

# **National Biomedical Computation Resource (NBCR): Developing End-to-End Cyberinfrastructure for Multiscale Modeling in Biomedical Research**

Wilfred W. Li <sup>&1</sup>, Nathan Baker <sup>¥3</sup>, Kim Baldridge <sup>&1</sup>, J. Andrew McCammon <sup>@1</sup>, Mark H. Ellisman <sup>~%1</sup>, Amarnath Gupta <sup>&1</sup>, Michael Holst <sup>#1</sup>, Andrew D. McCulloch <sup>\$1</sup>, Anushka Michailova <sup>\$1</sup>, Phil Papadopoulos <sup>&1</sup>, Art Olson <sup>~2</sup>, Michel Sanner <sup>~2</sup>, Peter W. Arzberger <sup>\*%1</sup>

California Institute for Telecommunications and Information Technology (Calit2)\*, Center for Research In Biological Systems (CRBS)<sup>%</sup>, Departments of Biochemistry & Molecular Biophysics<sup>¥</sup>, Bioengineering<sup>\$</sup>, Chemistry<sup>@</sup>, Mathematics<sup>#</sup>, Molecular Biology<sup>^</sup>, Neuroscience<sup>~</sup>, San Diego Supercomputer Center<sup>&</sup>, University of California, San Diego (UCSD)<sup>1</sup>, The Scripps Research Institute (TSRI)<sup>2</sup>, Washington University in Saint Louis (WUSTL)<sup>3</sup>

**Abstract** – Begun in 1994, the mission of the National Biomedical Computation Resource (NBCR) is to conduct, catalyze and enable multiscale biomedical research by harnessing advanced computation and data cyberinfrastructure through multidiscipline and multi-institutional integrative research and development activities. Here we report the more recent research and technology advances in building cyberinfrastructure for multiscale modeling activities.

The development of the cyberinfrastructure is driven by multiscale modeling applications, which focuses on scientific research ranging in biological scale from the subatomic, to molecular, cellular, tissue to organ level. Application examples include quantum mechanics modeling with GAMESS, calculation of protein electrostatic potentials with APBS and the finite element toolkit FEtk, protein-ligand docking studies with AutoDock, cardiac systems biology and physiology modeling with Continuity, and molecular visualizations using PMV and visual workflow programming in Vision. Real use cases are used to demonstrate how these multiscale applications may be made available transparently on the grid to researchers in the biomedicine and translational research arena, through integrative projects ranging from the understanding of detailed mechanism of action of HIV protease and integrase, to neuromuscular junction research in myopathy, to heart arrhythmia and failure, and to emerging public health threats, as well as through collaborative projects with other research teams across the world.

The adoption of service oriented architecture enables the development of highly reusable software components, and efficiently leverages the international grid development activities. We describe an end to end prototype environment, exemplified by the adoption of key components by the Telescience project, that allows existing applications to run transparently on the grid, taking advantage of open source software that provides a portal interface using GridSphere, with transparent GSI authentication using GAMA, a web service wrapper using Opal, a metascheduler using CSF4, a virtual filesystem using Gfarm, and a grid-enabled cluster environment using Rocks. Solutions to complex problems may be developed using workflow tools that coordinate different interoperable services. In addition, we also describe the development of ontology and semantic mediation tools such as Pathsys and OntoQuest for data integration and interoperability which may be efficiently coupled with the application services provided to the biomedical community. NBCR Website URL: <http://nbcr.net>.

\*Principal Investigator, email: [parzberg@ucsd.edu](mailto:parzberg@ucsd.edu), phone: +01 (858) 822-1079.

## **1. Introduction**

With the increasing availability of genome sequencing data, it is becoming apparent that knowing the parts is only a prerequisite to understand the big picture. While genomics and proteomics effort are producing data at ever increasing rate, the data are more of a descriptive nature rather than functional integration and interactions of the parts [1]. Traditionally most research and modeling activities have focused on a particular system level such as proteins, cells, tissues, organs, organ systems, up to the level of populations. Multiscale modeling, across the length scale from nanometers for molecules to meters for human bodies, as well as across the time scale from nano-seconds for molecular interactions to the length of human life, is crucial to the development of simulation systems for better understanding of human physiology and predictive capabilities for disease prevention and treatment [2].

Multiscale modeling studies derive mathematical models of structure-function relations at one scale and link to the level below through appropriate parameters. These models need to be based on widely adopted modeling standards, with necessary software tools for developing, visualizing and linking the models [3]. Multiscale modeling requires constant cross validation and feedback from experiments and models. Often the experiments provide the data for development and validation of the models, and the models could in turn provide predictions of behavior or require additional experiments which may lead to new discoveries [1]. Models may impose *a priori* physical constraints and represent complex processes, and provide quantitative predictions that may be verified experimentally [4]. Systems modeling are data-limited when mechanistic models are to be built, because experiments may be slow and difficult to validate. On the other hand, models across scales are compute-limited due to the “tyranny of scale”. Molecular dynamics simulations are often limited to a scale 5 to 6 orders of magnitude smaller than the time necessary for a real event to complete. In terms of an extreme case of computational challenge, the panel on Simulation Based Engineering Science (SBES) [5] noted that in the turbulence-flow modeling, the “tyranny of scale” prevents a solution for many generations, even with Moore’s law holding true.

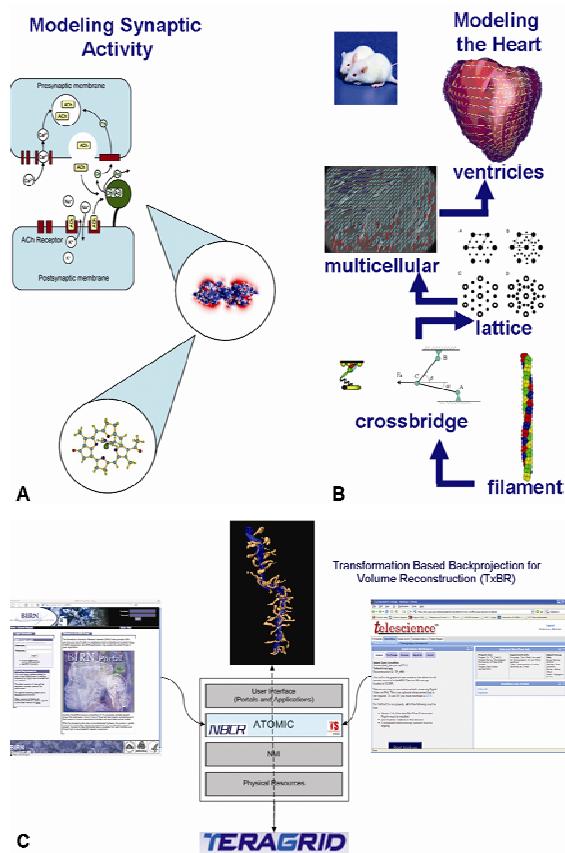


Figure 1. Multiscale modeling of A) the neuromuscular junction, B) physiology of the heart, C) electron microscopy of cellular structures are the main scientific drivers for NBCR cyberinfrastructure development efforts in support of translational research

have been focusing enabling a set of pathfinder examples across scales, through team work with world class scientists who are members of NBCR, and through collaborative projects with members of the broader biomedical and MSM communities.

## 2. Scientific Drivers and Tools

Understanding the workings of cells, tissues, organs, or entire organisms require researchers to pull together information from multiple physical scales and across multiple temporal scales. We highlight activities within our

In recognition of the importance and the severe obstacles to experiments that are cross-scale in space, time and state, there have been a number of workshops held and panel recommendations made. The multiscale modeling consortium or Interagency Modeling and Analysis Group (IMAG) [6], with the participation of a number of federal agencies including the National Institutes of Health (NIH), National Science Foundation (NSF), National Aviation and Space Agency (NASA), Depart of Energy, (DOE), Department of Defense (DoD) and the United States Department of Agriculture (USDA), aims to promote the development and exchange of tools, models, data and standards for the MultiScale Modeling (MSM) community. The NSF blue ribbon panel [5] recognizes that SBES applied to the multiscale study of biological systems and clinical medicine, or simulation based medicine, may bring us closer to the realization of P4 medicine (predictive and preventative, personalized, and participatory). The June 2005 PITAC report on computational science [7] and the 2005 National Research Council report [8], both specifically recommended increased and sustained support for infrastructure development to meet the computational challenges ahead.

The National Biomedical Computation Resource (NBCR, [9]), a national center supported by the National Center for Research Resources (NCRR, [10]), has a mission to “conduct, catalyze and enable multiscale biomedical research” by harnessing advanced computation and data cyberinfrastructure through multidiscipline and multi-institutional integrative research and development activities (Figure 1). NBCR has co-hosted several workshops and conferences that engage researchers and students from the MSM community [11] [12] [13]. It has been our experience that tools developed with one application in mind tend to be narrow and less flexible, let alone interoperable with other software. Therefore, we

collective and collaborative experiences at different or cross scales. These scientific examples are from long running complex multiscale research areas, which drive the development of our technology integration and infrastructure development efforts.

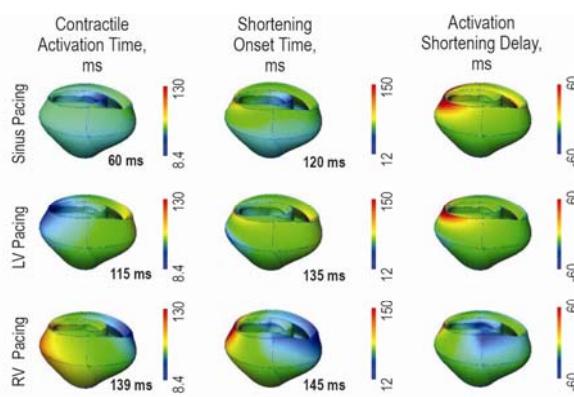


Figure 2. Effects of left and right ventricular pacing compared with normal sinus rhythm on the temporal sequences of electrical activation (left column) and mechanical shortening (middle column) in a three-dimensional model of the canine heart. Activation-shortening delays (right column) are heterogeneous, even during normal sinus rhythm. Simulations rendered using Continuity developed by NBCR.

## 2.1. Continuity 6

The need for integrative analysis in cardiac physiology and pathophysiology is readily appreciated. Common heart diseases are multi-factorial, multi-genic, and linked to other systemic disorders such as diabetes, hypertension, or thyroid disease. The coupling between the L-type Calcium channels (LCCs, also known as dihydropyridine receptors, or DHPRs) and ryanodine receptors (RyRs) are important in the excitation-contraction (E-C) of cardiac myocytes. The influx of calcium releases the calcium store in the Sarcoplasmic reticulum (SR), a phenomenon known as the calcium induced calcium release (CICR). In fact, the latest ionic models of cardiac myocytes include more than 20 ionic fluxes and 40 ordinary differential equations [1]. Computational methods and ionic models for cardiac electromechanics at different scales have also been developed and are available in the software package Continuity. Figure 2 shows an example of how Continuity is used to help develop dual pacemaker systems that are helping to save people's lives today [14]. Continuity is used by a number of researchers in the field of cardiac biomechanics, and receives regular acknowledgment in peer-reviewed publications [15,16]. Continuity 6 is continuously being improved to support larger scale simulations, e.g., using the MYMPI package [17], a standards-driven MPI libraries for Python, developed by NBCR to improve the parallel computation efficiency.

In ventricular myocytes, the dyadic cleft is a periplasmic space that spans about 10 nm between the voltage-gated LCCs/DHPRs on the transverse tubule (TT) membrane, and RyRs on the SR. Within the dyadic cleft, the small reaction volume and the exceedingly low number of reactant molecules means that the reaction system is better described by stochastic behavior, rather than continuous, deterministic reaction-diffusion "partial differential equations" [18,19]. This system is one of the focal points of NBCR research at the molecular scale, and for cross scale integrations, with the development of highly realistic 3D models based on electron tomography from NCMIR (The National Center for Microscopy and Imaging Research).

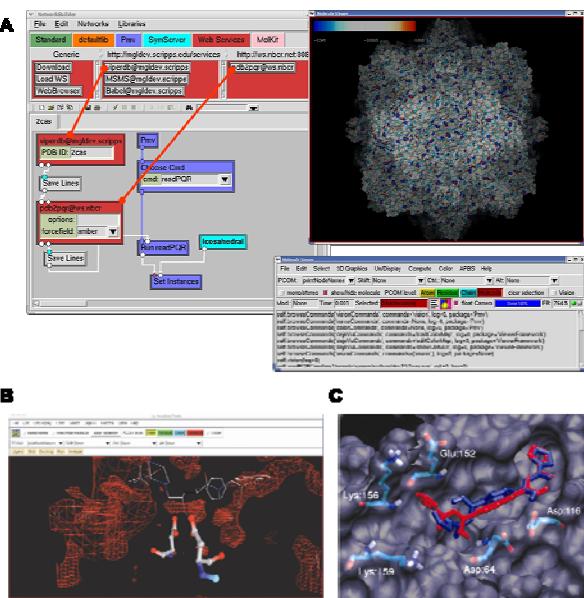


Figure 3. A) PMV is used to visualize the viral capsid proteins using the web service toolkit Opal based database access to the Viper Database. B) AutoDock Tools is used to visualize the ligand-protein interactions inside the HIV protease active site. C) Relaxed Complex Method and AutoDock are used to develop novel HIV inhibitors that have led to new drugs designed by Merck that are now in clinical trials.

## 2.2. PMV, Vision, AutoDock Tools & AutoDock

The Python Molecular Viewer (PMV) is a component based software package (Figure 3) written in Python and contains an accompanying visual programming tool called Vision [20]. PMV is among the first molecular visualization software package that takes advantage of grid services, using the web service toolkit Opal, developed by NBCR, to access remote databases and computational resources (Figure 3A). Many key packages from the

PMV/Vision framework have been reused in Continuity 6, the multiscale modeling platform for cardiac electrophysiology and biomechanics.

In addition, AutoDock Tools (ADT) [21] has been developed as a module inside PMV for the popular molecular docking software package AutoDock. AutoDock is a world famous software package, and has been used in developing inhibitors for many important diseases [22,23]. The FightAids@Home project has been using AutoDock to screen for HIV inhibitors, and is now running on the World Community Grid, an IBM philanthropic activity [24]. ADT greatly simplifies the preparation and post analysis procedures of AutoDock (Figure 3B).

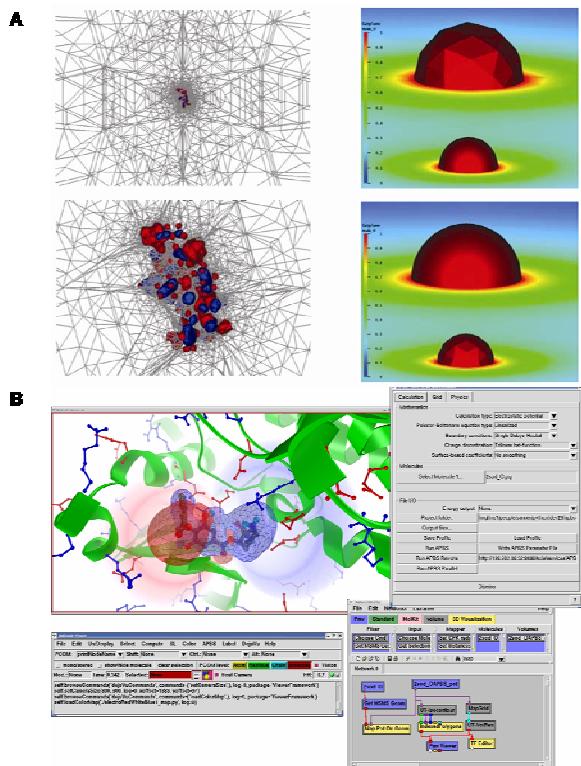


Figure 4. A) Application examples of FETk for solving the electrostatic potentials of biomolecules or rendering of isosurfaces of black holes in astrophysics problems. B) The electrostatic binding energy of PKA and balanol is visualized using PMV after remote distributed APBS calculations using NBCR strongly typed web service.

potential around a biomolecule, computed adaptively using FEtk, and a closeup of the adapted part of the simplex mesh. Right (Top/Bottom): Isosurfaces projected onto a cutting plane through two black holes in an astrophysics problem. FEtk was used to compute the initial bending of space and time around two massive black holes; this involved adaptively solving a coupled nonlinear elliptic PDE.

The primary FETk ANSI-C software libraries include MALOC (Minimal Abstraction Layer for Object-oriented C) for portability, SG (Socket Graphics) for networked OpenGL-based visualization, and MC (Manifold Code) for adaptively solving coupled systems of nonlinear PDE on two- and three-manifold domains. A 2D MATLAB-based prototyping tool called MCLite is also available for fast development of MC-based software. A related package, PMG (Parallel Algebraic Multigrid) is designed to numerically solve scalar nonlinear PDE problems such as the Poisson-Boltzmann Equation (PBE) on a class of regular domains using algorithms which have optimal or near-optimal space and time complexity.

APBS is a software package for the numerical solution of the Poisson-Boltzmann equation (PBE), one of the most popular continuum models for describing electrostatic interactions between molecular solutes in salty, aqueous media

### **2.3. Relaxed Complex Method and its Application in HIV Integrase Inhibitor Studies**

The use of computer models to screen for small molecules that will bind to and alter the activity of macromolecules has long been complicated by "induced fit" effects: the macromolecule may undergo shape changes during the formation of the complex with the small molecule. A new approach has been developed in which a variety of conformations of the target macromolecule are generated by molecular dynamics simulation or other methods, to generate structures that might include one or more that are representative of that of the "relaxed complex" with the small molecule [25]. To select the most relevant macromolecular conformations, rapid screening is done using AutoDock and ADT. The most stable complex structures from this screening are then subject to more refined analysis to yield the most probable structure of the complex and an associated estimate of the strength of binding (Figure 3C). More recently, applications of this method to an antiviral target for treatment of HIV infections [26] has aided scientists at Merck & Co. in developing a new class of drugs that are in Phase III clinical trial[27].

## **2.4. Finite Element Tool Kit (FETk) & Adaptive Poisson-Boltzmann Solver (APBS)**

The Finite Element Tool Kit (FEtk) [28] is an evolving collection of parallel adaptive multilevel finite element software libraries and supporting tools for solving coupled systems of partial differential equations (PDE) and integral equations (IE) [29]. The numerical libraries are written in an object-oriented form of ANSI-C (Figure 4A). Left (Top/Bottom): Potential contours of the electrostatic

[30]. Continuum electrostatics plays an important role in several areas of biomolecular simulation, including: simulation of diffusional processes to determine ligand-protein and protein-protein binding kinetics; implicit solvent molecular dynamics of biomolecules; solvation and binding energy calculations to determine ligand-protein and protein-protein equilibrium binding constants and aid in rational drug design; and biomolecular titration studies. APBS was designed to efficiently evaluate electrostatic properties for such simulations for a wide range of length scales to enable the investigation of molecules with tens to millions of atoms. APBS uses FEtk and PMG to solve the Poisson-Boltzmann equation numerically.

APBS is available to users as standalone applications or inside PMV for seamless access (Figure 4 B), and available in CHARMM or AMBER through the NCR iAPBS interface [31], or as web services using the NCR Opal toolkit [32]. In this example, the electrostatic binding energy for the protein kinase A complex with the inhibitor balanol was computed using APBS and displayed in PMV. The protein kinase A (PKA) is shown as a ribbon diagram with side chains or positively (blue) and negatively (red) charged amino acids displayed as Sticks and Balls. The inhibitor molecule is displayed using a fat Sticks and Balls representation and colored by atom types. The electrostatic binding energy is visualized by direct volume rendering with the shown transfer function, and using two isosurfaces.

A new software component from NCR that allows the efficient modeling of diffusion events across the neuromuscular junction in a steady state or time dependent manner called the Smoluchowski equation solver (SMOL) is also based on the FEtk and has been used in conjunction with APBS in a study of the tetrameric complex acetylcholine receptors [33].

### 3. Computational and data cyberinfrastructure to support multiscale modeling

In parallel to efforts focusing attention on the needs and the benefits on multiscale modeling, tremendous national and international investments have already been made to develop and deploy a cyberinfrastructure that will revolutionize the conduct of science [34]. This cyberinfrastructure consists of distributed computational, data storage, observational, and visualization resources, including human resources, connected by a network infrastructure and a software layer (middleware), that will “bring access of resources (at one end) to researchers (at another) and allow researchers to

conduct team science as part of normal conduct of science [35], in an end-to-end cyberinfrastructure”. Cyberinfrastructure, and grid, are often used interchangeably nowadays [36].

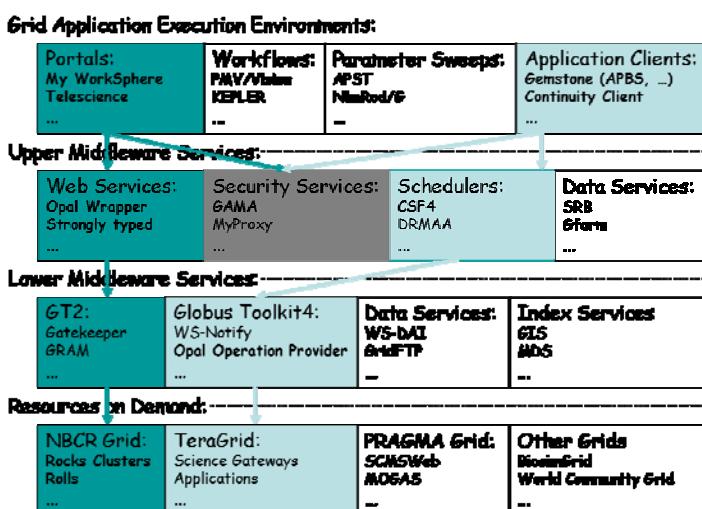


Figure 5. Selected Components of NCR Software Service Stack. Colored blocks and arrows indicate possible routes for distributed job execution. The different green shades show possible routes to access physical computational resources, and the arrows indicate direction of workflow. The grey shaded region indicates shared security services among the different workflows.

modeling needs to have the capability to handle the representative data commonly encountered in MSM, by representative users who need to accomplish tasks that are representative of the nature of MSM research and development. The scientific drivers for multiscale modeling at NCR are diverse, and spans the scale from subatomic electron charge density flows in development of better photodynamic therapies [39], to cellular and whole heart

An "end-to-end" cyberinfrastructure for multiscale

physiology modeling [1], with information integration technology providing the mediator necessary for interoperability. All these projects drive the development and integration effort we are leading to provide the cyberinfrastructure necessary and required for the research objectives to be met. In addition, NBCR works closely with other projects at UCSD, such as NCMIR, BIRN [40], CAMERA [41], Optiputer [42], JCSG [43], and establish important collaborations with national projects such as TeraGrid [44], Open Science Grid [45], and international projects such as PRAGMA [46] and OMII [47].

The development of infrastructure that can support diverse applications using distributed physical resources, and remain easy to use and scalable has ushered in the service-oriented architecture as the dominant *modus operandi* for reasons eloquently stated in [48]. To enable scientific applications on the grid, many different approaches have been adopted [49]. We have taken a minimalist approach, to select the most stable components, achieve the greatest leverage, and develop smart glues that are reusable components, within the service oriented approach (Figure 5). We'll highlight some key components developed by or with critical contributions from NBCR, and then discuss how they are used to support multiscale modeling efforts at NBCR, and are available to the MSM community in general.

Upper middleware services are those that make the development of distributed applications significantly easier, with support for higher levels of abstraction and standardization. For example, NBCR Opal based application web services provide job management and scheduling features based on the Globus toolkit. An application developer may begin using the grid quickly with the basic knowledge of web service development, as shown by the use cases in PMV, My WorkSphere, and Gemstone user environments. Lower middleware services are those that have stabilized over the years and serve as the foundation for the development of more sophisticated and transparent modes of access. However, as often dictated by performance requirements, a user application may access lower layers directly. This is much less desirable unless the integration is based on the service oriented architecture. Below we describe some key tools and technology developed or co-developed by NBCR actively to support our user communities. For a more comprehensive description of all the tools and software, please see NBCR download site [50].

### 3.1. *Opal – a simple yet powerful web service wrapper*

Opal is developed by NBCR as a Java based toolkit that automatically wraps any legacy applications with a Web services layer that is fully integrated with Grid Security Infrastructure (GSI) based security, cluster support, and data management [32,51]. The advantage of using OPAL is that the application may be launched using any Web service client, because the WSDL defines using the standard protocol how the service may be accessed, and it provides the basic HTTP access to results, as well as any metadata that describes how the results may be handled (Figure 6A). This ‘deploy once, and used by many’ feature of a Web service is a key ingredient of achieving interoperability. Because OPAL manages the interaction with the grid architecture, the grid is transparent to the user. In addition, workflow tools like Kepler [52] may easily compose web services based workflows through a common interface (Figure 6B).

(Figure 6)

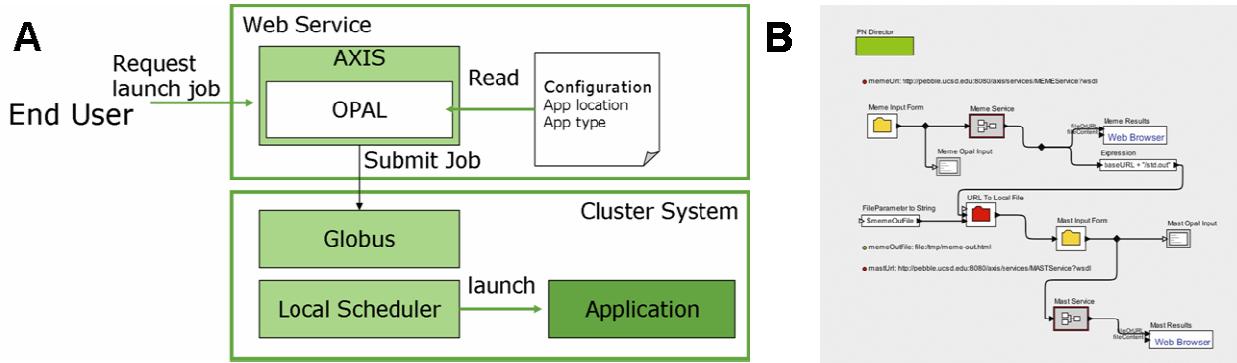


Figure 6. A) Opal allows rapid deployment of applications as web services using user provided configuration options. B) Workflow programs such as Kepler may orchestrate Opal based web services with reusable actors.

### 3.2. Grid Account Management Architecture (GAMA)

GAMA, co-developed by NBCR [53], is a GSI-based security service that manages X.509 user credentials on behalf of users, and supports SOAP-based applications (Figure 7A). The server component leverages existing software packages such as CAS, MyProxy, and CACL, with the GAMA version 2 supporting other CA packages such as NAREGI. A portlet component provides the administrative interface to the server. As security is a sensitive and critical issue in the production use of the cyberinfrastructure, GAMA allows any (Figure 7) organization to create their own certificate authority, manage their user certificates, secure the SOAP communications using HTTPS and mutual authentication, and integrate seamlessly with portals, and rich clients. For example, GAMA is used by GridSphere based portals [54], Gemstone [55] and PMV/Vision.

### 3.3. Application TO Middleware Interface Component (ATOMIC)

The ATOMIC bundle (Figure 7B), developed for the Telescience Project by NBCR, provides a fabric for seamless interoperability among user interfaces (web portals and applications) and externally addressable grid resources (instruments and computers) inside a grid portal environment [56]. There is a unified programming API for basic scientific research which liberates the users and programmers from having to learn about the grid. TeleAuth is a specialized version of GAMA. TeleWrap provides seamless access to remote data transparently, and has been used successfully in the Telescience portal and EMWorkspace. TeleRun provides a higher level of abstraction than Opal-provided services, and may further shield developers from the details of grid execution environment. ATOMIC also provides persistent session information during a particular Telescience session, so that all user portlets have access to the same session information regardless of the specific portlet being used.

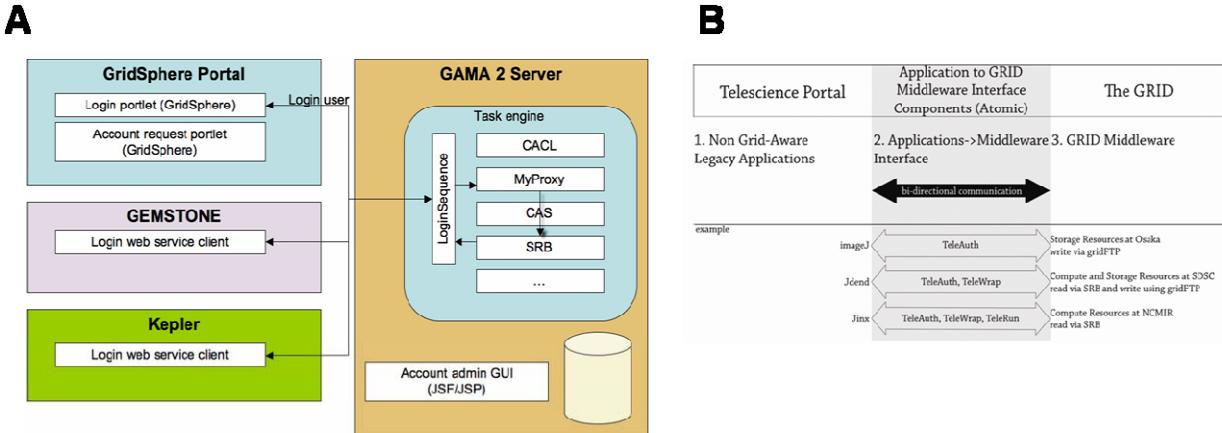


Figure 7. A) GAMA provides web service based access for GSI and proxy management. B) ATOMIC leverages GAMA for TeleAuth and adds transparent data access and job execution for the Telescience portal environment.

### 3.4. MYMPI

The development of MYMPI [57] is an effort of NBCR, driven by the requirement of Continuity, to improve interprocess communications between Python, FORTRAN [17]. It is a Python module used with a normal Python interpreter. MyMPI is an implementation following the standard specifications of MPI, supporting more than 30 of the commonly used MPI calls. The syntax of the calls matches the syntax of C and FORTRAN calls closely to support mixed Python with the other languages. MYMPI will also allow access from Python other application packages such as the Finite Element Toolkit (FEtk) [58], a collection of important software libraries and differential equation solver packages, with critical contribution from NBCR for continued development.

### 3.5. Gemstone, My Worksphere as Portals to Grid Services

Gemstone [59] is developed with major support from the NSF National Middleware Initiative, in close collaboration with NBCR, especially in the interaction with strongly typed and Opal based web services (Figure 8 A). For example, Gemstone provides an interface to GAMESS [60] and APBS strongly typed web services [30], both developed by

NBCR. It also supports the GARNET visualization web service from the NMI (NSF Middleware Initiative) project. Opal based web service for PDB2PQR [61], a utility package for APBS calculations, is also accessible from Gemstone as web service based workflows. Gemstone utilizes the open source Mozilla engine, and uses XML User Interface Markup Language (XUL) to describe user interface. *My WorkSphere* is a prototype GridSphere based portal environment that leverages portlets that are JSR168 standard compliant. As an example, we've deployed MEME [62] as a portlet using Opal, as part of a generalized Cyberinfrastructure for bioinformatics applications [51], with Gfarm [63] and CSF4 [64] as the other key middleware components.

### 3.6. Rocks and Rock'n Roll

In an effort to make cyberinfrastructure more readily available to scientists and engineers, it is necessary to not only develop different middleware to support legacy applications, but also to make the different software packages easy to deploy into existing infrastructure. Rocks cluster environment toolkit [65] (Figure 8B) has proven to be invaluable for NBCR to build the basic infrastructure, deploy our software stack, and make our infrastructure replicable by others. NBCR has contributed critically to the development of the Condor roll [66], (a roll is a mechanism, similar to the Red Hat RPM though fully automated, for building reproducible cluster and grid environment). Other rolls available from NBCR include the APBS, MEME, GAMA, AutoDock, and PMV. Additional rolls for SMOL [67], FEtk will be available soon (Figure 8 B).

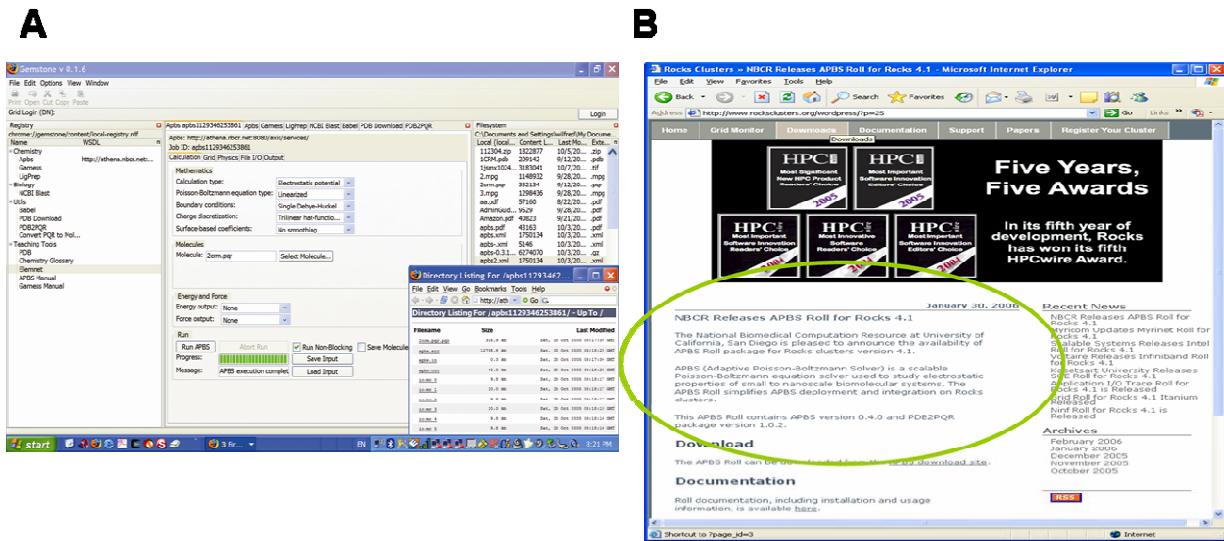


Figure 8. A) Gemstone offers a light weight mode of access to distributed resources using Opal based web services or strongly typed web services. B) Rocks cluster environment software has provided a replicable infrastructure for easy deployment and production use.

### 3.7. Networks, pathways, ontologies and graph query engines

PathSys is a general-purpose, scalable, graph data warehouse of biological information, complete with a graph manipulation and a query language, a storage mechanism and a generic data-importing mechanism through schema-mapping. The client software, named BiologicalNetworks, supports the navigation and analyses of molecular networks (Figure 9B). As systems biology integrates information from many data sources, distributed or integrated, there is an increasing demand for managing, querying and automated reasoning using ontology concepts and description logic. OntoQuest is developed to provide the extended mapping schemes for storing OWL (web ontology language) ontologies into backend databases (Figure 9A). It is aimed to guide a user to explore ontological datasets, and eventually make non-preconceived, impromptu discoveries.

Both PathSys and OntoQuest take advantages a new semantic-aware RDF (resource description framework) algebra, which supports the inference of complex relationships represented in ontological hierarchies. OntoQuest is being developed using the Cell Centered Database (CCDB) and neuroscience ontology, whereas PathSys is driven by research in yeast signaling pathways. The technology is extensible to the data integration between distributed data and

computation services described with appropriate semantics and ontology terms (Figure 9). The technology being developed under OntoQuest may also be used to provide additional semantic annotations to Opal based web services to improve automated service discovery and utilization.

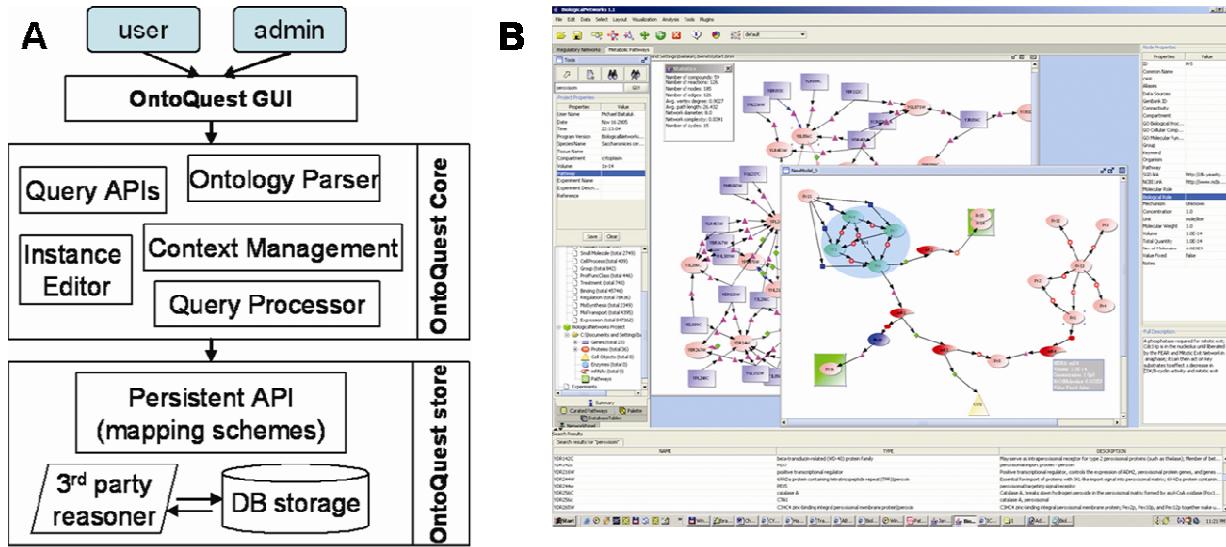


Figure 9. A) OntoQuest architecture: ontology based query engine and database environment. B) BiologicalNetworks is a client interface to the PathSys database warehouse, with support for many xml data formats.

## 4. Summary & Discussion

There are a plethora of tools and innovative approaches to the development of cyberinfrastructure to support multiscale modeling activities. However, better and more robust approaches will always come out of close collaborations between computer scientists and biomedical researchers, as well as other field specialists. The interactions will educate all the groups to be fully aware of the requirements and challenges of the state of the art technology, and make routine use of the grid possible today. In addition, the development of new tools that support applications in different fields and through international collaborations greatly reduces the collective cost for global computational grids. The service oriented approach is gaining momentum and greatly facilitates the development of a knowledge based global economy.

### 4.1. Collaborations needed to build robust grid computing platform

There are often conflicting requirements between biomedical research and grid communities, with the former accustomed to a “one experiment at a time” approach, and the latter desiring systems that handle large number of jobs simultaneously. The reality is that current grid computing systems are still difficult to learn, only relatively stable, and limited by the technologies available in hardware, software, and programming languages, as noted in [3]. It is a challenge to both communities to design better software and use them effectively. With these caveats in mind, one may attempt to select from the available tools, and build a robust platform to make routine use of the grid possible, through close collaborations with developers of all components involved where possible. By working with different applications and addressing the common needs and individual requirements, reusable components may be isolated without sacrificing the customized environments demanded by users. The structure of resources supported by the National Center for Research Resources, such as NBCR and NCMIR, requires collaborative project to guide the development of the tools, as noted above.

### 4.2. The Iterative cycle of science and technology development

The number of tools showed in Figure 5, along with those represented only as “...”, offer many possible combinations to build end to end problem solving environments, often with overlapping features. For every tool listed, there are solid alternatives under different use cases. As shown in the usage scenarios, the ultimate choice of tools depends on the specific problems to be solved and the target audience of the designed environment. As demonstrated by the

number of tools that are using Opal based services, the service oriented approach provides the flexibility in customized front-end tools with transparent access to underlying distributed computation resources. The demands of multiscale modeling applications will drive the development of the computation and data cyberinfrastructure, which in turn enables simulation based biology and medicine. The ensuing challenges in the search for solutions of more difficult problems will further this iterative developmental cycle of technology and science, with continuous and immediate impact to the health and well being of the public.

#### **4.3. Future directions**

In developing collaborations and ultimately a user base to conduct research, having a persistent infrastructure is required. By this we mean access to tools and software for researchers to experiment with. The National Centers for Research Resources provide that type of infrastructure for the community. In the case of NBCR, access to the tools and services discussed in this article can be found at the NBCR website.

While the experiences and developments discussed above have been informed by our local activities, the concepts of cyberinfrastructure are not bound by national borders. Examples of large scale projects and smaller team projects have been commented on in prior CTWatch issues (e.g. February 2006). Furthermore, it is essential that we actively participate in the international arena, both to ensure the various CI efforts can interoperate (e.g., when we want to use unique resources in other countries) and to reduce reproduction of tools that already exist.

Another exciting development that will link biomedical research with the environment, across various scales, the newly funded CAMERA project, which will provide a community resource that links genetic information from a metagenomics effort, the Global Ocean Survey, with environmental factors such location, temperature and chemistry of the sampled environment.

The critical role played by NBCR to bridge the multiscale modeling research community and cyberinfrastructure is an important one, and will require collaborative efforts from other projects to expedite the translational/medical potential of multiscale biomedical and cyberinfrastructure research.

## **5. Acknowledgment**

The authors are all members of the National Biomedical Computation Resource, which is supported by the NIH National Centers for Research Resources award P41-RR08605. Major support for APBS is provided by NIH grant GM069702 to NAB.

## **6. References**

1. McCulloch AD, Paternostro G (2005) Cardiac systems biology. *Ann N Y Acad Sci* 1047: 283-295.
2. Hunter PJ, Borg TK (2003) Integration from proteins to organs: the Physiome Project. *Nat Rev Mol Cell Biol* 4: 237-243.
3. Hunter P, Nielsen P (2005) A strategy for integrative computational physiology. *Physiology (Bethesda)* 20: 316-325.
4. Hunter PJ, Li WW, McCulloch AD, Noble D (2006) Multi-scale Modeling Standards, Tools Databases for the Physiome Project. Computer: In Press.
5. Oden JT, Belytschko T, Fish J, Hughes TJ, Johnson C, et al. (2006) Revolutionizing Engineering Science through Simulation.
6. IMAG: (2004) Interagency Opportunities in Multi-Scale Modeling in Biomedical, Biological, and Behavioral Systems 2006: <http://www.nsf.gov/pubs/2004/nsf04607/nsf04607.htm>.
7. PITAC: (2005) Computational Science: Ensuring America's Competitiveness. <http://www.nitrd.gov/pitac/reports/index.html>.
8. NRCNA: (2005) Catalyzing Inquiry at the Interface of Computing and Biology. <http://www.doegenomestolife.org/pubs/NRCCComputingandBiology/index.shtml>.
9. NBCR: (2005) National Biomedical Computation Resource. <http://nbcr.net>.
10. NCRR: (2006) National Center for Research Resources. <http://www.ncrr.nih.gov>.

11. MCMBR: (2004) Multiscale Computational Modeling for Biomedical Research. <http://nbcr.net/physiome>.
12. McCulloch AD, Arzberger PW, Hunter P (2005) Computational Physiology: From Genome to Physiome. *The Physiologist* 49: 94-95.
13. DCBC: (2002) Data and Collaboratories in the Biomedical Community. <http://nbcr.net/Collaboratories/index.html>.
14. Usyk TP, McCulloch AD (2003) Relationship between regional shortening and asynchronous electrical activation in a three-dimensional model of ventricular electromechanics. *J Cardiovasc Electrophysiol* 14: S196-202.
15. Trayanova N (2005) In pursuit of the elusive culprit. *Heart Rhythm* 2: 729-730.
16. Xie F, Qu Z, Yang J, Baher A, Weiss JN, et al. (2004) A simulation study of the effects of cardiac anatomy in ventricular fibrillation. *J Clin Invest* 113: 686-693.
17. Kaiser TH, Brieger L, Healy SN. (2006) MYMPI - MPI Programming in Python. International Conference on Parallel and Distributed Processing Techniques and Applications; Las Vegaspp. In Press.
18. Koh X, Srinivasan B, Ching HS, Levchenko A (2006) A 3D Monte Carlo Analysis of the Role of Dyadic Space Geometry in Spark Generation. *Biophys J* 90: 1999-2014.
19. Soeller C, Cannell MB (2004) Analysing cardiac excitation-contraction coupling with mathematical models of local control. *Prog Biophys Mol Biol* 85: 141-162.
20. Sanner MF, Stoltz M, Burkhard P, Kong X-P, Min G, et al., editors (2005) Visualizing Nature at Work from the Nano to the Macro Scale.: John Wiley & sons, Ltd. 7-11 p.
21. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 46: 34-40.
22. Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* 9: 1-5.
23. Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* 10: 293-304.
24. WCG: (2005) World Community Grid: FightAIDS@Home. [http://www.worldcommunitygrid.org/projects\\_showcase/viewFaahFaq.do?shortName=faah](http://www.worldcommunitygrid.org/projects_showcase/viewFaahFaq.do?shortName=faah).
25. Lin JH, Perryman AL, Schames JR, McCammon JA (2003) The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* 68: 47-62.
26. Schames JR, Henchman RH, Siegel JS, Sottriffer CA, Ni H, et al. (2004) Discovery of a novel binding trench in HIV integrase. *J Med Chem* 47: 1879-1881.
27. NIH: (2006) Clinical Trials Database: . <http://clinicaltrials.gov/ct/show/NCT00293254?order=1>.
28. FEtk: (2005) Finite Element Toolkit. <http://www.fetk.org>.
29. Holst MJ (2001) Adaptive Numerical Treatment of Elliptic Systes on Manifolds. *Advances in Computational Mathematics* 15: 139-191.
30. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98: 10037-10041.
31. iAPBS: (2006) iAPBS Interface. <http://nbcr.net/forum/viewforum.php?f=17>.
32. Krishnan S, Stearn B, Bhatia K, Baldridge K, Li WW, et al. (2006) Opal: Simple Web Service Wrappers for Scientific Applications. International Conference for Web Services. In Press.
33. Zhang D, Suen J, Zhang Y, Song Y, Radic Z, et al. (2005) Tetrameric mouse acetylcholinesterase: continuum diffusion rate calculations by solving the steady-state Smoluchowski equation using finite element methods. *Biophys J* 88: 1659-1665.
34. Atkins D, Kroegemeier K, Feldman S, Garcia-Molina H, Klein M, et al. (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Arlington, VA: National Science Foundation.
35. Hey T, Trefethen AE (2005) Cyberinfrastructure for e-Science. *Science* 308: 817-821.
36. Graham SL, Snir M, Patterson CA (2004) Getting Up to Speed: The Future of Supercomputing: The National Academies Press. 308 p.
37. Foster I, Kesselman C, editors (2004) The Grid 2: Blueprint for a New Computing Infrastructure. 2 ed. San Francisco: Morgan Kaufmann Publishers, Inc.
38. Bal H, Casanova H, Dongarra J, Matsuoka S (2004) Application-Level Tools. In: Foster I, Kesselman C, editors. The Grid 2. 2 ed. Amsterdam: Elsevier.

39. Yerushalmi R, Scherz A, Baldridge KK (2004) Direct experimental evaluation of charge scheme performance by a molecular charge-meter. *J Am Chem Soc* 126: 5897-5905.
40. BIRN: (2005) Biomedical Informatics Research Network. <http://www.nbirn.net>.
41. CAMERA: (2006) Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis. <http://camera.calit2.net/>.
42. OptIPuter: (2005) A Powerful Distributed Cyberinfrastructure to Support Data-Intensive Scientific Research and Collaboration. <http://www.optiputer.net>.
43. JCSG: (2005) Joint Center for Structural Genomics. <http://www.jcsg.org>.
44. TeraGrid: (2004) TeraGrid. <http://www.teragrid.org>.
45. OSG: (2006) Open Science Grid. <http://www.opensciencegrid.org/>.
46. PRAGMA: (2004) Pacific Rim Applications and Grid Middleware Assembly. <http://www.pragma-grid.net>.
47. OMII: (2006) Open Middleware Infrastructure Institute. <http://www.omii.ac.uk/>.
48. Foster I (2005) Service-oriented science. *Science* 308: 814-817.
49. Abramson D, Lynch A, Takemaya H, Tanimura Y, Date S, et al. (2006) Deploying Scientific Applications to the PRAGMA Grid testbed: Strategies and Lessons. Sixth IEEE International Symposium on Cluster Computing and the Grid, 2006 CCGrid 06; Singapore. 1: pp. 241-248.
50. NBCR: (2006) NBCR Software and Tools Description and Download Site. <http://nbcr.net/tools.php>.
51. Li WW, Krishnan S, Mueller K, Ichikawa K, Date S, et al. (2006) Building cyberinfrastructure for bioinformatics using service oriented architecture. Sixth IEEE International Symposium on Cluster Computing and the Grid Workshops Singapore. 2: pp. 39-46.
52. KEPLER: (2006) The Kepler project. <http://kepler-project.org/>.
53. Bhatia K, Chandra S, Mueller K. (2006) GAMA: Grid Account Management Architecture. 1st IEEE International Conference on e- Science and Grid Computing; Melbourne, Australiapp. In Press.
54. Gridsphere: (2004) GridSphere. <http://www.gridsphere.org>.
55. Baldridge K, Bhatia K, Greenberg JP, Stearn B, Mock S, et al. (2005) GEMSTONE: Grid Enabled Molecular Science Through Online Networked Environments. Life Sciences Grid Workshop; Singapore. World Scientific Press.
56. Lin AW, Dai L, Ung K, Peltier S, Ellisman MH. (2005) The Telescience Project: Applications to Middleware Interaction Components. The 18th IEEE International Symposium on Computer-Based Medical Systems; Dublin, Irelandpp. 543-548.
57. MYMPI: (2006) My MPI. <http://nbcr.net/forum/viewforum.php?f=16>.
58. FEtk: (2006) Finite Element Toolkit. <http://www.fetk.org>.
59. Baldridge KK, Bhatia K, Greenberg JP, Stearn B, Mock S, et al. (2005) Grid-Enabled Molecular Science through Online Networked Environments. Life Sciences Grid Workshop; Singapore. World Scientific Press. 1.
60. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, et al. (1993) General atomic and molecular electronic structure system. *Journal of Computational Chemistry* 14: 1347-1363.
61. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: W665-667.
62. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34: W369-373.
63. Li WW, Arzberger PW, Yeo CL, Ang L, Tatebe O, et al. (2005) Proteome Analysis using iGAP in Gfarm. The Second International Life Science Grid Workshop 2005; Grid Asia 2005, Singapore. World Scientific Press. 1.
64. Wei X, Li WW, Tatebe O, Xu G, Hu L, et al. (2005) Integrating Local Job Scheduler - LSF with Gfarm. *Lecture Notes in Computer Science* 3758: 197.
65. ROCKS: (2005) Rocks Cluster Distribution. <http://www.rocksclusters.org>.
66. Litzkow M, Livny M, Mutka M. (1988) Condor - a hunter of idle workstations. Proceedings of the 8th International Conference of Distributed Computing Systems; June 1988pp. 104-111.
67. Song Y, Zhang Y, Shen T, Bajaj CL, McCammon JA, et al. (2004) Finite element solution of the steady-state Smoluchowski equation for rate constant calculations. *Biophys J* 86: 2017-2029.