Grid Computing for hybrid-QM/MM platform: *Construction and Applications*

Haruki Nakamura

Laboratory of Protein Informatics

Research Center for Structural and Functional Proteomics

Institute for Protein Research

Osaka University

http://www.protein.osaka-u.ac.jp/rcsfp/pi/

BioGrid at Osaka *http://www.biogrid.jp*

Started from 2002

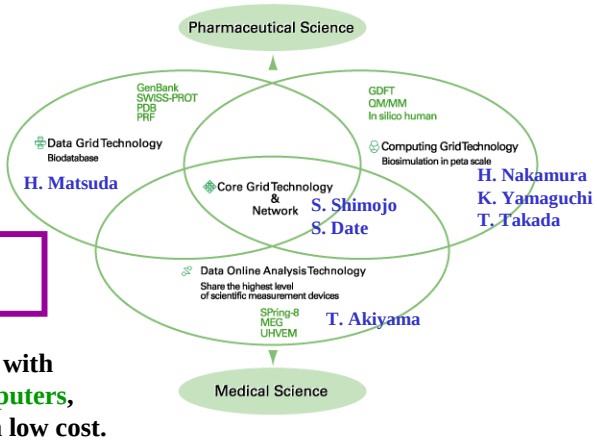
Goals:Grid Technology Development for:
Biotechnology
(drug discovery) and Medical Sciences.

Leader: Shinji Shimojo (CMC, Osaka Univ.)

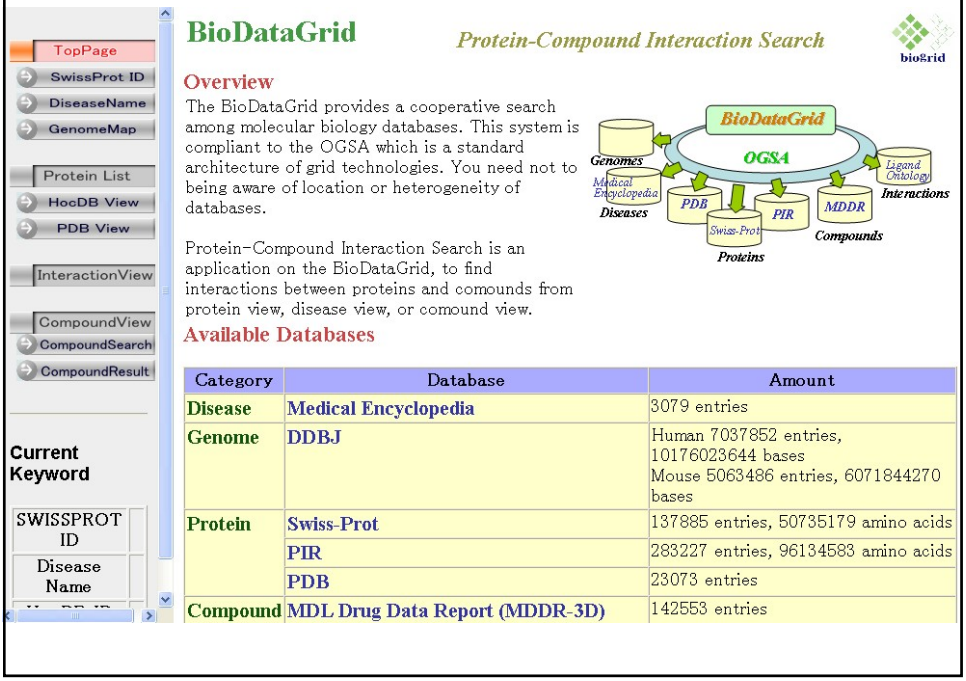
Government Support (MEXT): 5years

Integration of Data Grid and Computing Grid

to access Integrated and Standardized Databases with High-performance Computers, from anywhere and with low cost.



BioDataGrid by Prof. Hideo Matsuda (Osaka Univ.)



Category	Database	Amount
Disease	Medical Encyclopedia	3079 entries
Genome	DDBJ	Human 7037852 entries, 10176023644 bases Mouse 5063486 entries, 6071844270 bases
Protein	Swiss-Prot	137885 entries, 50735179 amino acids
	PIR	283227 entries, 96134583 amino acids
	PDB	23073 entries
Compound	MDL Drug Data Report (MDDR-3D)	142553 entries

PDBj (Protein Data Bank Japan)

http://www.pdbj.org/

Data Curation/Editing/Distribution

(About 20% of the world data are processed for Asia and Oceania.)

1) Construction of Advanced PDB database

.**Development of XML description for PDB data.PDB-XML.**

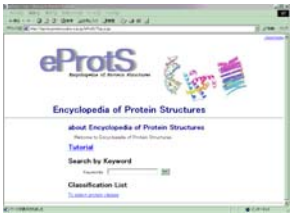
.**Addition of Exp. Condition and Biochem. Function Information**

2) Construction of Secondary Databases

.**Protein Molecular Surface Database, *e*F-site (Nakamura & Kinoshita.**

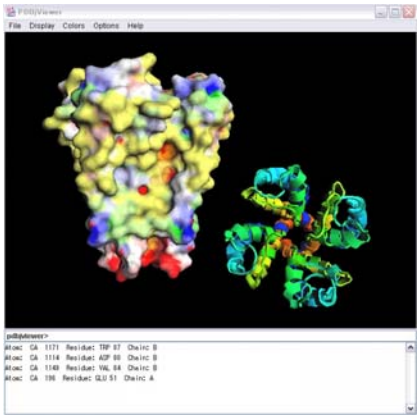
.**Protein Dynamics Database, *Pro*Mode (Wako & Endo)**

.**encyclopedia of Protein Structures, eProtS (Nakamura & Ito.**

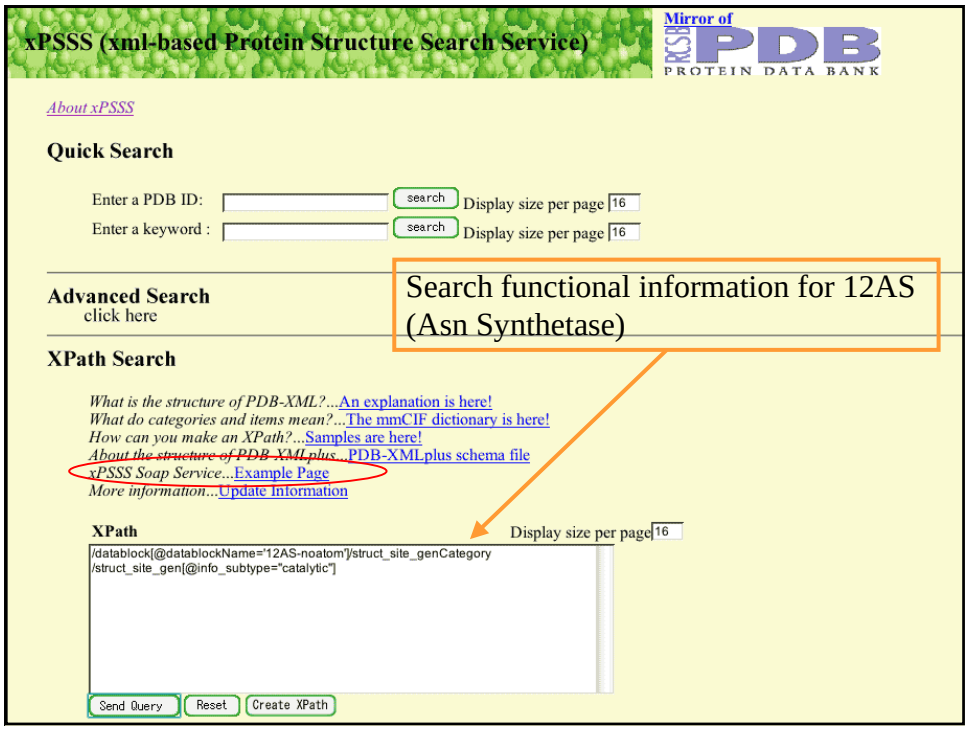


PDBjViewer ver 1.0

http://www.pdbj.org/PDBjViewer/



- Offers interactive molecular visualization with the RasMol type commands.
- Can be used both as a Stand-alone and as an Applet with Java3D.
- Any polygons defined by XML can be displayed and manipulated quickly.
- Can parse a PDB-XML file and display the molecule



Multiscale Simulation@BioGrid and Biosimulation Project

(Development of models for disease and drug action)

Tissue & Organera simulation
by H. Kitaoka

Cell simulation
by S. Suzuki & Y. Kurachi

***Ab initio* Protein Fold Simulation**
by S. Takada

Protein molecular Simulation
by H. Nakamura

Electronic state Simulation
by K. Yamaguchi & T. Takada

Physiome

Cellome

Proteome

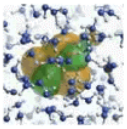
In silico drug screening

Biochemical reactions

BioPfuga:

Biosimulation Platform United on Grid Architecture

A platform where individual application programs at the different levels are united to execute a hybrid computation. In particular, *BioPfuga* is a platform for biosimulation.



http://www.biogrid.jp/

BioPfuga

Development of *BioPfuga* (Biosimulation Platform United on Grid Architecture)

The usual usage of the grid architecture is running one computation on many distributed CPUs through a rapid network. However, in order to analyze much more complicated biological systems, composed of simulations at different levels, along the new paradigm for biological science, more integrated computational approaches are required. The individual programs should be driven on their own corresponding machines on the grid system. For this purpose, we have designed and developed a new platform, *BioPfuga* (**Biosimulation Platform United on Grid Architecture**) where individual applications, corresponding to the different levels of bio-simulations, are united and executed as a hybrid application.

Construction of *BioPfuga*

- Propose and Design Communication between program components by XML description (BMS-ML: BioMolecular Simulation-ML) , and Creation of the libraries to handle it.

- Divide programs into a set of many components

→ Different program modules are then implemented as the service programs based on the OGSA (Open Grid Service Architecture) mechanism.

XML description for exchanging data

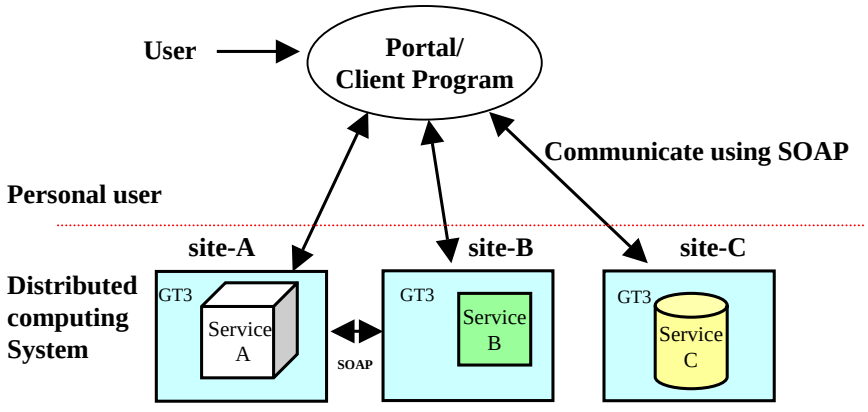
Items described with XML (BMS-ML)

1) **Contents** (atomic coordinates, charges, MO, etc)
2) **Unit** (AU, kcal/mol, angstrom, etc)
3) **Description Form** (Text, hexadecimal form, Base64 form)
4) **Data Type** (Character, Integer, Floating, Double Precision, etc)
5) **Data Form** (Scalar, Vector, Tensor, etc)
6) **Data Size**
7) **Data**

Example of BMS-ML *form = Base64* . (x4/3)

```
<?xml version="1.0" encoding="UTF-8"?>
<bms_data size="1" count="200" unit="AU" form="b64">
   <bms_content> sample_b64.xml </bms_content>
   <bms_comment> Sample of BMS-ML (hexdec) </bms_comment>
   <bms_array_count>4</uds_array_count>
   <bms_array seq="1" element="character" length="76">
pLWkq6TipMik0qS1pLekzrrupMOkv6XQpaSlyqXqpcehvKW/pPKlqKXzpbOhvKXJp
LekxqS9pM6l</bms_array>
   <bms_array seq="2" element="character" length="76">
x6G8pb+k8qPYo82jzLfBvLCkx6XVpaGlpKXrpMvK3cK4pLmk66XXpe2lsKXppeCkz
qXGpbmlyKTH</bms_array>
   <bms_array seq="3" element="character" length="76">p
Lmho6XXpe2lsKXppeCkrMC1vu+ky8awpKSkxqSkpOu+7LnnpM+ks6TOyrikrMC1p
Lekr8bJpOGk</bms_array>
   <bms_array seq="4" element="character" length="40">xqSkpOukz6S6pMekuaG
jpMmkpqTHpLek56Smoak=</bms_array>
</bms_data>
```

Configuration of a computing system on a Grid environment with several different program modules.



(GT3: Globus Toolkit version 3.02 of OGSA-Open Grid Service Architecture)

Integration of Multiscale Simulations

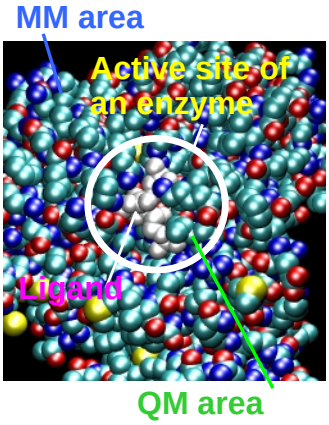
Molecular Structure Simulation

With Grid computing architecture

Electronic Structure Simulation

Quantum mechanics (QM) and Molecular mechanics (MM) simulation coupled on *BioPfuga*

Coupled simulation on Grids



Hamiltonian of total system

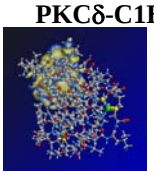
$$H_{total} = H_{QM}(x,x) + H_{QM/MM}(x,y) + H_{MM}(y,y)$$

Calculate in QM area

Calculate in MM area

Simulation of Electronic structure (1)

PKCδ-C1B



856 atoms, 8672 AO's

AMOSS ***Ab initio*** **Molecular Orbital Simulation for Supercomputer**
developed by NEC Quantum Chemistry Group

Rapid computation for huge molecular systems (20,000 basis for RHF, 1,000 basis for CASSCF and MP2), with high parallel performance

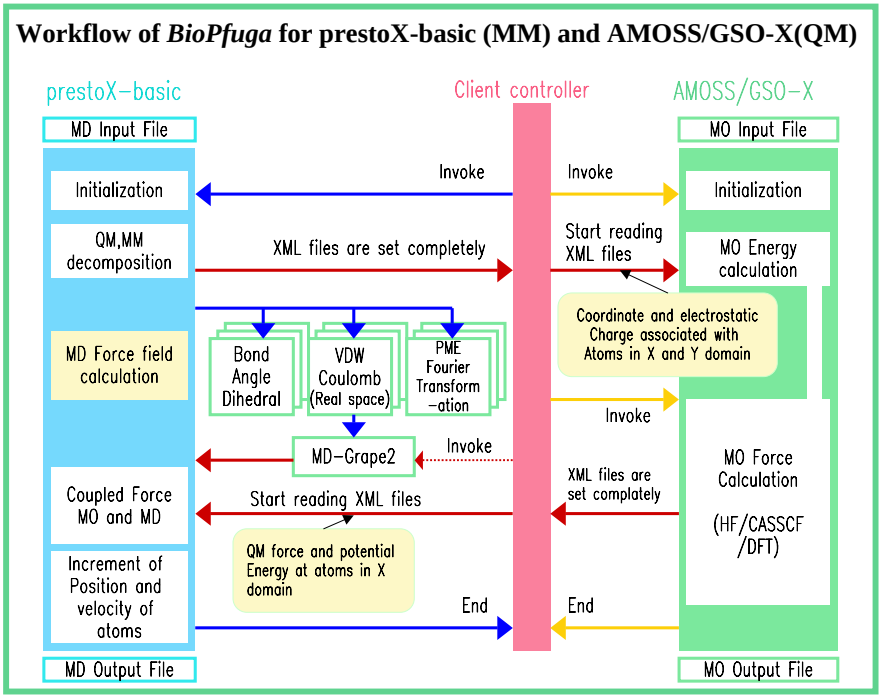
Simulation of Electronic structure (2)

GSO-X **generalized spin density function theory (DFT)**
developed by Shusuke Yamanaka & Kizashi Yamaguchi, at Osaka Univ.

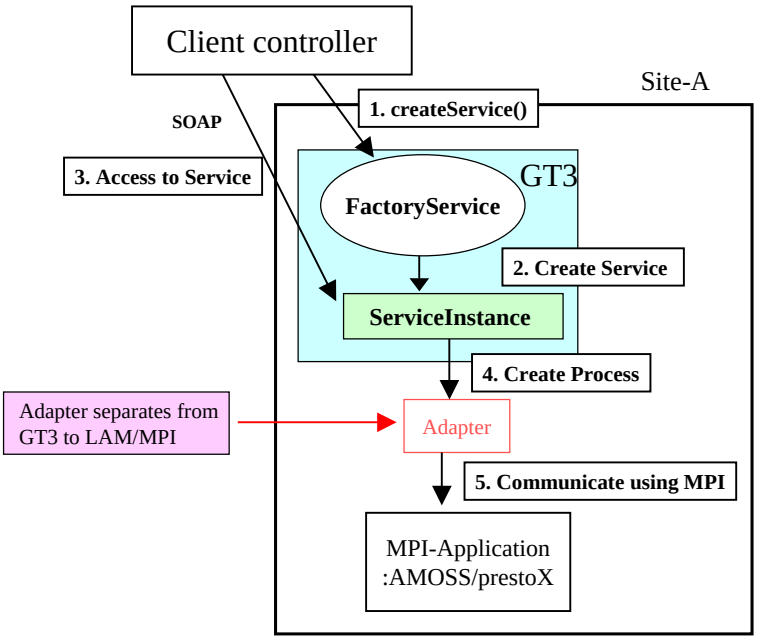
Simulation of Protein-solvent structure

prestoX-basic **Protein Engineering Simulator eXtended –basic versin-**
developed by Yoshifumi Fukunishi at JBIRC-AIST and Haruki Nakamura at Institute for Protein Research, Osaka Univ.

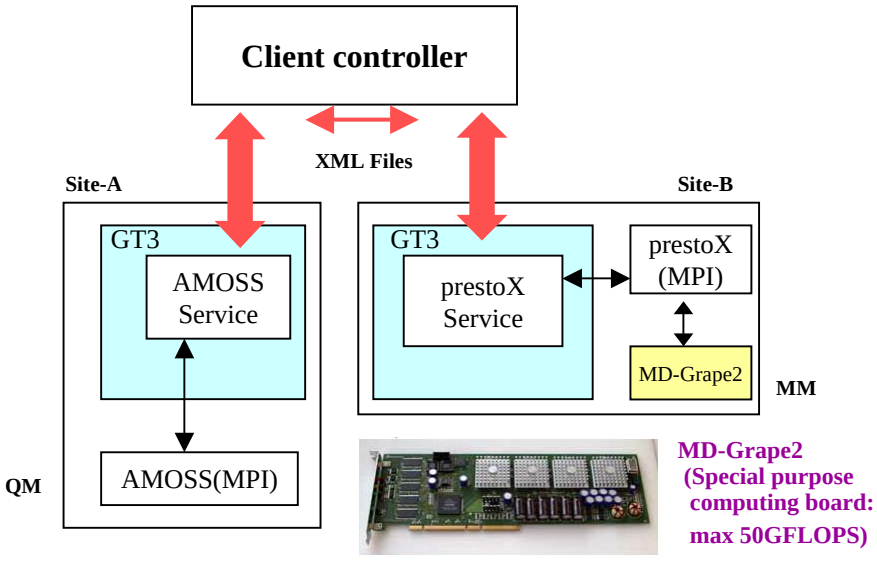
Workflow of *BioPfuga* for prestoX-basic (MM) and AMOSS/GSO-X(QM)



Architecture of the Grid Service



A hybrid-QM/MM with *BioPfuga* on the Grid Service

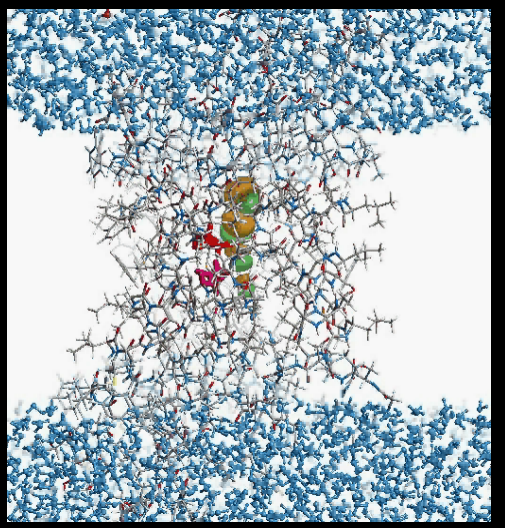


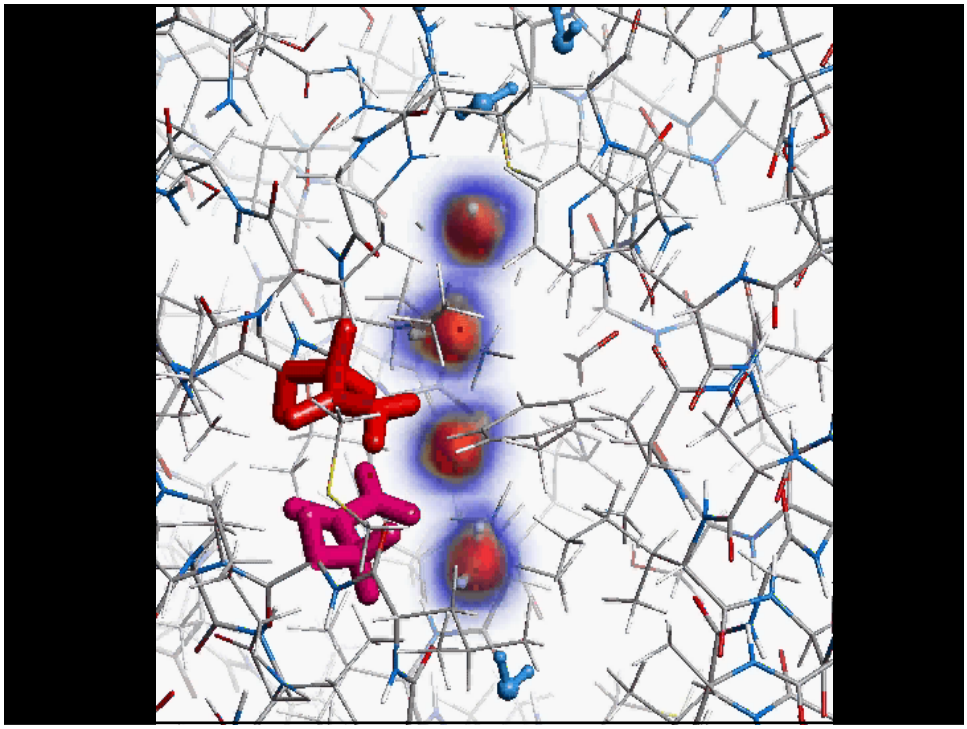
Summary

- For distributed computing environments, we have expanded the QM/MM system using architecture of Globus Toolkit Ver3.0.2.
- The three big programs, AMOSS/GSO-X and prestoX, which form the engines of the QM/MM simulations, were integrated into service components with slight modifications.
- The intermediate data communication between AMOSS/GSO-X and prestoX is performed by using XML.
- Communication between Grid service and MPI-module became possible, with an adapter module insulating MPI-modules from the Grid Service.
- Users at a great distance can access individual services as they are collected on a desk top.

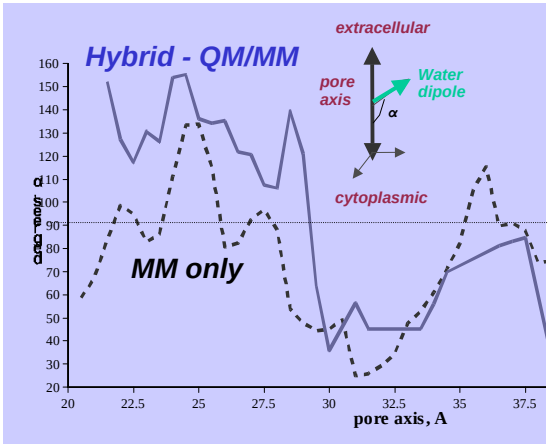
Molecular system: Aquapolin-1 with water

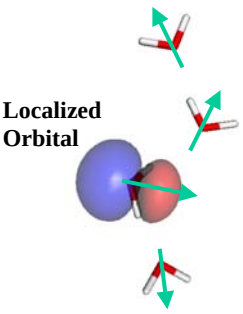
- 4 water molecules near the NPA residues of bovine aquapolin-1 were treated by Quantum Mechanics (MINI-4).
- Other 4,393 water molecules were treated by Molecular Mechanics with the TIP3P model.
- Total 16,962 atoms were simulated without truncation at 310 K.
- Canonical ensemble.





Orientation of the Water Molecules during the hybrid-QM/MM simulation





Hybrid-QM(HF or DFT)/MM
for a hydrogen bond with π-electron.

Hydrogen-bonds associated with the π-electron cloud

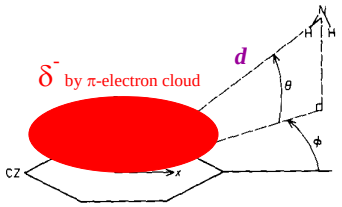
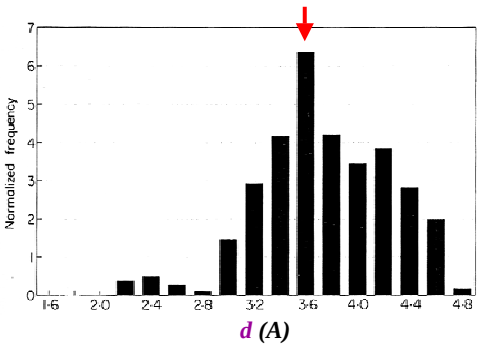


Figure 2. Definition of the spherical polar angles θ, ϕ for an interacting aromatic–amino pair. Reference Phe-ring lies on the x–y plane. An NH2 group is shown above the plane of the ring. θ defines the position of the NH2 above the x–y plane. ϕ defines the position of NH2 in the x–y plane.

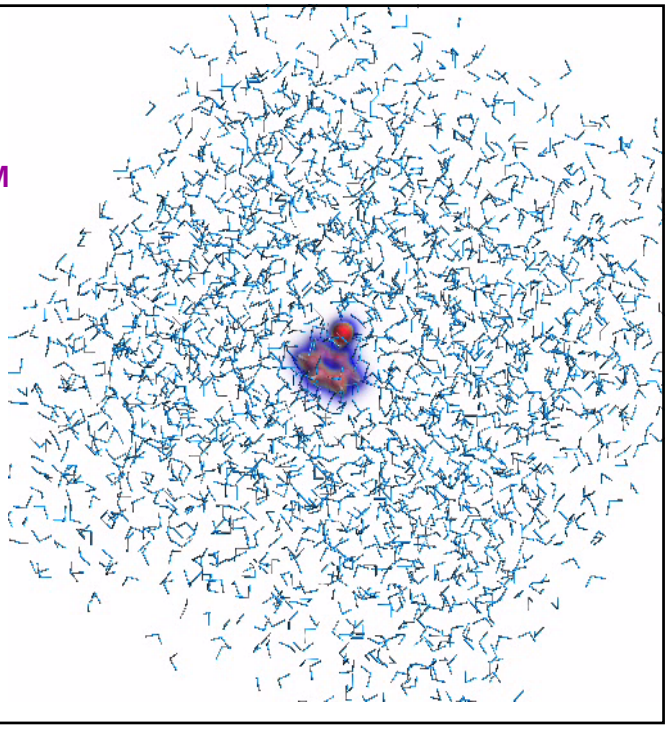


Burley & Petsko (1987), Levitt & Perutz (1988)
Singh & Thornton (1990)

UDFT (BLYP)

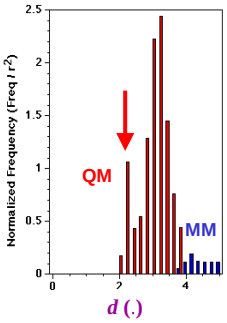
Hybrid-QM(DFT)/MM
for a hydrogen bond with π-electron.

Benzene 1
H_2O 1
TIP4P 1935
PME
(40x40x40.^{3})

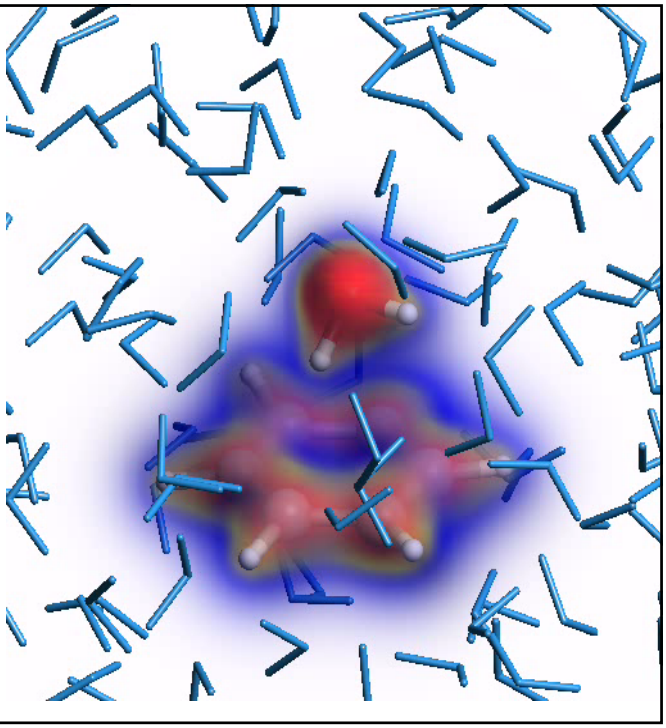


UDFT
(BLYP)

Hybrid-QM(DFT/HF)/MM for a hydrogen bond with π-electron.



Probability of H atom of the water located at *d* from the center of the benzene ring.

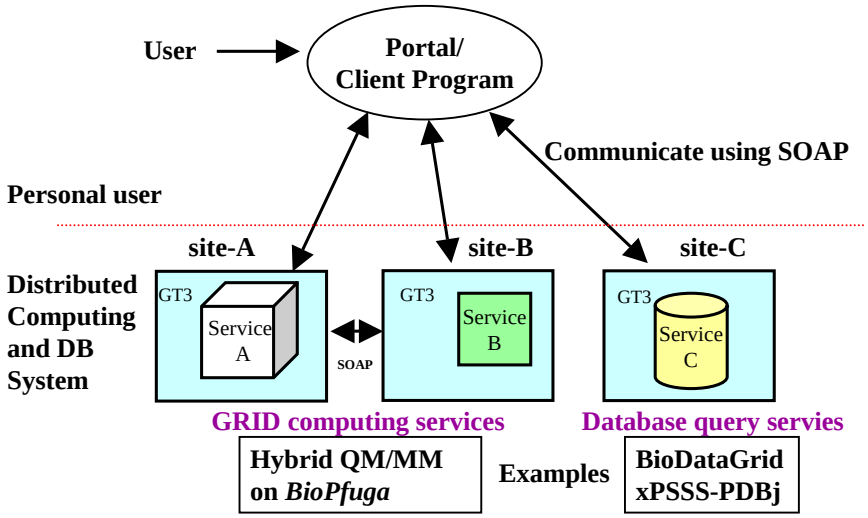


Comments for Multiscale Computations

- **Simlations at different levels can be integrated based on the GRID architecture:**
  1) **Usage of simultaneous computation services like a hybrid-QM/MM on *BioPfuga*.**
  2) **Usage of database query services for the pre-calculated or extracted data from other databases.**

- **Standard descriptions by XML are required.**

Configuration of computing systems and Database query systems on a Grid for multiscale simulations.



Collaborators

Yasushige Yonezawa, Rossen Apostolov
(Institute for Protein Research, Osaka University)
Kohei Ichikawa, Susumu Date, Hideo Matsuda, Shinji Shimojo
(Graduate School of Information Science and Technology, Osaka University)
Shusuke Yamanaka, Kizashi Yamaguchi
(Graduate School of Science, Osaka University)
Kazuto Nakata, Toshihiro Sakuma, Toshikazu Takada
(NEC Co. Ltd.)
Takashi Korosawa, Shirun Ho (Hitachi Co. Ltd.)
Yoshifumi Fukunishi, Yoshiaki Mikami (JBIRC-AIST)
Nobutoshi Ito, Kaori Kobayashi, Kengo Kinoshita (PDBj)

Institute for Protein Research, Osaka University