simple search in a limited conformational subspace. Since individual explorers are independent of each other, they can be executed independently, which makes the basis of an efficient distributed calculation. The program was designed in such a way, that it can be run on a heterogeneous cluster such as the Metacentrum resources or other grids.

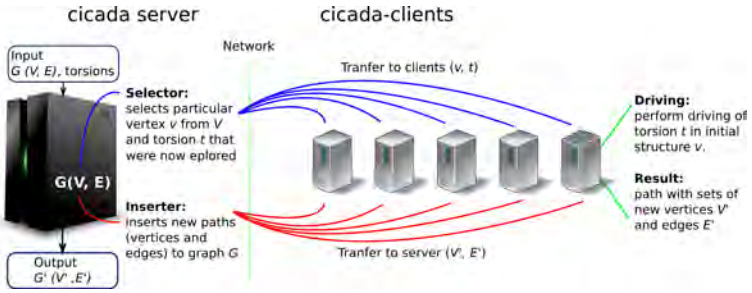


Fig. 16. CICADA program design

3.5 Other software

Several other programs and utilities have been developed, namely dynutil, cats, and qmutil. For example, the dynutil precycle utility, together with Infinity,

enables effective automatization of job resubmission. This allows utilization of MetaCentrum resources limited to short and normal queues for performing very long molecular dynamics simulations. In addition, tools for effective parallel analysis of molecular dynamics trajectories have been developed and used in the calculation of binding free energies using the MM/PBSA method.

4 Hardware

NCBR contributed to the common MetaCentrum infrastructure with about 230 CPUs in 2010. These resources were organized in two clusters: orca (18 nodes, 2x dual-core Opterons @2.6 GHz) and perian (10 nodes, 2x quad-core Intel Xeon @3 GHz; 10 nodes, 2x quad-core Intel Xeon @2.5 GHz). Temporarily unused resources were available to other MetaCentrum users via the queues short, normal, backfill, and preemptible. This setup allowed preferred access to these resources for NCBR members.

5 Conclusion

Research conducted at NCBR in the field of computational chemistry uses state-of-the art computational approaches, which requires up-to-date computational resources. There are diverse requirements for job executions, from a huge number of small single CPU jobs, to a smaller number of highly parallelized jobs. These different requirements are fulfilled within the MetaCentrum framework. Thus, we can conclude that the computational resources and services provided by MetaCentrum are helpful for conducting effective research in the field of computational chemistry at the National Centre for Biomolecular Research.

Acknowledgements. The access to the MetaCentrum supercomputing facilities provided under the research intent MSM6383917201 is highly appreciated. This work has been supported by the Ministry of Education of the Czech Republic (MSM0021622413, LC06030) and the Czech Science Foundation (203/09/0748, 301/09/H004, 305/09/0457). The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement no. 205872.

6 Selected publications containing results obtained with the use of MetaCentrum resources

6.1 Articles

1. Reblova, K., Strelcova, Z., Kulhanek, P., Besseova, I., Mathews, D., Van Nostrand, K., Yildirim, I., Turner, D., Sponer, J.: An RNA Molecular Switch: Intrinsic Flexibility of 23S rRNA Helices 40 and 68 5'-UAA/5'-GAN Internal Loops Studied by Molecular Dynamics Methods. Journal of Chemical Theory and Computation. 6, 910-929 (2010).

2. Kolman, V., Kulhanek, P., Sindelar, V.: Inclusion of Carboxyl Function Inside of Cucurbiturils and its Use in Molecular Switches. Chemistry-an Asian Journal. 5, 2386-2392.
3. Wiesner, J., Kriz, Z., Kuca, K., Jun, D., Koca, J.: Influence of the Acetylcholinesterase Active Site Protonation on Omega Loop and Active Site Dynamics. Journal of Biomolecular Structure & Dynamics. 28, 393-403 (2010).
4. Precechtelova, J., Novak, P., Munzarova, M., Kaupp, M., Sklenar, V.: Phosphorus Chemical Shifts in a Nucleic Acid Backbone from Combined Molecular Dynamics and Density Functional Calculations. Journal of the American Chemical Society. 132, 17139-17148 (2010).
5. Mishra, N., Kriz, Z., Wimmerova, M., Koca, J.: Recognition of selected monosaccharides by Pseudomonas aeruginosa Lectin II analyzed by molecular dynamics and free energy calculations. Carbohydrate Research. 345, 1432-1441 (2010).
6. Maier, L., Solomek, T., Pipiska, M., Kriz, Z., Necas, M., Marek, R.: Structural study of 8-azole derivatives of protoberberine alkaloids experimental and quantum chemical approach. Tetrahedron. 66, 9277-9285 (2010).
7. Standara, S., Malinakova, K., Marek, R., Marek, J., Hocek, M., Vaara, J., Straka, M.: Understanding the NMR chemical shifts for 6-halopurines: role of structure, solvent and relativistic effects. Physical Chemistry Chemical Physics. 12, 5126-5139 (2010).
8. Wiesner, J., Kriz, Z., Kuca, K., Jun, D., Koca, J.: Why acetylcholinesterase reactivators do not work in butyrylcholinesterase. Journal of Enzyme Inhibition and Medicinal Chemistry. 25, 318-322 (2010).

6.2 Bachelor, diploma, and doctoral theses

1. Slavík, J. Pocitacove modelovani glykolurilovych struktur. Bachelor Thesis, MU Brno (2010).
2. Benešová, B. Studium interakci substituovanych cyklickych nitronu s fragmenty DNA metodami molekuloveho modelovani. Bachelor Thesis, MU Brno (2010).
3. Ionescu, C.M. Fast methods of charge calculation. Diploma Thesis, MU Brno (2010).
4. Kulaš, A. Conformational behavior study of the BCLA lectin and its deletion mutants. Diploma Thesis, MU Brno (2010).
5. Mládek, A. Quantum-chemical studies of nucleic acids. Diploma Thesis, MU Brno (2010).
6. Adam, J. Structure-functional analysis and molecular modeling of interactions of bacterial lectins and saccharides. Ph.D. Thesis, MU Brno (2010).
7. Novák, P. Computational tools for the analysis of biomolecular data obtained by NMR and molecular dynamics simulations. Ph.D. Thesis, MU Brno (2010).
8. Precechtělová, J. Torsion Angle Dependence of ^{31}P NMR Parameters in a Nucleic Acid Backbone Studied by Density Functional Tudory. Ph.D. Thesis, MU Brno (2010).

DFT Study of the Oxametallacycle Formation on Platinum Clusters

Jakub Šebera[1,2,3] and Stanislav Záliš[4]

[1] Institute of Macromolecular Chemistry, AS CR, v. v. i., Heyrovský Sq. 2, 162 06 Prague 6, Czech Republic
[2] Institute of Physics, Czech Academy of Sciences, 182 21 Prague, Czech Republic
[3] Institute of Organic Chemistry and Biochemistry AS CR, v.v.i., Flemingovo náměstí 2, CZ, 16610 Prague, Czech Republic
[4] J. Heyrovský Institute of Physical Chemistry, AS CR, v.v.i., Dolejškova 3, 18223, Prague 8, Czech Republic
sebera@fzu.cz

1 Introduction

The control of the selectivity of the (electro)catalytic processes represents the main challenge in the field of the 21st century. The issue of selectivity is of particular importance in electrode processes involving small organic molecules featuring both single as well as double bonds. Interesting are oxidative processes connected with oxygen transfer, e.g., oxidation of aliphatic alcohols or insertion of oxygen in to double bond [1]. In the latter case there is a need to develop electrocatalytic materials for selective oxygen insertion. Possible materials are the platinum nanoparticles and clusters which are known to facilitate reactions mechanism unknown on conventional catalysts [2]. In the case of the ethylene oxidation experimental studies indicated the preferential carbon dioxide formation on platinum clusters [3].

2 Computational Details

The electronic structure of the model systems presumed as the intermediate intermolecular interactions in the process of ethylene oxidation, was studied for complexes of oxametallacycle adsorbed on platinum clusters. The present study is based on the cluster approach where platinum catalytic center is described as finite cluster. Reactants and transition states are considered as one molecular system in the electronic ground or activated state. Energies of all possible products are calculated on the basis of separated systems between clusters and small organic products. The electronic structure of these molecular systems was studied at the non-empirical all-electron level using density functional theory (DFT) within basis set with polarisation functions 6-31G(d). Furthermore, in the case of the platinum atoms, the LANL2DZ basis set was used. At the first step, the platinum clusters were optimized without any symmetry restriction. The most applied DFT functional in quantum chemistry B3LYP was used in our study.

B3LYP is recommended for calculations of thermochemistry data for reaction energy ΔG_R and activation energy $\Delta G^{\ddagger}$. Gibbs free energy G will by a key quantity for thermochemistry characteristic of different channels of investigated reaction. All quantum chemistry calculations were performed in Gaussian 03.

3 Results and Discussion

DFT method was used for the modeling of platinum cluster Pt_{21} and its interaction with organic molecules. The electronic configuration of platinum atom is: Pt [Xe] $4f^{14}$ $5d^9$ $6s^1$. Optimized state geometry was calculated for the lowest spin multiplicity of platinum cluster Pt_{21} (^{7}A). Two lowest lying spin multiplicity states beyond ground state were (^{9}A) and (^{15}A), the energy difference between ground state and the two lowest lying states were 0.1 eV (vertical energy). The averaged interatomic distance in pyramidal Pt_{21} (^{7}A) was 2.79 Å (2.58–3.29 Å) which is in agreement with experimental value of bulk distance of Pt(111) 2.775 Å.

Atomic oxygen is bonded to the clusters in two ways (plane interaction—oxygen is bonded to the three atoms, edge interaction—oxygen is bonded to the two atoms). For platinum cluster, the averaged distances of Pt-O are: plane interaction – 2.024 Å (nonuplet state) and edge interaction – 1.953 Å (septet and nonuplet state).

The mechanisms of the O and C_2H_4 adsorption at the platinum cluster, the formation of Pt_{21}-Et-O intermediates were studied. The interaction of the adsorbed oxygen with ethylene leads to the stable surface oxametallacycle intermediates which are formed analogously to ones described in the case of ethylene oxide reaction at Ag(111) surface, see Figure 1. The interaction with the surface depends on multiplicity and the reaction site (the plane steps or edges). Oxametallacycle intermediate leads to the formation of oxirane or acetaldehyde products. The DFT calculations of $\Delta G^{\ddagger}$ show that acetaldehyde product has small energetic barrier than formation of oxirane product. The acetaldehyde product is further oxidized to the CO_2. The results of the DFT calculations are in accordance with experimental results of electrocatalytic oxidation of ethylene on platinum. In the case of Pt electrodes the reaction produces solely the carbon dioxide.

4 Conclusions

DFT calculations of the oxidation of ethylene on platinum nanostructures lead to a formation of acetaldehyde as a major product. In the experimental studies, the ethylene oxidation proceeds carbon dioxide. It is known that further oxidation of acetaldehyde forms carbon dioxide [4]. DFT results support experimental findings [3] and indicate possible reaction mechanisms of catalytic reactions. DFT calculations points to the different reactivity on individual types of surfaces.

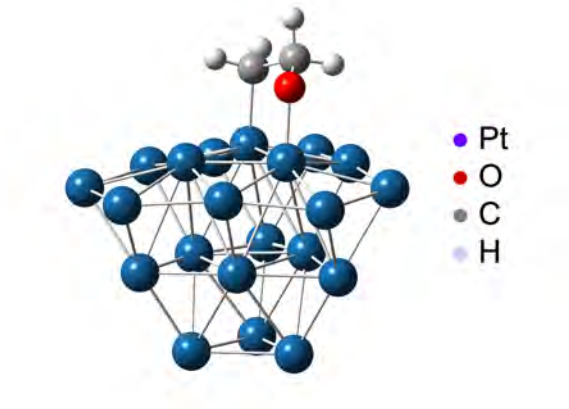


Fig. 1. DFT optimized structure of oxametallacycle intermediate interacting with Pt_{21} cluster.

Acknowledgements

This work was supported by European COST action and the Academy of Sciences of the Czech Republic (grant KAN100400702).

The access to the MetaCentrum computing facilities provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" LM2010005 funded by the Ministry of Education, Youth, and Sports of the Czech Republic is highly acknowledged.

References

1. K. Otsuka, I. Yamanaka, *Catal. Today*, 41, 311-325 (1998)
2. T. Ishida, M. Haruta, Angew. *Chem. Int. Ed.*, 46, 7154–7156 (2007)
3. A. Toghan, L. M. Rösken, R. Imbihl, *ChemPhysChem*, 11, 1452-1459 (2010)
4. S. Linic, M. A. Barteau, *J. Am. Chem. Soc.*, 125, 4034-4035 (2002)

Used Programs and Applications

Gaussian 03 and Gaussian 09

List of Publications Dedicated to MetaCentrum

1. A. Cannizzo, A. M. Blanco-Rodriguez, A. El Nahhas, J. Šebera, S. Záliš, A. Vlček, Jr., and M. Chergui: "Femtosecond Fluorescence and Intersystem

Crossing in Rhenium(I) Carbonyl-Bipyridine Complexes", *J. Am. Chem. Soc.*, **130**, 8967-8974, (2008)
2. J. Šebera, S. Nešpůrek, I. Kratochvílová, S. Záliš, G. Chaidogiannos and N. Glezos: "Charge carrier mobility in sulphonated and non-sulphonated Ni phthalocyanines: experiment and quantum chemical calculations", *Eur. Phys. J. B*, **72**, 385-395, (2009)
3. I. Kratochvílová, T. Todorciuc, K. Král, H. Němec, M. Bunček, J. Šebera, S. Záliš, Z. Vokáčová, V. Sychrovský, L. Bednárová, P. Mojzeš, B. Schneider: "Charge Transport in DNA Oligonucleotides with Various Base-Pairing Patterns", *J. of Phys. Chem. B*, **114**, 5196-5205, (2010)
4. J. Šebera, H. Hoffmannová, P. Krtil, Z. Samec and S. Záliš: "Electrochemical and density functional studies of the catalytic ethylene oxidation on nanostructured Au electrodes", *Catal. Today*, **158**, 29-34, (2010)

List of Conference Presentations Prepared Using MetaCentrum Resources

1. ESF DYNA Conference on Electronic Excited States in Condensed Phases: S. Záliš, J. Šebera, A. Vlček, *Modeling of singlet and triplet charge-transfer excited states in rhenium (I) and ruthenium (II) carbonyl-bipyridine complexes*, Book of Abstracts. Richmond : Organic Reactions Catalysis Society, 18.07.2008-22.07.2008, Italy, 86, (2008)
2. International Conference on Coordination Chemistry: S. Záliš, J. Šebera, A. Vlček, *Quantum chemical modeling of lowest lying excited states in transition metal carbonyl-dimine complexes*, Book of Abstract. Santa Fe, S. 71-71, /38./. 20.07. 2008 - 25.07. 2008, Jerusalem (2008)
3. Modeling and Design of Molecular Materials: J. Šebera, Y. Matsuda, Y. Tanaka, V. Sychrovský, *Computational study and analysis of infrared spectra of uridinemonophosphate*, Book of Abstract, Wroclaw, Poland, July 4-8, S 115 (2010)
4. Modeling and Design of Molecular Materials: J. Šebera, S. Záliš, *First-principles analysis of formation and reactivity of oxametallacycle intermediate on Au and Pt clusters*, Book of Abstract, Wroclaw, Poland, July 4-8, S 116 (2010)
5. Central European Symposium on Theoretical Chemistry: J. Šebera, S. Záliš, P. Kubát, Pavel, K. Lang, T. Polívka, *TD-DFT investigation of S_1 and S_2 singlet states of TMPyP(n) and complexes of TMPyP4 with sulfonated calix[m]arenes*, Book of Abstracts, Bratislava : Comenius University, 12. 09. 2010–15. 09. 2010, Nový Smokovec, S121-122 (2010)

Virtual Organization for the Pierre Auger Observatory

Jiří Chudoba[1,2]

[1] Institute of Physics
Academy of Sciences of the Czech Republic
[2] CESNET, z.s.p.o.

Abstract. The Pierre Auger Observatory (PAO) studies the highest energy cosmic air showers. The EGI grid is used for computationally demanding simulations of showers. CESNET operates the core servers for the Virtual Organization *auger* and manages membership in this VO. Output files are stored on geographically distributed grid Storage Elements (SE), we maintain the central database with the file location.

1 Cosmic Rays and Their Detection

Cosmic rays were discovered almost 100 years ago by Victor Hess. His measurements during balloon flights proved that intensity of the ionizing radiation increases with altitude. Later experiments in 1938 conducted by Pierre Auger in the Alps revealed that particles may arrive to distant detectors at the Earth surface in almost the same time. This led to a conclusion that detected particles originate from the same source particle which creates a shower of secondary particles after its first interaction in high altitudes of atmosphere. Pierre Auger estimated the energy of the primary particle up to 10^{15} eV from the number and surface distribution of secondary particles. The energy spectrum of cosmic rays falls steeply with energy. While we measure every second per 1 square meter 1 particle with energy 10^{11} eV, we must wait the whole year to detect a particle with energy 10^{16} eV on the same area and there is less than 1 particle per km^2 per year with energy above 10^{19} eV. In 1962 experiment Volcano Range detected a giant shower with estimated energy of primary particle 3×10^{20} eV. These energies are much higher than energies reachable by the most energetic accelerators built by humans. The LHC design energy is 1.4×10^{13} eV in the centre of mass energy system, while the center of mass energy of a 10^{20} eV proton interacting with a target nucleon in the earth's atmosphere corresponds to 4.5×10^{14} eV.

The sources and accelerator mechanism for such high energies remain unknown. It was shown shortly after the discovery of cosmic microwave background radiation, that very high energy particles loose energy due to the interaction with this radiation and we expect sharp fall in the energy spectra above so called GZK limit 5×10^{19} eV. Results of the experiment AGASA, which measured more than 10 events with energy above 10^{20} eV, contradict the existence of the GZK limit. Observations of the experiment HiRes were in agreement with the GZK limit.

Fig. 1. Surface detector array and view directions of fluorescence detectors.

2 Pierre Auger Observatory

Experiments AGASA and HiRes used two different techniques for cosmic ray shower detection. The AGASA deployed 111 scintillation detectors positioned in 1 km distance covering the total area 100 km^2. The HiRes used 2 fluorescence detectors to measure light emitted by secondary particles in the atmosphere. This technique requires dark clear nights for measurements while scintillators on the ground can measure continuously.

The Pierre Auger Observatory combined these two approaches. It deploys over 1,600 surface detectors separated by 1.5 km covering the total 3,000 km^2. Surface detectors (SD) are water tanks with 12,000 l of water monitored by 3 photomultipliers. Some particles traversing a SD emit Cherenkov light which is detected by photomultipliers and the signal is sent for further processing. Such signals from several SDs during a short time window signify a shower caused by primary particle with high energy. Fluorescence detectors (FD) are located in 4 buildings on the border of the SD array overlooking the atmosphere above SDs. The total number 24 FDs from original design was recently increased by addition of more FDs pointing to higher altitudes. The details of the detector parameters can be found in [1] and [2].

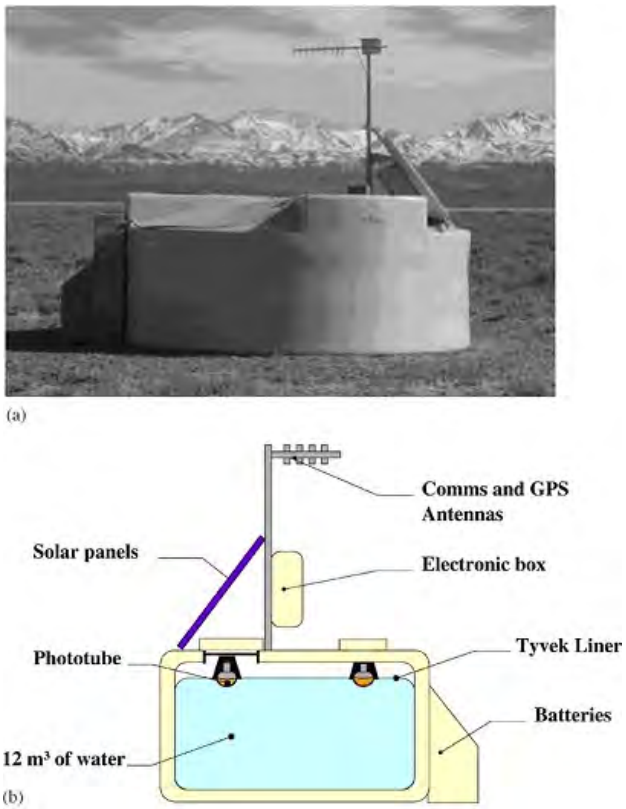


Fig. 2. (a) A photograph of an EA water tank; (b) schematic view of an EA tank (picture taken from [AugerDet]).

3 Monte Carlo Simulations

Energy and arrival direction of primary particles are obtained from several measured quantities. The energy reconstructed from the surface detectors is calibrated by fluorescence detectors to get smaller model dependence. Several models are used for the hadronic and electromagnetic part of the shower. Comparison of detailed simulations with observed events is essential for determination of chemical composition of primary particles. Several models are implemented in the CORSIKA simulation program [3]. The computing time needed for simulation steeply grows with energies and for energies above 10^{20} eV typically takes tens of hours on modern processors even with certain simplifications in the shower description (shower thinning). Availability of shower libraries with different energies, arrival directions, chemical composition and hadronic models was limited by computing capacity of local computing facilities in collaborators institutes.

4 Virtual Organization *auger*

The computational requirements of Monte Carlo simulations lead us to the idea of using the distributed resources via grid. The shower simulation is a trivially parallel task; each job can process one or more showers. Results are stored in different files, one shower per file. These requirements were similar to the requirements of the particle physics experiments which prototyped distributed computing on the now gLite based grid. We settled the Virtual Organization (VO) *auger* in 2006 within the project EGEE. The central servers for this VO are still operated by CESNET and we manage membership in this VO. The CESNET team developed custom registration portal for new VO members. Each member is registered not only in the standards VOMS server, but also in the PERUN database, from where we create user account on the User Interface (UI) server. The UI can be used by all users to submit jobs and retrieve their output, although due to scalability issues we recommend to most active users to use their own hardware.

We also developed a framework for a job submission necessary for a bulk production on the grid. This task was later passed on the Auger team from Granada University. We continue to do the user support task and central data management tasks.

5 Grid Usage by the VO *auger*

Grid based on the gLite middleware is now supported by several organizations. The European Grid Initiative (EGI) in Europe, Open Science Grid in USA and Gisela in Latin America are the most relevant for the VO *auger*. CESNET as a partner of the EGI project provides not only central services, but also computational server for the VO *auger*. 15 different computing sites were used by the VO *auger* in 2010. The CESNET gLite cluster contributed by almost 3% to the total used time. The VO *auger* was one of the most active VOs with steady increase of the resources.

The output files are stored on the distributed grid Storage Elements. The distribution of output files depends on the availability of SEs, their free space and limits for the VO *auger*. All files are registered in the central catalogue LFC operated by CESNET. Although files are stored on relatively reliable disk arrays protected by RAID5 or RAID6 configuration, it happens seldom that a set of files is lost or corrupted. In such case we delete files from the central catalogue and inform the production team that some jobs should be rerun. Another data management operation is due to changes in sites. Some sites declare their SEs as obsolete and leave limited time to users to transfer their files to different locations. Currently there are no tools directly in the middleware and a set of scripts is used for file replication and catalogue updates.

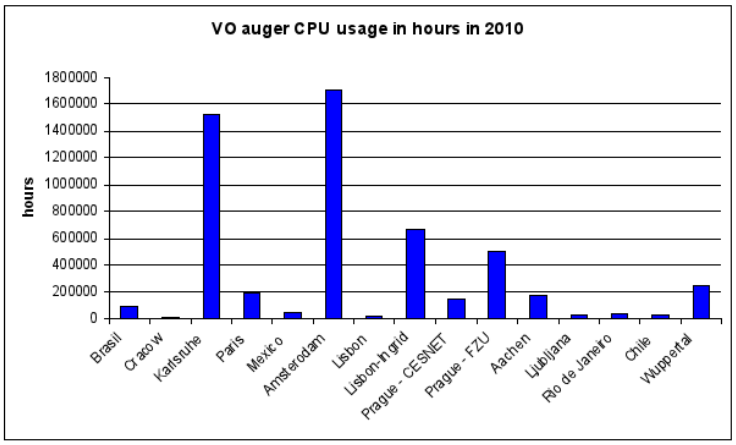


Fig. 3. VO *auger* CPU usage in hours in 2010.

6 Conclusions

The gLite based grid s heavily used by the Pierre Auger Observatory for Monte Carlo simulations. CESNET in cooperation with Institute of Physics of the Academy of Sciences of the Czech Republic founded the Virtual Organization *auger* and provides central services for the its operation. CESNET also contributes to the available computing and storage capacity.

References

[1] Properties and performance of the prototype instrument for the Pierre Auger Observatory, J. Abraham et al. [Pierre Auger Collaboration], *Nuclear Instruments and Methods*, **A523** (2004), 50

[2] The Fluorescence Detector of the Pierre Auger Observatory, J. Abraham et al. [Pierre Auger Collaboration], *Nucl. Instr. and Methods in Physics Research*, **A620** (2010) 227

[3] CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers, D. Heck et al., Report **FZKA 6019** (1998), Forschungszentrum Karlsruhe; `www-ik.fzk.de/corsika/physics_description/corsika_phys.pdf`

The Pierre Auger Collaboration publications in 2010

Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter, The Pierre Auger Collaboration, Astroparticle Physics 34 (2010) 314–326

The fluorescence detector of the Pierre Auger Observatory, The Pierre Auger Collaboration, NIM A 620 (2010) 227–251

The northern site of the Pierre Auger Observatory, The Pierre Auger Collaboration, New Journal of Physics 12 (2010) 035001

Measurement of the Depth of Maximum of Extensive Air Showers above 10^{18} eV, The Pierre Auger Collaboration, Physical Review Letters, 104, 091101 (2010)

Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory, The Pierre Auger Collaboration, Physics Letters B 685 (2010) 239–246

A study of the effect of molecular and aerosol conditions in the atmosphere on air fluorescence measurements at the Pierre Auger Observatory, The Pierre Auger Collaboration, Astroparticle Physics 33 (2010) 108–129

Trigger and aperture of the surface detector array of the Pierre Auger Observatory, The Pierre Auger Collaboration, Nuclear Instruments and Methods in Physics Research A613 (2010) 29–39

Author Index

Ivana Křenková, Petr Hanousek, and Luděk Matyska (Eds.)

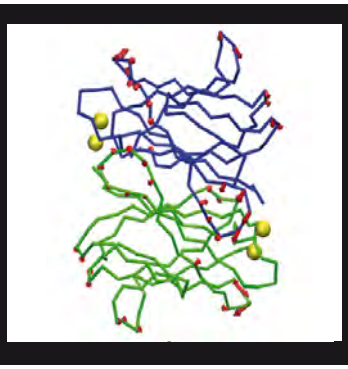
MetaCentrum Yearbook 2010

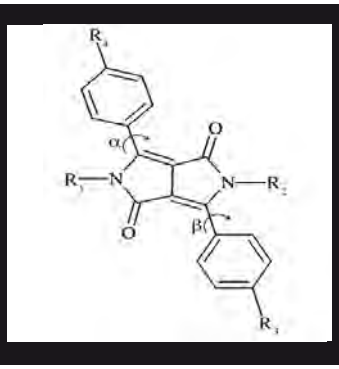
Publisher: CESNET, z. s. p. o., Zikova 4, 160 00 Prague, Czech Republic
Printing: PBtisk, s.r.o., Dělostřelecká 344, Příbram, Czech Republic
Edition: first
Year of publication: 2011

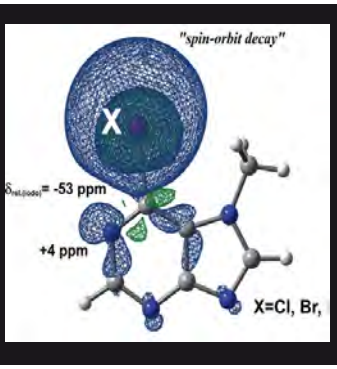
Typesetting, data conversion and design by David Antoš

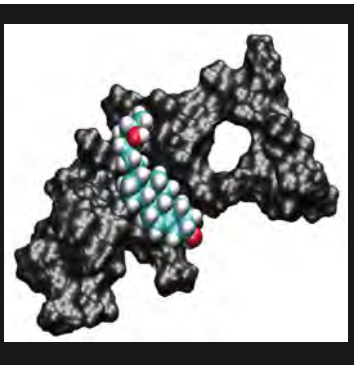
Cover design: Ivana Křenková

ISBN 978-80-904689-3-1









MetaCentrum Yearbook 2010
ISBN 978-80-904689-3-1

http://metavo.metacentrum.cz
http://metacentrum.cz

Printed by PBtisk, s.r.o., Příbram, Czech Republic
Not for sale

ISBN 978-80-904689-3-1
9 788090 468931