

The 1st France-Korea Particle Physics Laboratory Workshop, Lyon

Status of Activities in KISTI within the WISDOM collaboration

Soonwook Hwang, Jincheol Kim, Sehoon Lee KISTI

On behalf of the WISDOM collaboration

















KISTI's Achievements within the WISDOM project in 2008

- Completed development of a prototype version of DrugScreener-G software suite
- Performed a part of the data challenge for human diabetes type II in collaboration with LPC/IN2P3 (target protein: alpha-amylase)
- Improved the performance of WISDOM Production Environment by adopting AMGA to store production data
 - Also added more enhanced features, such as enhanced monitoring and bookkeeping, more fault-tolerant operation by CE, RB blacklisting, etc.
- Development of new enhanced features of AMGA in collaboration with the AMGA team of CERN IT Division and INFN-Catania team
 - Native SQL support
 - WS-DAIR support



Introduction to DrugScreener-G

 DrugScreener-G: Integrated Environment for Grid-enabled Large-scale Virtual Screening

😼 Drugscreener-G Product		_ _ N×	C Drugscreener-G Product
Elle Edit Window Drugscr			Elle Edit Window Drugscreener-G Editor-G Help
h · h · + ≽ II			2 - 2 - 1 + -> 🖬 📕 🥴 🖄
	🔮 Drugscreener-G Docking Editor 🕴	🗖 🗖 🅦 Ligand List 🖉 🗖	🔁 Project Exp 🛛 🖓 🗖 💽 😥 Drugscreener-G Docking Editor 😒 👘 🗖
a + a + + >		Ligand Dem 1 Ligand Rem 1 Ligand Rem2	Project Exp., 32 Ducking Duck
	Ceneral wayeer recent paymer processor and the paymer Ceneral wayeer recent paymer processor and the paymer Ceneral Properties No operations to display at this time.		



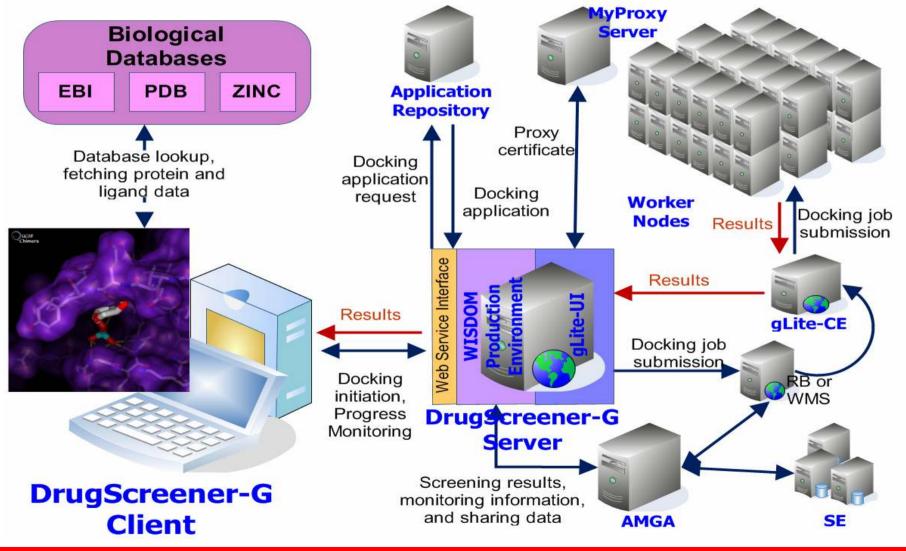
Motivation and Vision

- Motivation
 - Scientists and researchers in drug discovery are usually not familiar with new concepts and techniques of computing and information technology such as Grid computing.
 - New concepts and technical details to utilize Grid computing and infrastructure for large-scale virtual screening should be hidden in an integrated environment for Gridbased virtual screening.
 - To provide an integrated environment easy to integrate various virtual screening methods and extensible to adopt new virtual screening methods easily.
- Vision
 - Software platform to provide a user-friendly integrated environment for Gridenabled large-scale virtual screening for users without much knowledge of Grid computing to exploit full power of Grid computing infrastructure for drug discovery
- Target users: Bioinformaticians, Biologists, Drug Chemists

4



How does DrugScreener-G work with the EGEE infrastructure?



5



- DrugScreener-G Client
 - Docking project editing and management, monitoring, postprocessing and analysis
- DrugScreener-G Server
 - Communicates with a client by web service
 - Management of jobs and data on Grid, or local computing farm and storage devices
- Plug-in is the basic unit of functional components on both of client and server



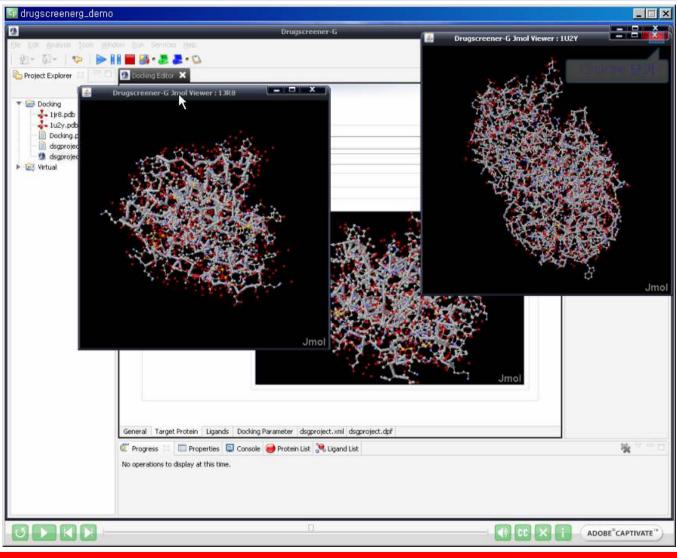
Notable features of DrugScreener-G – Accessing biological data

Congescreener-G Rioduct File Edit Window Drugscreener-G Editor-G Help	_0_	KI ∰ 200 Degen	Despacement-3 Web Service – WSDervich # E81 Despacement-3 can open the result as a Resource. Press Finish to open the Me.	
[3 + 2 +] +] +] → H ■ S @ D		🔗 🔗 👪 📻	Fieranne (joilt 102) pdb	
E Project Exp	🗖 🗖 🎘 Ligand List 🔗 🗖	Martin Sector C. Desport.	Preview:	
Docking degroject.mil degroject.mil	y (in example y (in example	82 305 Pack Mark Mark Sch 27 82 305 Pack Mark Pack Mark 83 305 Pack Mark Pack Mark 93 305 Pack Mark Pack Mark 94 305 Pack Mark Pack Mark 95	A Construction of the second s	
K Back Mexit>	200 💭 Tao Entropadoro Diagonaceles Entropa environ 201970: International Internation International International			
General ranger and same occurry and compression	Image: Control of the section of	December 2		
No operations to display at this time,				·····································
	Target proble ID : ["uby	Provise Presia		an 1993 (2 · A 漢칭 민 7 · · · · · · · · · · · · · · · · · ·
	Froein Massication E 21 Al 9 Froein data Joennaac Froein data Joennaac Froein data Joennaac Froein data Joennaac	월 (1539 PD) Shazar Sightor - Marson Hamme Dupleer 1940 - 전신은 고려있는 영국가 2013 - 2014 - 전 북성		
		* 1 월 월 월 월 월 월 월 월 월 월 월 월 월 월 월 월 월 월		
	2011. Tool Adulta Shar	Help] Strachaw Sammary [Bology & Sherristry] Platestals & resthode] Seg	sence Details' Geometry	
		Download Felex Learn more: (n Point Sequence Door 10 2210/yd/ c3/4 Point Sequence Control 2010 2010/yd/ c3/4	55 Eiological Molecule	
	Maran ar Vill Internat S (utfield	Display hiles In situ extension as an approach far Display holecule Title Title Title Title Title	. 🔊	
	Dange Die Beging Dange Die Beging Dange Die Beging	Decema Like Incorrect Like Nemao, S., IJ, C., Domager, I., Wredning Structure Analysis Authors T.M. Beaum, A., Owerell, C.M. Braver,	No	
	Properties No operations to display at this free,	Ninnao, S., Damager, T., Li, C., Wrodniag, T.M., Beguni, A., Overall, C.M., Brayer, D. Utiline, S. C. Conto, S. C.		
	12 프린 1 dath 목서스 코 Docan	To size the 90 exception 100500001 52		
	Preventer and a second	click as a set of the intege. History Deposition 200–17-20 Reference 2004-09-07 Reportmental track XRAY DIFFRADTION case 3 [CD3]	Display Options	
	LightScribe EDracione Xming	Resolution R Volue R Free Space Group		
	Legisch Results	Cost 4-192 5000 100.0	QuLAPOB All langues * Españle ordinologie filling al valocator	
	1923 0000	Unit Cell Linger 3 33.00 E 53.90 E 132- Augles (* 133-00 E 53.00 E 53.		
	Margosot. 20030120 Turneler	2a		
	Protection Suckyard, Enteredd			
	19.44 🖓 🕐 🐾 " 👞 Dox — 17. (1.) 💭 DugSmenner, 🛛 🗙 instruktionner, i 🖻 Brugscreener		annos (201 A / S & (1) € (102 C)	3 95

France-Korea Particle Physics Laboratory Workshop 2009

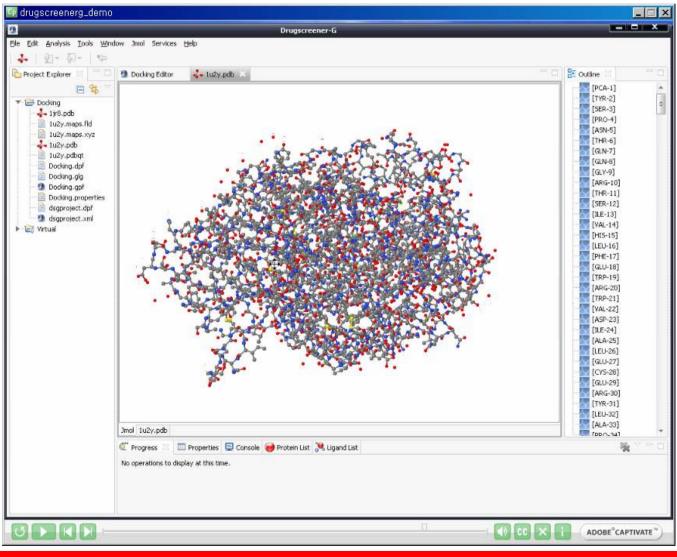


Notable features of DrugScreener-G – Protein visualization and modeling





Notable features of DrugScreener-G – Protein visualization and modeling





Notable features of DrugScreener-G – Ligand database management

🚾 drugscreenerg_demo			
2		Drugscreener-G	
Eile Edit Analysis Iools Wind	ow <u>R</u> un Services <u>H</u> elp		
1 월 - 월 - 🐤 🕨	1 1 3 • 2 2 • ©		
Project Explorer	Docking Editor ×		20 📴 Outine 🕄 📃 20
847	Ligands		An outline is not available.
Cocking properties Docking properties dsgproject.dpf dsgproject.xml EX Virtual	Ligand data Database location Data library	Test Library 2 Find	
	Ligand data list		
	General Target Protein Li	2 Luk 2 2 Luk 3 2 Luk 4 2 Luk 5 2 Luk 7 2 Luk 7 2 Luk 7 2 Luk 7 2 Luk 7 2 Luk 8 2 Luk 9 2 Luk 9 2 Luk 9 2 Luk 9 2 Luk 3 2 Luk 4 2 Luk 5 2 Luk 7 2 L	rary 2 List
	C Progress Properti	es 📮 Console 🥮 Protein List 🎘 Ligand List	1 T T T T T T T T T T T T T T T T T T T
	No operations to display at the		

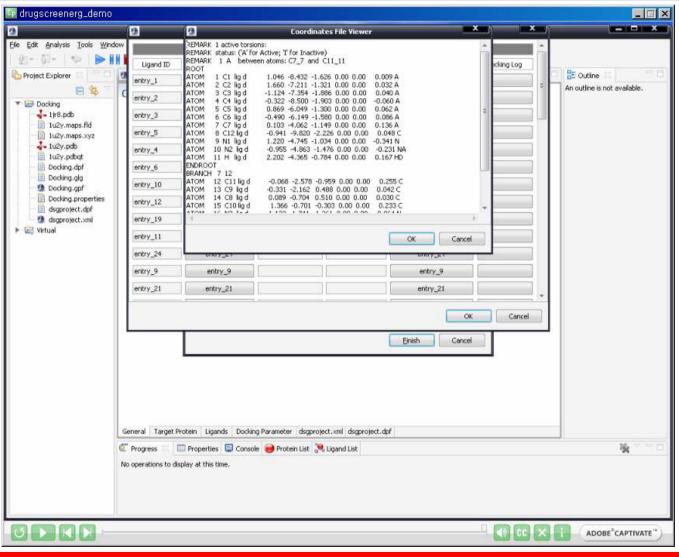


Notable features of DrugScreener-G – Progress monitoring

gscreenerg_den	no			
r daulusis Toola	Drugscreener-G Windom Bun Sanniss Helb			
9	Progress Report		×	
Project Performance	e Monitoring			1
	i6eb316-ab09-4484-a9b7-b20b40eba03d_150.183.250.204_Docking_Docking			
Total Progress	lob Status			-
Job Information	Select the Job			
Total Jobs : 1				
Job not Done : 1	.00			
Progress			_	
			_	
			Refresh	
			Refresh	
]
	General Target Protein Ligands Docking Parameter dsgproject.xml dsgproject.dpf			
	General Target Protein Ligands Docking Parameter disproject.xml disproject.dpf C Progress III Properties III Console 📦 Protein List IIII Ligand List			
			Cancel	
	🖉 Progress 🔅 🔲 Properties 📮 Console 🥪 Protein List 💸 Ligand List		Cancel	
	🖉 Progress 🔅 🔲 Properties 📮 Console 🥪 Protein List 💸 Ligand List	0	Cancel	
	🖉 Progress 🔅 🔲 Properties 📮 Console 🥪 Protein List 💸 Ligand List		Cancel	



Notable features of DrugScreener-G – Screening result view





Future work: more convenient features and tools for drug discovery users

- Support for R language for post-processing and analysis
 - Summary table of screening results as a data frame in R
- Workflow support
 - Motuer workflow engine for EGEE
- Consensus docking support
- Multi-step docking support with workflows
- Multiplatform support for Grid computing (Long Term Goal)
 - Based on the EGEE-DEISA interoperability model
 - Plans to develop resource selector of DrugScreener-G server for:
 - ✓ Globus Toolkit 4.0
 - ✓ UNICORE



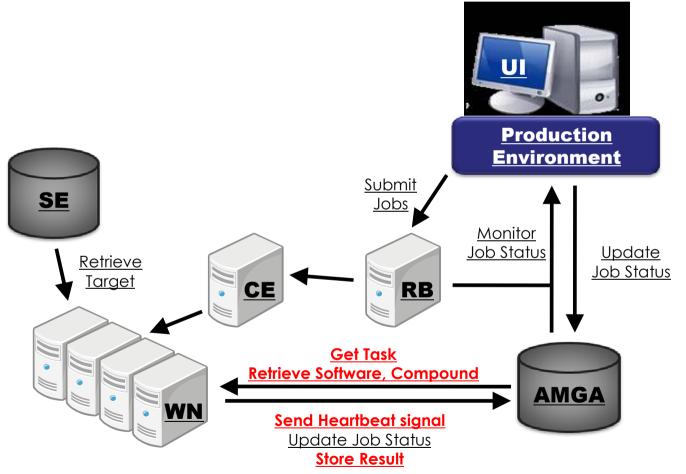
Future Work: Use of Molecular Dynamics

- Molecular Dynamics as a virtual screening tool
 - Screening drug candidates considering flexible conformation of proteins and compounds in solution with MD
 - Generating protein conformations with MD for flexible docking (Ensemble-based virtual screening with relaxed complex scheme)
- New developments for integration of Molecular Dynamics (future work)
 - User-friendly GUI for input file generation
 - Job management engine for Molecular Dynamics (as a plug-in for both of client and server)
 - Well-known molecular dynamics application support (CHARMM, AMBER, NAMD, etc.)



Performance enhancment of the WISDOM Production Environment

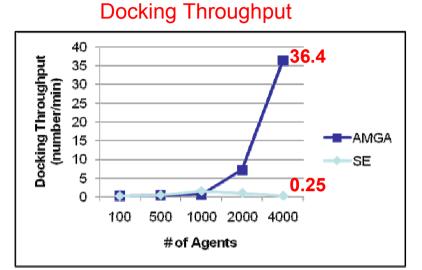
Use of AMGA for task distribution and file store



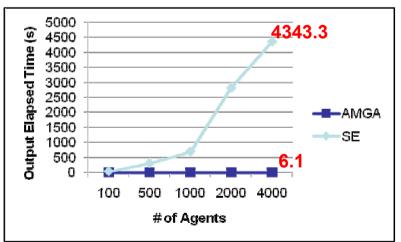


The exploitation of AMGA for storing ligands and docking results

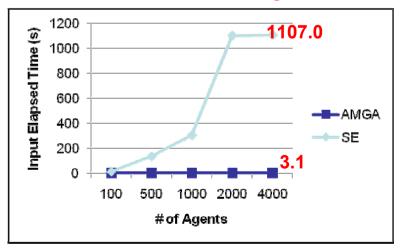
- Why is AMGA used for the input/output data storing purpose?
 - AMGA performs way much better than SE in terms of throughput and latency







Time taken to retrieve a ligand



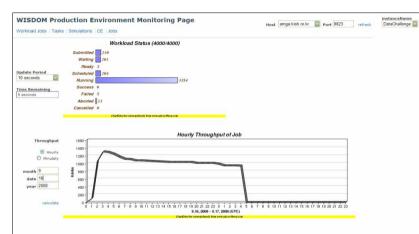


Additional Features to the Improvement of WISDOM PE by KISTI

- Flexible fault tolerance
 - Can recover a previous instance using data on AMGA
 - Can increase/decrease the number of agents of a running previous instance
- Improved Job distribution efficiency
 - Dynamic management of the list of available RB and CE
 - While the initial job submission keeps going, we can check the status of submitted jobs and resubmit failed jobs
 - Direct collecting the statistical data of sites from WN
 - Its own ranking based on previous statistical data
 - Timeout feature (submission timeout, heartbeat timeout)
- Enhanced monitoring and bookkeeping



Enhanced monitoring and bookkeeping



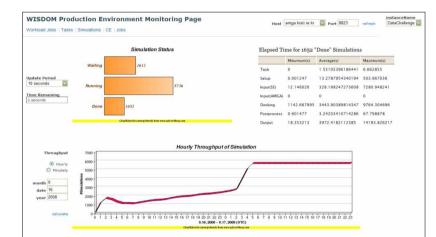
Workload status and Job throughput monitoring

About Us (Site Map (Privacy Policy) Contact Us (02008 Korea Institute of Science and Tec

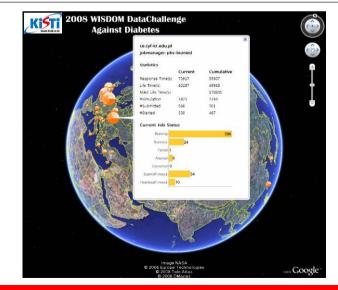
	ID	Name	Total CPU	Free CPU	Waiting Jobs	Running Jobs	Response Time	Life Time	#Simulations	#Submitted	#Started	Score
	TOT	TAL	62376	22537	12964	10995			246151	153395	43824	
	98	ce.gina.sara.nl jobmanager-pbs-medium	496	43	0	581	6027	50285	9777 (82)	1510 (500)	1232 (419)	9777
	17	ce.cyf-kr.edu.pl jobmanager-pbs-biomed	552	19	4	170	27991	49948	9527 (32)	1490 (474)	1018 (238)	9527
	86	lcgce02.gridpp.rl.ac.uk jobmanager-lcgpbs-grid500M	2583	1585	11	380	2182	50236	8806 (26)	1472 (457)	1419 (456)	8806
	128	ce2.ppgrid1.rhul.ac.uk jobmanager-pbs-biomed	400	1	165	0	9361	49480	6352 (0)	1541 (454)	444 (0)	6352
Update Period 60 seconds	15	gridgate.cs.tcd.ie jobmanager-pbs-egee	691	156	123	437	8611	54533	6082 (36)	1427 (449)	900 (181)	6082
Time Remaining	85	<u>ce.reef.man.poznan.pl</u> jobmanager-pbs-biomed	552	258	0	0	24133	37798	5962 (0)	1461 (452)	349 (13)	5962
efresh	79	gazon nikhef nl jobmanager-pbs-qlong	1152	1	12	385	9710	37822	5757 (79)	1443 (488)	863 (377)	5757
	144	serv03.hep.phy.cam.ac.uk jobmanager-lcgcondor-biomed	152	0	6	143	22011	26739	4993 (64)	1190 (453)	633 (140)	4993
	80	trekker.nikhef.nl jobmanager-pbs-qlong	1152	0	12	385	13784	42626	4988 (37)	1481 (461)	768 (275)	4988
	103	w-ce02.grid.sinica.edu.tw jobmanager-kgpbs-biomed	1480	156	118	100	18367	55944	4791 (38)	1476 (466)	414 (38)	4791
	126	ce01.esc.gmul.ac.uk jobmanager-kgsge-kg_long_x86	1674	801	0	318	3810	50104	4291 (18)	1268	1107	4291

Job distribution and execution status with CE info

HOULUS | She Map | Hiwacy Houcy | Contact US | VICUUS Notes installate of collecte and recting



Status summery and simulation throughput monitoring





Data challenge against Diabetes Type II with the improved WISDOM PE

- Data challenge to Human pancreatic amylase inhibitor (1u2y), a target protein of diabetes type 2, with 308310 chemical compounds
- Achieved significant improvement in efficiency and throughput

	WISDOM-II	DIANE	KISTI
Total number of dockings	156,407,400	308,585	308,310
Estimated duration on 1 CPU	413 years	16.7 years	39.0 years
Duration of the experiment	76 days	30 days	2.4 days
Cumulative number of Grid jobs	77,504	2,580	103,583
Maximum number of concurrent CPUs	5000	240	7,370
Number of used Computing Elements	98	36	127
Crunching Factor	1983	203	5937
Distribution Efficiency	39%	84%	81%



KISTI's contribution to AMGA included in a new version of AMGA

- AMGA1.9 released on Oct. 6, 2008
 - Support for the import of existing relational tables
 - Native SQL support
 - Multi-threaded Server
- Currently preparing AMGA 2.0 release
 - To be released by the end of this year
 - WS-DAIR support
 - Stability of WS-DAIR server needs to be improved before release



- KISTI's achievements in FKPPL in 2008
 - Completed development of a prototype version of DrugScreener-G software suite
 - Performed a part of the data challenge for human diabetes type II in collaboration with LPC/IN2P3 (target protein: alpha-amylase)
 - Improved the performance of WISDOM Production Environment by adopting AMGA to store production data
 - Also added more enhanced features, such as enhanced monitoring and bookkeeping, more fault-tolerant operation by CE, RB blacklisting, etc.
 - Development of new enhanced features of AMGA in collaboration with the AMGA team of CERN IT Division
 - ✓ Native SQL support
 - ✓ WS-DAIR support



Thank you for your attention!

Soonwook Hwang Senior Researcher, Korea Institute of Science and Technology Information <u>hwang@kisti.re.kr</u> (Soonwook Hwang)



Questions?





Backup slides