



Molecular dynamics applications

Presented by

Vinod Kasam

BioinfoGRID workshop, April 2007, Geneva

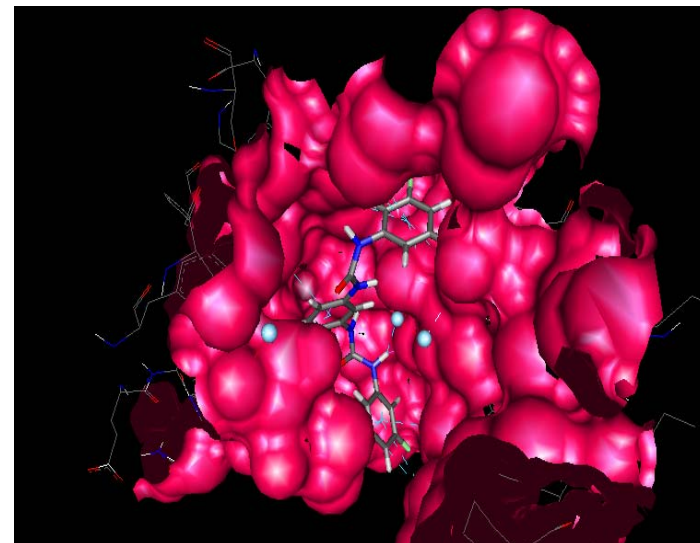




- Molecular dynamics
- Computational grids
- Amber software
- Production environment
- Deployment strategy
- Initial results
- Issues
- Future steps



- Molecular Dynamics = computation of the motion of atoms within a molecular system using molecular mechanics
- Molecular Dynamics is commonly used for drug design and drug discovery
 - Molecular modelling of drugs
 - Measurement of binding energies between ligands and biological targets
- Grids offer promising perspectives for in silico drug discovery
 - Identification of drug candidates using computing tools
 - Virtual screening = rapid assessment of large libraries of chemical structures in order to guide the selection of likely drug candidates





Criteria for MD software selection

- Criteria for MD software selection
 - Local expertise
 - Type of simulations to perform (protein folding or re-ranking after virtual screening, etc)
 - Licensing issues
 - Choice of infrastructure (Linux/Unix/Windows)
 - Robustness, reliability and available literature
- Motivations for the choice of Amber9
 - Very well suited for protein ligand complex simulations
 - Deployed on many operating systems including linux
 - Very well known and fully documented MD software
 - Running collaboration with groups at Univ. Modena and SCAI Fraunhofer experienced with Amber
 - **Issue: licensed software**
- Other MD software considered: Gromacs,



Criteria for grid infrastructure selection

- Grids of supercomputers and grids of PCs are relevant environment for MD computations
 - Grids of PCs have limited memory (2GB today on EGEE): limitation for large molecules and memory greedy software
 - Supercomputers are harder to access
- Criteria for selection
 - Local expertise
 - Type of simulations to perform (memory requirements, CPU time)
 - Accessibility
 - Production environment
- Infrastructure selected: EGEE
 - Easy access to resources
 - No memory limitation
 - Previous experience and existing production environment



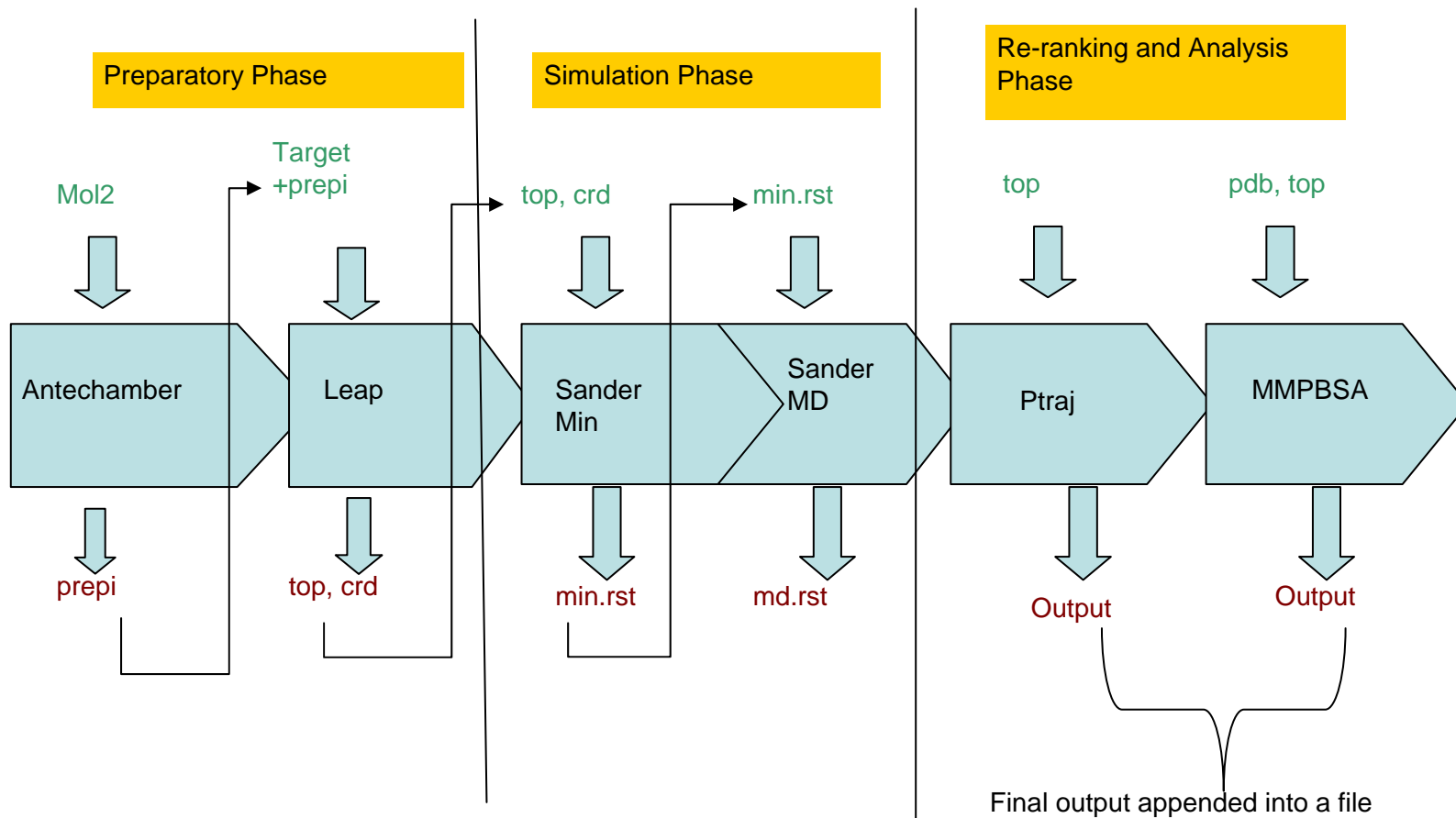
- New large scale docking deployments organized by WISDOM collaboration in 2006

Two major initiatives

- Data challenge on Plasmepsin (July 2005 -August 2005)
 - Second data challenge on malaria (October 2006 – January 2007)
-
- Production of real data within the framework of WISDOM

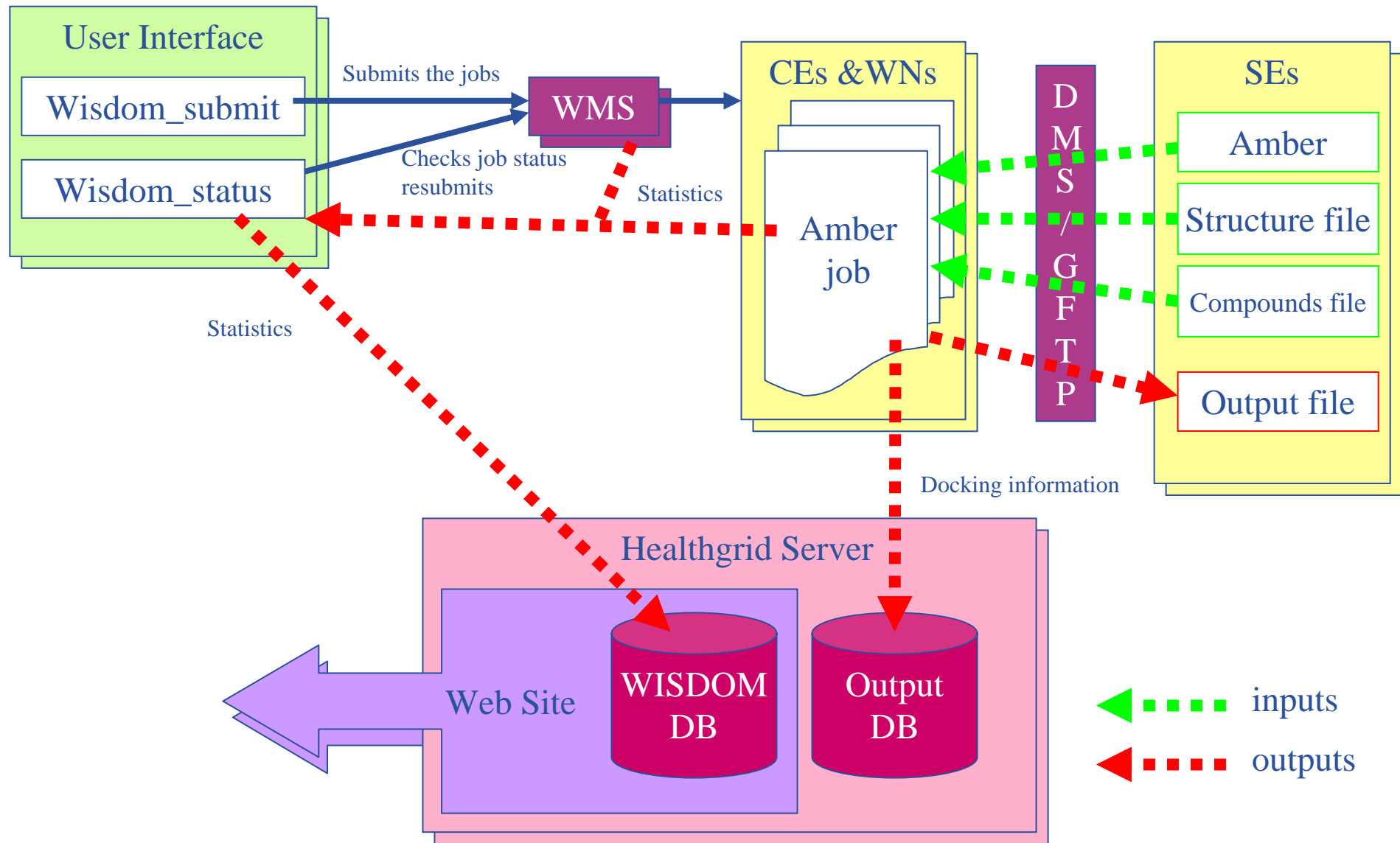


Amber software architecture





- MM-PBSA is a program which is used to estimate energies and entropies from the snapshots contained within trajectory files.
- The calculations are organized and spawned by a Perl script
- mm-pbsa, collects statistics and formats the output in tabular form.
- The analysis is primarily based on continuum solvation models.





- Deployment is segmented into 3 phases
 - Store all the necessary input on the Storage Element of the grid
 - 5000 chemical compound multiframe made into 100 fragments
 - Target (1LEE)
 - Bash Script to run the Amber job
 - Run the jobs
 - Running the jobs on Clermont cluster by using WISDOM production environment
 - Checking the status of the jobs
 - Done automatically (developed for WISDOM data challenges)



- Licensing
 - Amber is commercial software with an academic license for computers, like a cluster, of an institute.
 - The license is integrated in the software package (so no license server required)
 - Amber can be easily used on the EGEE grid
- Issues
 - One license per grid site deploying Amber
 - Use of Amber restricted to grid users coming from institutes owning a license
- Issue addressed
 - Access granted to all the nodes where Amber is installed



Microsoft Excel - Table_24_complexes_2

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

Tapez une question

A1 Analyse of MMPBSA results

Analysis of MMPBSA results

		From Pbtot and Gbtot				H-bonds Interactions						
		From Pbtot		From Gbtot								
File Name	RMS	PbTot	GbTot			ASH 36	ASP 216	VAL 80	SER 81	SER 220	PHE 296	TYR 1
1	test2_0162	1.47	-32.11	-54.51	No	LIG N3-ASP OD1 LIG N5-ASP OD1	Yes	Yes	Yes	No	Yes	
2	test2_1511	1,18	-31,68	-49,72	LIG N1-ASH OD2	LIG N1-ASP OD1 LIG N2-ASP OD1	Yes	Yes	Yes	No	No	
3	test2_1866	1,6	-32,47	-50,41	LIG N2 ASHOD1	LIG N1 ASP OD1 LIG N1 ASP OD2	Yes	Yes	Yes	No	No	
4	test2_1941	1,75	-33,17	-52,2	LIG N3-ASH OD1 LIG N3-ASH OD2	LIG N4-ASP OD1 LIG N4-ASP OD2 LIG N1-ASP OD1	Yes	Yes	Yes	No	No	
5	test2_2020	2,01	-33,57	-57,99	LIG N1-ASH OD1 LIG N2-ASH OD2	LIG N2-ASP OD1	Yes	Yes	Yes	No	No	

20 selected / Feu1 / Feu2 / Feu3

Prêt

démarrer | Inbox for kas... | ACM: Copyri... | 2 Windows... | 2 Microsoft... | 3 SSH Sec... | X ~ | X ImageMagick... | Microsoft Exc... | FR | 08:56



Initial results

BioinfoGRID

Microsoft Excel - Table_24_complexes_2

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

Tapez une question

Répondre en incluant des modifications... Terminer la révision...

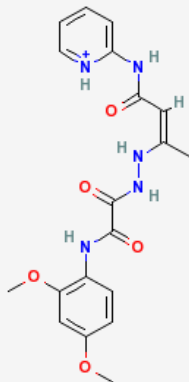
A1 Analyse of MMPBSA results

	A	B	C	D	E	F	G	H	I	J	K	L
16	6	test2_2062	1,36	-31,64	-57,6	LIG N2-ASH OD1	LIG N1-ASP OD1	Yes	Yes	No	No	No
17	7	test2_2339	2,03	-31,6	-47,36	LIG N1-ASH OD1 LIG N2-ASH OD1	LIG N3-ASP OD1 LIG N4-ASP OD1	Yes	Yes	No	No	No
18	8	test2_3293	1,76	-32,73	-51,57	ASH OD2-LIG N2	LIG N3-ASP OD1 LIG N3-ASP OD2	Yes	Yes	Yes	No	No
19	9	test2_3759	1,97	-29	-66,37	LIG N5-ASH OD1	LIG N2-ASP OD1 LIG N3-ASP OD1	Yes	Yes	No	No	No
20	10	test2_3762	1,62	-22,19	-59,4	LIG N5-ASH OD1	LIG N3-ASP OD1	No	Yes	Yes	No	No
21	11	test2_3781	1,18	-11,02	-57,66	LIG N1-ASH OD1 LIGN1-ASH OD2 ASH OD2-LIG O5	LIG N1-ASP OD1	Yes	Yes	No	No	Yes

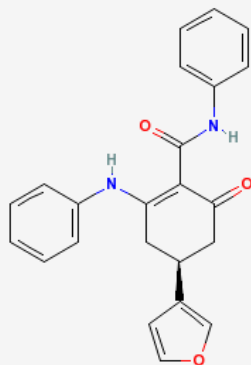
20 selected / Feu1 / Feu2 / Feu3

Prêt

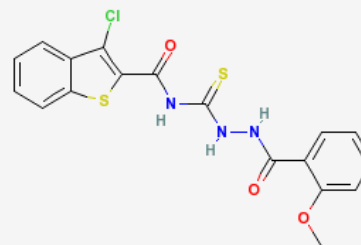
démarrer | Inbox for kasam... | ACM: Copyright... | 2 Windows M... | 2 Microsoft Of... | 3 SSH Secure ... | 2 XWin | Microsoft Excel - ... | FR | 08:57



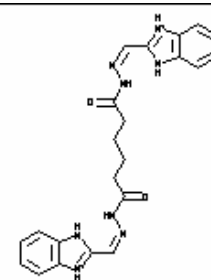
Result_0162



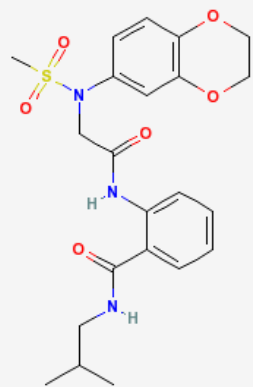
Result_1511



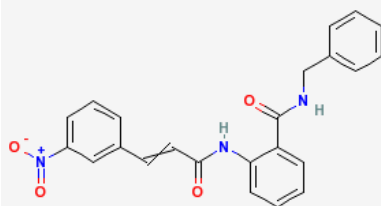
Result_1866



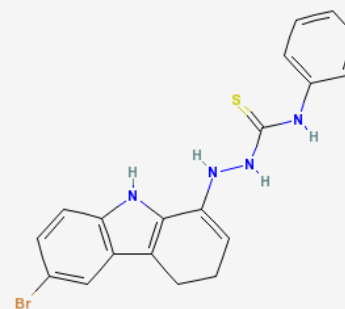
Result_1941



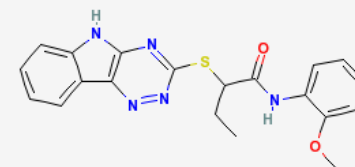
Result_2020



Result_2062



Result_2339

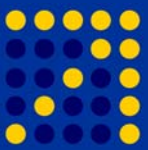


Result_3293



Achievements so far and Future steps

- Achievements
 - Successfully deployed Amber software on grid infrastructure
 - The best 5000 compounds coming from WISDOM-I are tested
- Future steps
 - Deployment of the same procedure at other grid nodes
 - Fine tuning of the parameters
 - Simulations with full protein flexibility
 - The bests compounds from WISDOM-II will be refined by using the same procedure
 - Finally selected compounds will be tested in the experimental laboratories



- *BIOINFOGRID* <http://www.bioinfoGRID.eu>
- Giulio Rastelli (Univ. Modena)
- Ana Ferrari (Univ. Modena)
- Jean Salzemann
- Astrid Maass (Scai-Fraunhofer)
- Nicolas Jacq
- Vincent Breton
- Ana Da Costa
- and the WISDOM collaboration