BioinfoGRID
Bioinformatics Grid Application for life science

Results of WISDOM and Deployment of Molecular Dynamics on EGEE infrastructure

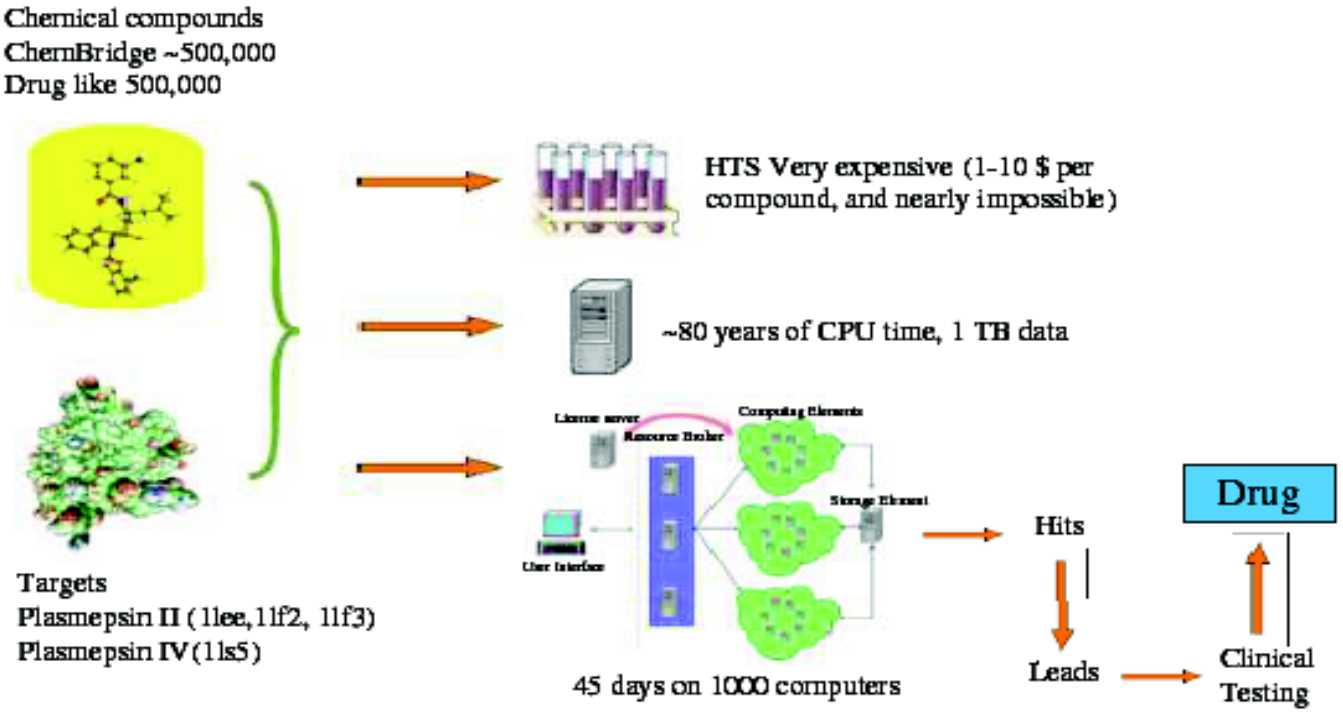
Vinod Kasam

CNRS - IN2P3

Information Society and Media

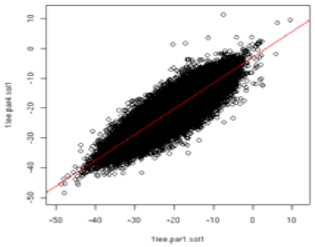
WISDOM

- Random screening approach
- First high throughput virtual screening approach on EGEE grid
- 500,000 compounds, 5+3 targets, 2 different software
- 1 TB of output
- Several strategies for result analysis
- Identification of three scaffolds, one is novel

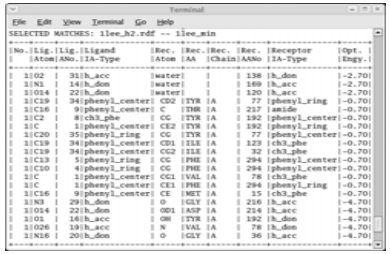


Strategies in result analysis

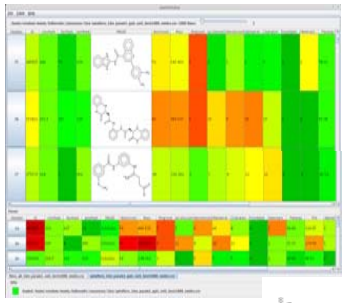
- Results based on Scoring



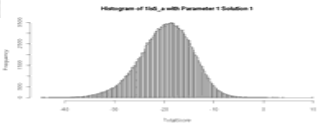
- Results based on match information

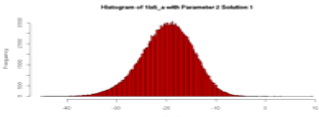


- Results based on consensus scoring

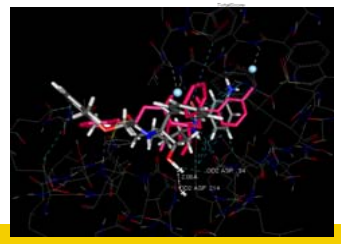


- Results based on different parameter settings



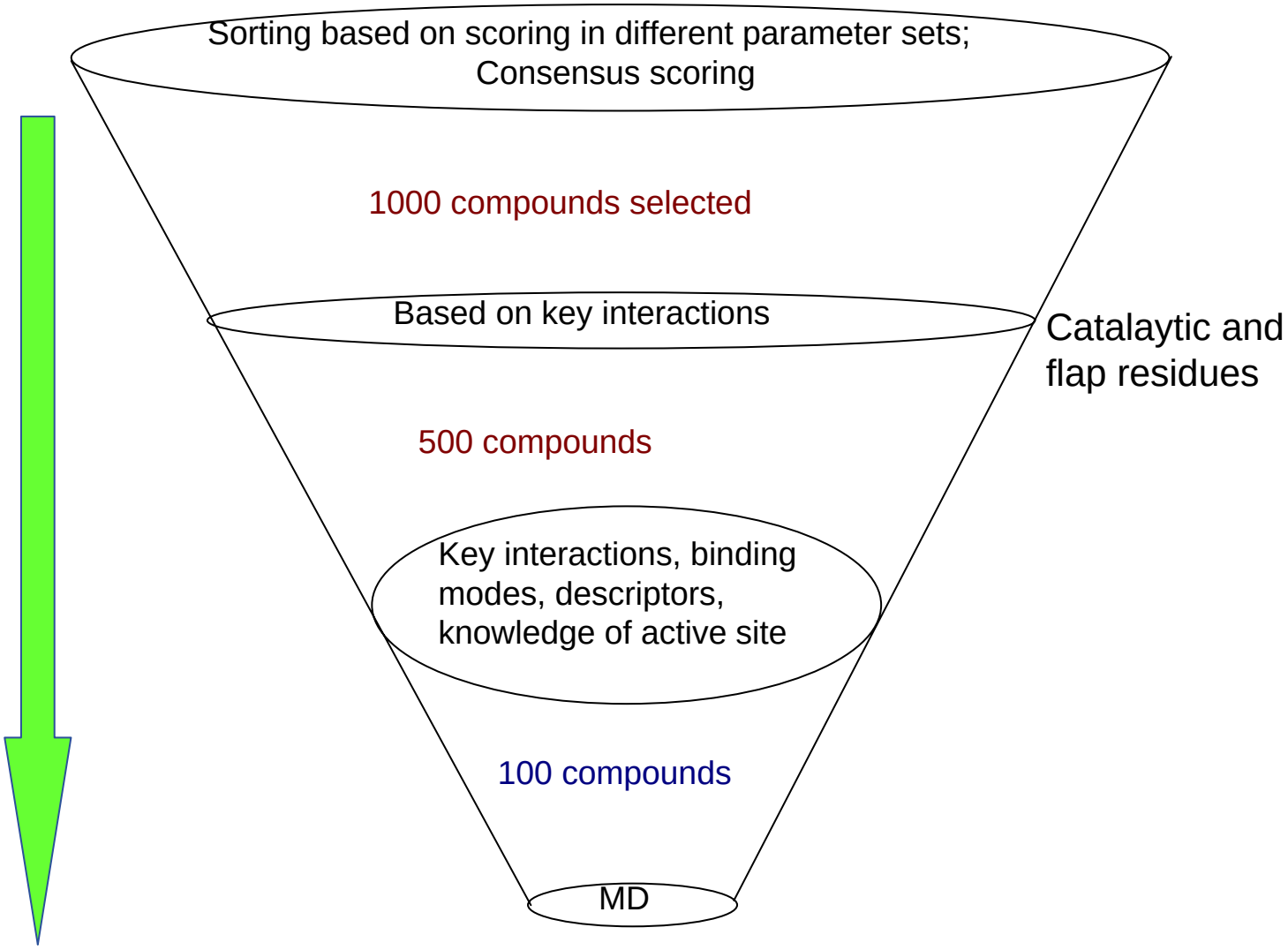


- Results based on knowledge on binding site



Filtering process

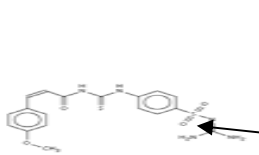
500, 000 chemical compounds



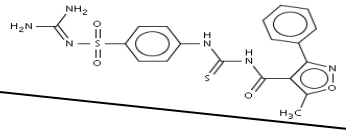
50 compounds to be tested in experimental lab

Top 10 compounds by scoring in Parameter 1

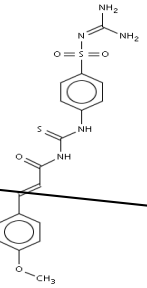
BioinfoGRID



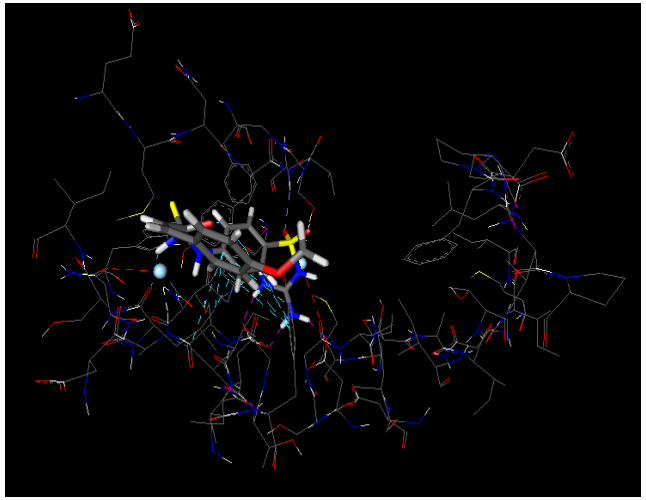
1. WISDOM-490500



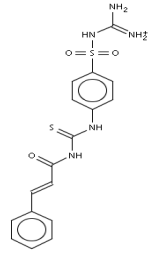
2. WISDOM-491901



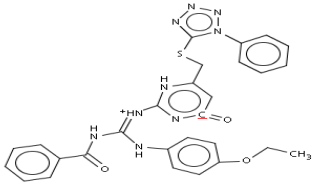
3. WISDOM-490515



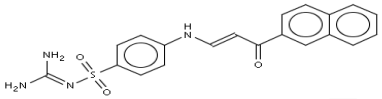
Top scoring but poor binding mode



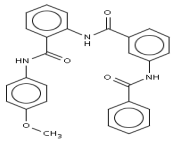
4. WISDOM-490514



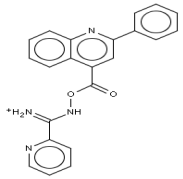
5. WISDOM-462271



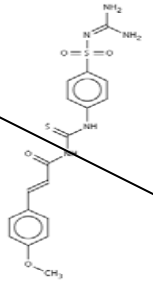
6. WISDOM-235118



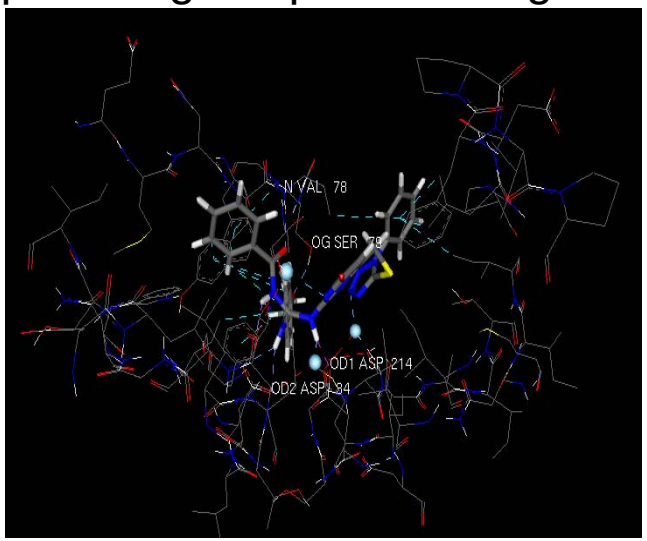
7. WISDOM-278345



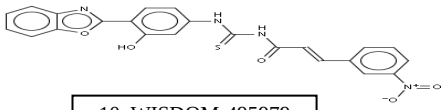
8. WISDOM-360604



9. WISDOM-490502

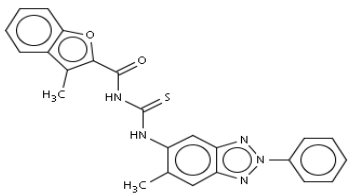


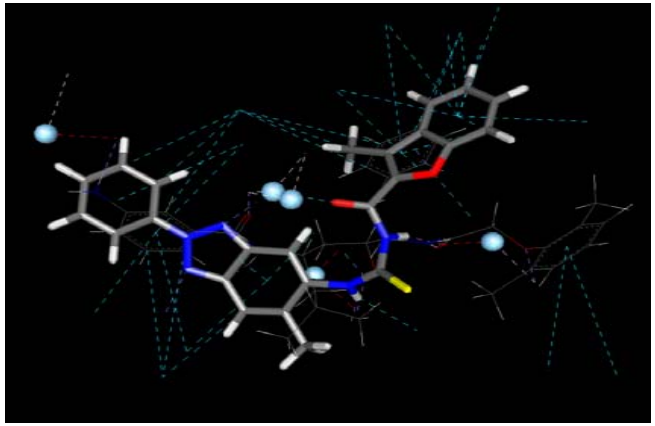
Top scoring, good binding mode, interactions to key residues

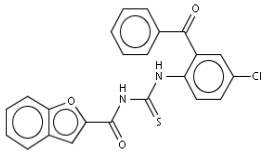


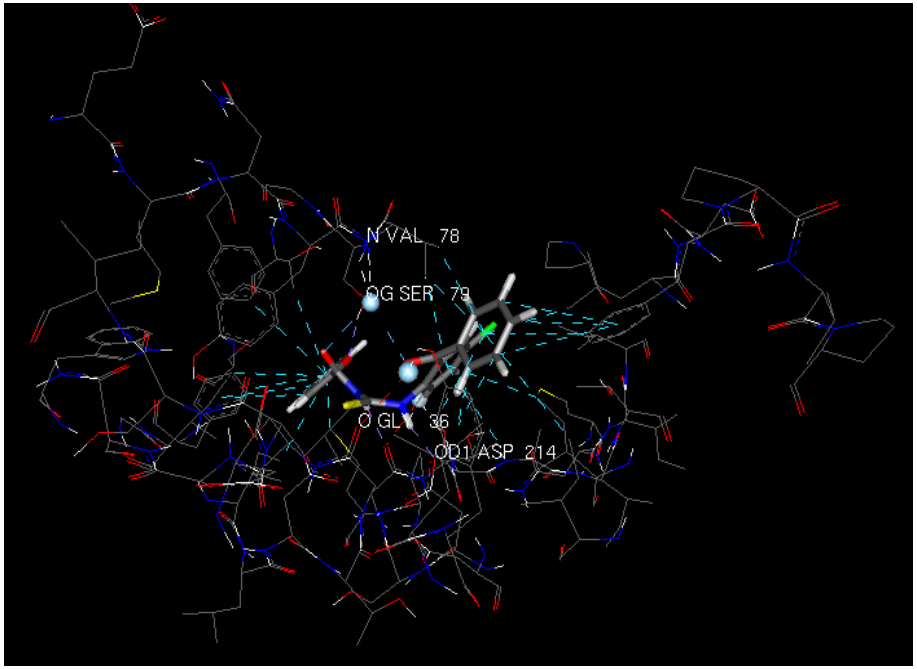
10. WISDOM-495979

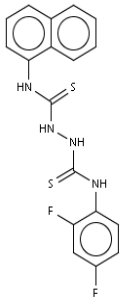
Compounds for MD - Thiourea compounds

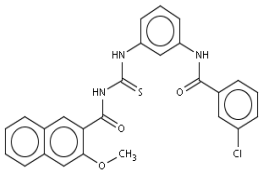




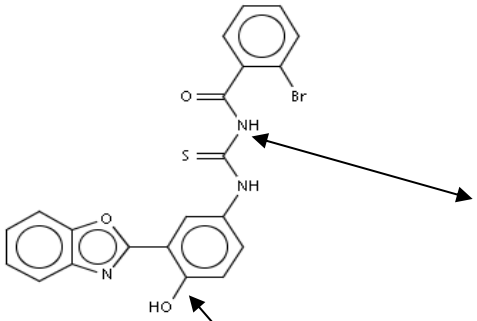


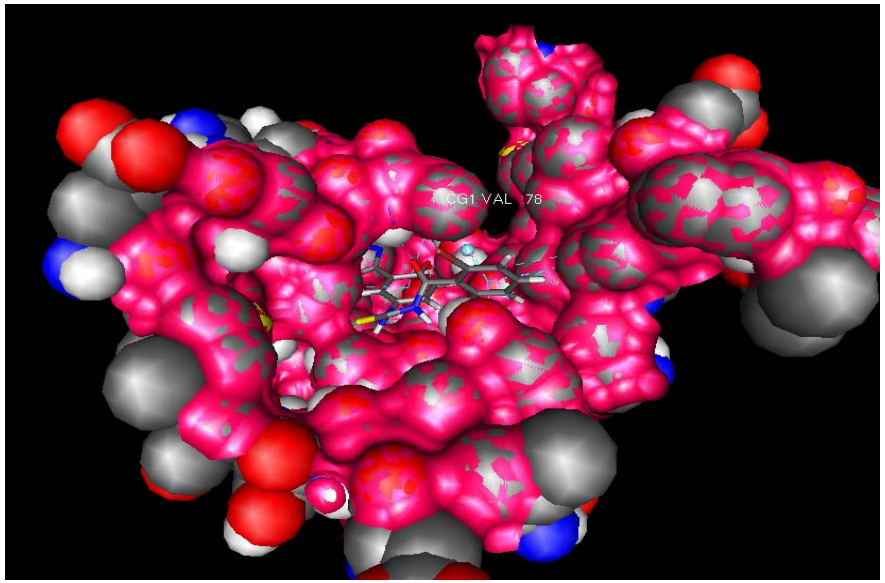


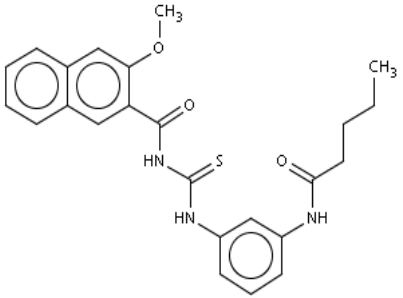




Compounds for MD - Thiourea compounds





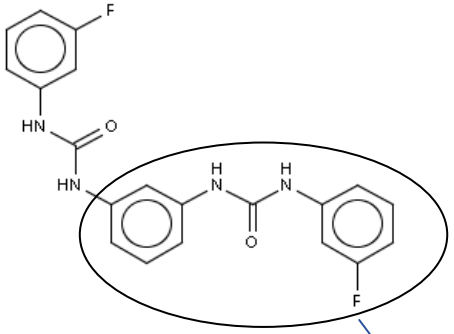


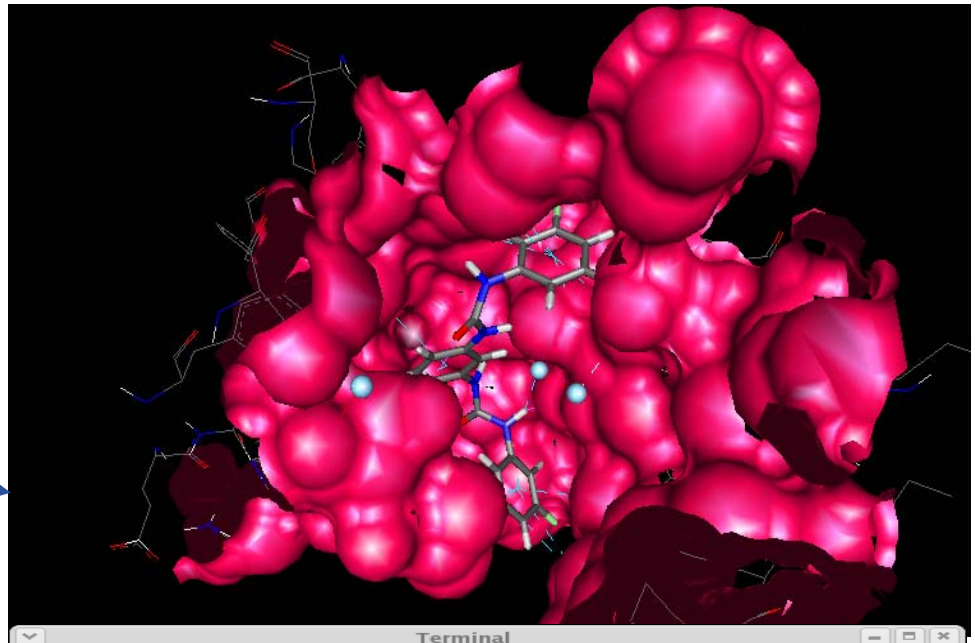
Terminal

File Edit View Terminal Tabs Help

No.	Lig. Atom	Lig. ANo.	Ligand IA-Type	Rec. Atom	Rec. AA	Rec. Chain	Rec. AANo	Receptor IA-Type
1	C19	24	phenyl_ring	CG	PHE	A	120	phenyl_center
1	C16	21	phenyl_center	CD1	ILE	A	32	ch3_phe
1	C16	21	phenyl_center	CE	MET	A	15	ch3_phe
1	C9	13	phenyl_center	CG2	VAL	A	78	ch3_phe
1	C1	1	phenyl_ring	CG	TYR	A	192	phenyl_center
1	C3	3	phenyl_ring	CG	PHE	A	294	phenyl_center
1	C2	2	phenyl_ring	CG	PHE	A	294	phenyl_center
1	C1	1	phenyl_center	CE1	PHE	A	294	phenyl_ring
1	C1	1	phenyl_center	CG2	VAL	A	78	ch3_phe
1	C1	1	phenyl_center	CE2	TYR	A	192	phenyl_ring
1	C1	1	phenyl_center	CD1	ILE	A	300	ch3_phe
1	C1	1	phenyl_center	CG1	VAL	A	78	ch3_phe
1	O1	8	h_acc	N	VAL	A	78	h_don
1	N1	9	h_don	O	GLY	A	36	h_acc
1	N2	12	h_don	OD1	ASP	A	34	h_acc
1	C20	25	phenyl_ring	CG	PHE	A	120	phenyl_center
1	C16	21	phenyl_center	CZ	PHE	A	120	phenyl_ring
1	C16	21	phenyl_center	CE1	PHE	A	120	phenyl_ring
1	C16	21	phenyl_center	CG2	THR	A	114	ch3_phe
1	C16	21	phenyl_center	CZ	PHE	A	111	phenyl_ring
1	C16	21	phenyl_center	CD1	ILE	A	123	ch3_phe
1	N3	20	h_acc	OG	SER	A	79	h_don
1	C9	13	phenyl_center	CD2	TYR	A	77	phenyl_ring
1	O3	28	h_don	O	GLY	A	216	h_acc
1	C14	18	phenyl_ring	CG	TYR	A	77	phenyl_center

Compounds for MD - Urea compounds





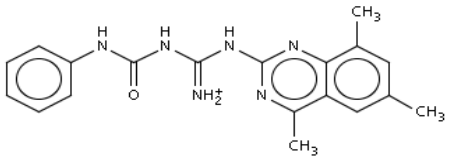
Terminal

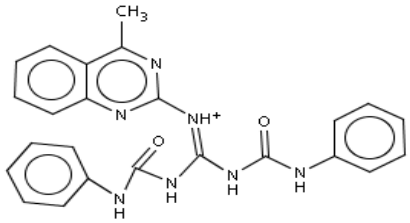
File Edit View Terminal Tabs Help

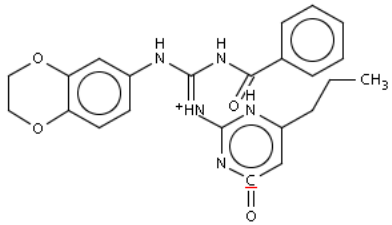
No.	Lig. Atom	Lig. ANo.	Ligand IA-Type	Rec. Atom	Rec. AA	Rec. Chain	Rec. AANo	Receptor IA-Type
1	N4	21	h_don	water			120	h_acc
1	C18	25	phenyl_ring	CG	PHE	A	294	phenyl_center
1	C15	22	phenyl_center	CE1	PHE	A	294	phenyl_ring
1	C15	22	phenyl_center	CG2	VAL	A	78	ch3_phe
1	C8	11	phenyl_center	C	THR	A	217	amide
1	C8	11	phenyl_center	C	GLY	A	216	amide
1	C8	11	phenyl_center	CD1	ILE	A	32	ch3_phe
1	C8	11	phenyl_center	CG2	ILE	A	32	ch3_phe
1	C8	11	phenyl_center	CE	MET	A	15	ch3_phe
1	O1	9	h_acc	OG	SER	A	79	h_don
1	N1	7	h_don	O	GLY	A	216	h_acc
1	C2	2	phenyl_ring	CG	TYR	A	77	phenyl_center
1	C1	1	phenyl_center	CD1	ILE	A	123	ch3_phe
1	C1	1	phenyl_center	CD2	TYR	A	77	phenyl_ring
1	C1	1	phenyl_ring	CG	TYR	A	77	phenyl_center
1	N3	18	h_don	OD2	ASP	A	34	h_acc
1	N3	18	h_don	OD1	ASP	A	34	h_acc
1	C15	22	phenyl_center	CE2	TYR	A	192	phenyl_ring
1	C15	22	phenyl_center	CG1	VAL	A	78	ch3_phe
1	N4	21	h_don	OD1	ASP	A	214	h_acc
1	C20	27	phenyl_ring	CG	TYR	A	192	phenyl_center
1	C15	22	phenyl_center	CD1	ILE	A	300	ch3_phe

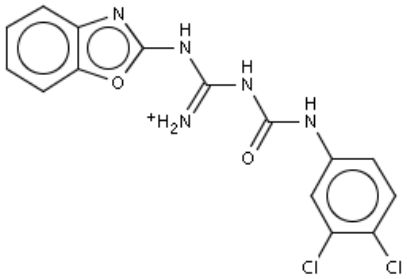
Note: Diphenyl urea compounds are well in agreement with literature (Walter Reed compounds)

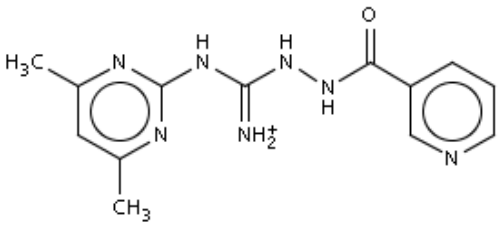
Compounds for MD- Guanidino compounds







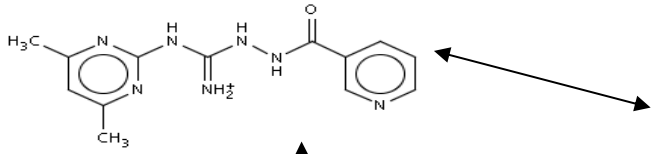


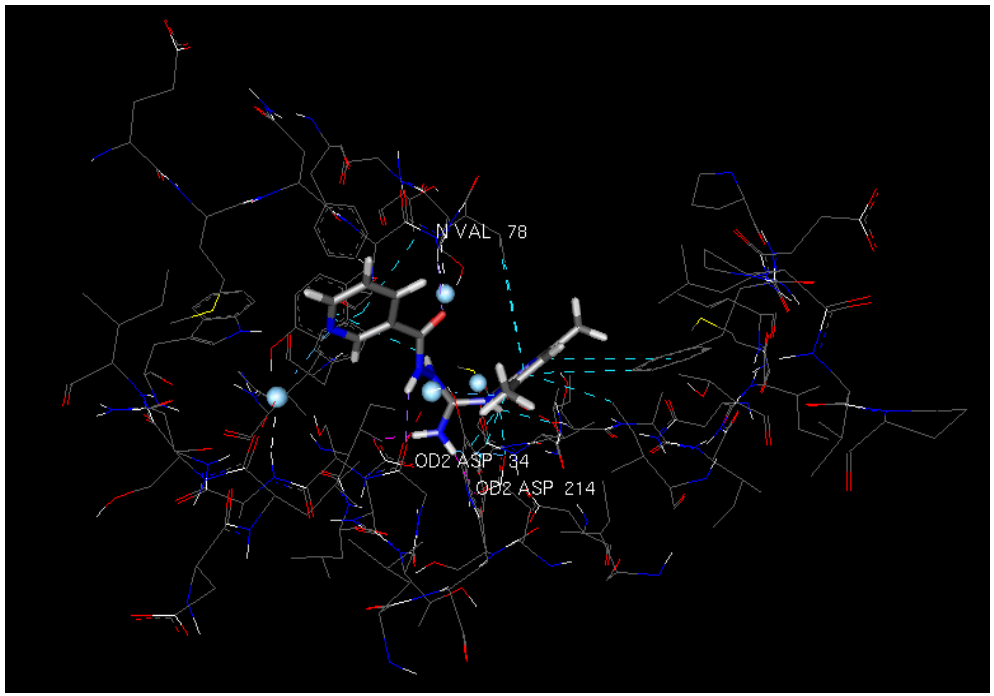


Note: Guanidino compounds are likely to be novel, so far, not identified as inhibitors for Plasmepsins.

Compounds for MD - Guanidino compounds

BioinfoGRID





Terminal

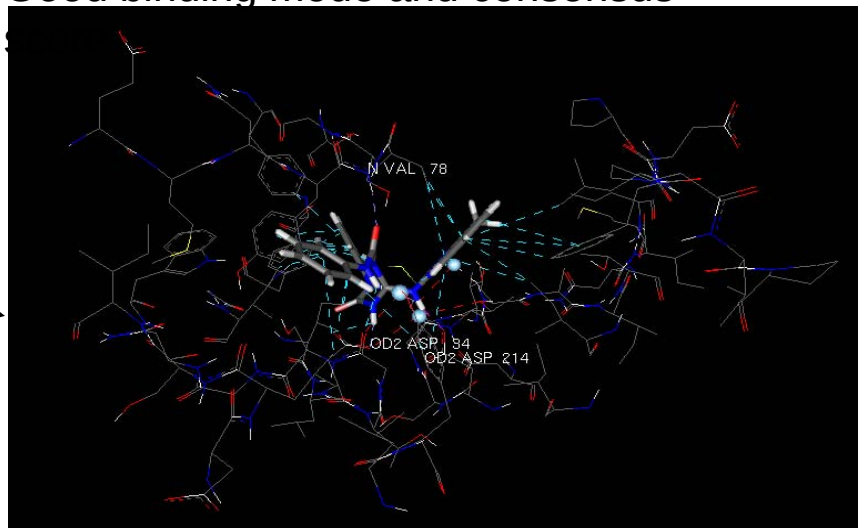
File Edit View Terminal Tabs Help

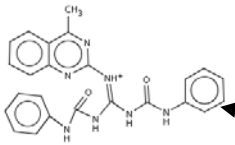
No.	Lig. Atom	Lig. ANo.	Ligand IA-Type	Rec. Atom	Rec. AA	Rec. Chain	Rec. AANo	Receptor IA-Type
1	N1	5	h_acc	water			58	h_don
1	N7	19	h_acc	water			39	h_don
1	N7	19	phenyl_center	C	TYR	A	77	amide
1	C7	13	amide	CG	TYR	A	77	phenyl_center
1	C13	21	ch3_phe	CG	TYR	A	192	phenyl_center
1	N1	5	phenyl_center	CE1	PHE	A	294	phenyl_ring
1	N1	5	phenyl_center	CG2	VAL	A	78	ch3_phe
1	N1	5	phenyl_center	CD1	ILE	A	300	ch3_phe
1	N1	5	phenyl_center	CE2	TYR	A	192	phenyl_ring
1	N1	5	phenyl_center	CG1	VAL	A	78	ch3_phe
1	C3	3	phenyl_ring	CG	PHE	A	294	phenyl_center
1	N3	8	h_don	OG1	THR	A	217	h_acc
1	N3	8	h_don	OD1	ASP	A	214	h_acc
1	N4	10	h_don	OD1	ASP	A	34	h_acc
1	N4	10	h_don	OD2	ASP	A	214	h_acc
1	N4	10	h_don	OD1	ASP	A	214	h_acc
1	C12	20	phenyl_ring	CG	TYR	A	77	phenyl_center
1	N7	19	phenyl_center	CD2	TYR	A	77	phenyl_ring
1	O1	14	h_acc	N	VAL	A	78	h_don
1	N6	12	h_don	O	GLY	A	36	h_acc

Note: Satisfied all criteria, good binding mode, interactions to key residues, good score, appropriate descriptors.

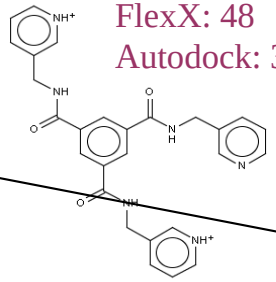
Compounds from consensus scoring

Good binding mode and consensus

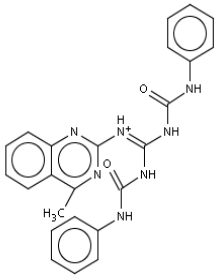




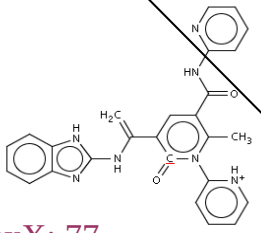
FlexX: 30
Autodock: 24



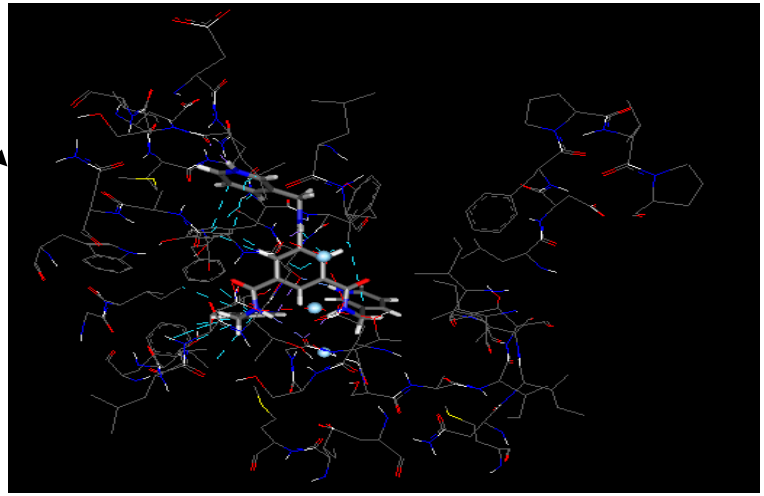
FlexX: 48
Autodock: 33



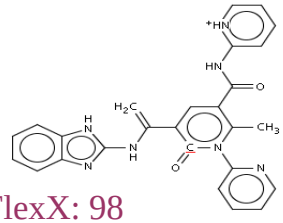
FlexX: 30
Autodock: 97



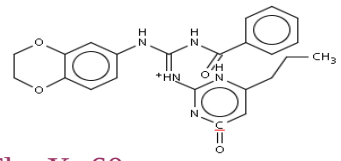
FlexX: 77
Autodock:
130



Good score but bad binding mode



FlexX: 98
Autodock:
110



FlexX: 60
Autodock:160

Compounds for MD and their descriptors

BioinfoGRID

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
WISDOM	smiles	score	Atomcoun	Mass	Ring coun	Accsiteco	donorsitec	Aliphaticat	Chaincour	Fusedaliph	Heteroaromaticringco	PSA	Aliphaticrii	Aromaticri	LogP	TotoalC	Rotatabl	
280991	c1c[nH+]c	-38.763	34	373.305	2	4	5	10	10	0	1	51.03	126.44	0	2	4.05	0.03	3
380406	c1ccc2c(c	-38.103	46	429.449	5	5	3	7	7	0	2	58.8	132.62	0	5	4.65	-0.05	3
378548	Cc1c(cc2c	-39.747	51	441.506	5	5	2	8	8	0	2	63.77	117.07	0	5	4.65	0.01	3
193748	c1ccc(cc1	-38.285	45	434.896	4	6	2	9	9	0	1	55.61	103.43	0	4	6	-0.04	4
242452	c1cc(c(cc	-40.407	46	454.887	3	10	3	13	13	0	0	60.08	148.14	0	3	5.09	-0.04	5
313614	c1cc2c(cc	-38.034	40	388.46	3	4	4	10	10	0	0	55.76	112.3	0	3	4.95	-0.06	3
312057	c1c(cc2c((	-38.412	37	382.41	3	6	4	13	10	0	0	58.45	130.76	1	2	3.64	-0.04	3
384677	COc1c(cc	-39.681	54	489.974	4	7	3	12	12	0	0	61.04	111.55	0	4	5.89	0.04	5
310954	Cc1cc(c(c	-38.2	44	487.756	2	9	3	16	16	0	0	58.74	140.3	0	2	4.58	0	6
243118	Cc1ccc(cc	-37.174	50	403.498	3	5	3	11	11	0	0	55.78	102.32	0	3	5.55	0	4
382373	Cc1c(cccc	-37.633	51	439.51	5	4	3	7	7	0	2	62.63	113.39	0	5	4.86	0.01	3
385534	COc1c(cc	-37.723	52	471.507	5	7	2	9	9	0	1	64.07	113.16	0	5	5.21	0.04	4
372757	c1cc(c(cc	-37.082	43	468.324	4	5	3	8	8	0	1	54.79	119.48	0	4	5.39	-0.04	3
373697	Cc1c(cc(c	-40.574	47	466.898	4	9	2	11	11	0	1	65.01	145.07	0	4	5.83	0.02	4
373762	Cc1c(cccc	-37.302	51	453.942	3	7	3	13	13	0	0	59.71	111.55	0	3	5.35	0.01	5
242449	Cc1cc2c((	-38.131	50	446.48	4	8	2	11	11	0	1	65.29	145.07	0	4	5.12	0	4
492970	CCCCC(=	-39.791	56	435.54	3	6	3	15	15	0	0	48.55	111.55	0	3	5.06	-0.07	7
475515	Cc1c(cccc	-39.502	53	464.495	3	10	3	15	15	0	0	64.69	157.37	0	3	4.79	0.01	6
404128	Cc1n(nc(c	-37.418	56	470.544	4	6	2	11	11	0	1	61.98	117.34	0	4	5.76	0.03	6
326015	c1ccc(cc1	-38.106	46	420.442	3	9	3	12	12	0	0	56.09	148.14	0	3	4.57	-0.06	5
329771	c1ccc(cc1	-37.112	44	375.445	3	5	4	9	9	0	0	47.32	116.31	0	3	4.55	-0.06	4
386759	c1cc2c(cc	-39.937	49	443.479	4	8	3	11	11	0	1	58.25	147.72	0	4	5.5	-0.05	5
430276	c1cc(cc(c	-37.045	44	382.364	3	6	4	10	10	0	0	56.5	82.26	0	3	5.24	-0.05	4
313546	Cc1c(oc2c	-36.875	41	353.329	4	6	3	11	8	0	1	53.43	101.83	1	3	2.65	0.02	3
109865	Cc1c(cccc	-37.293	35	303.743	2	5	3	9	9	0	0	42.06	70.23	0	2	3.44	0.01	3
416361	c1c(ccc(c	-36.661	35	398.239	3	8	3	10	10	0	1	54.17	98.47	0	3	4.12	-0.02	3
120595	c1c(ccc(c	-36.37	35	333.726	3	7	3	11	8	0	0	46.12	88.69	1	2	2.65	-0.02	3
437779	c1cc(c(cc	-36.029	32	415.613	2	5	3	9	9	0	0	40.76	70.23	0	2	3.96	-0.04	3
89351	C=C(c1c((	-36.869	35	269.299	2	4	4	8	8	0	0	34.41	73.39	0	2	2.13	-0.06	4
178145	c1cc(c(c(c	-36.123	31	325.15	2	7	3	9	9	0	1	42.32	83.12	0	2	1.52	-0.04	3
170421	c1c(cc2c((	-35.821	34	300.27	3	7	3	10	7	0	1	42.59	101.58	1	2	0.82	-0.03	3
178319	Cc1ccc(cc	-36.553	42	313.351	2	5	3	11	11	0	0	46.48	79.46	0	2	3.13	0	4
170305	CC(=O)c1	-36.106	37	315.299	2	7	3	11	11	0	0	44.29	87.3	0	2	1.9	0	4
73901	CC1(OCC	-40.888	40	293.318	2	6	3	15	9	0	0	69.14	88.69	1	1	1.29	-0.01	4
81354	c1cc(cc(c	-37.093	31	290.705	2	6	3	8	8	0	1	38.26	83.12	0	2	1.66	-0.04	3
315095	Cc1c2c(cc	-43.098	55	472.923	4	8	3	12	12	0	1	61.19	100.19	0	4	4.92	0.01	5
462971	CCOc1cc(	-45.625	67	568.631	5	12	4	18	12	0	1	72.37	174.69	1	4	4.87	-0.04	8
52923	Cc1cc(c2c	-44.035	47	349.41	3	5	5	10	10	0	1	54.28	104.53	0	3	4.11	0	2
261841	Cc1c2ccc	-43.119	55	440.477	4	6	5	11	11	0	1	63.13	122.01	0	4	4.6	0.04	3
300822	CCCc1[nH	-41.736	56	434.468	4	9	4	20	10	0	0	53.43	115.02	2	2	3.24	-0.07	4
305608	COC(=O)(	-41.596	56	464.451	4	10	4	22	12	0	0	62.03	141.32	2	2	1.65	0.05	5
392786	c1cc2c(cc	-40.486	36	365.194	3	6	5	9	9	0	1	50.21	104.78	0	3	4.38	-0.04	2
316830	Cc1[nH]c(	-38.111	48	412.849	3	9	4	17	11	0	0	53.29	105.79	1	2	3.32	-0.03	3
396606	Cc1nc(nc2	-37.822	46	392.455	4	8	4	8	8	0	1	54.39	117.95	0	4	3.76	0.04	2
17970	Cc1nc(nc(	-37.727	37	286.313	2	7	5	9	9	0	2	44.02	117.42	0	2	-0.29	0.03	3
261841	Cc1c2ccc	-43.119	55	440.477	4	6	5	11	11	0	1	63.13	122.01	0	4	4.6	0.04	3
491148	Cc1[nH]c(	-38.111	48	412.849	3	9	4	17	11	0	0	53.29	105.79	1	2	3.32	-0.03	3
49805	c1cc(c(cc	-39.53	36	326.287	3	8	4	9	9	0	1	48.85	138.57	0	3	2.89	-0.02	3
30030	CCn1c(nc	-18.408	43	306.385	4	2	2	5	5	0	2	42.14	48.92	0	4	3.8	-0.05	4

Overall approach

- Cost effective
- Data and knowledge sharing
- Possibility of novel compound identification

- Difficult to handle large volumes of data (Data management)

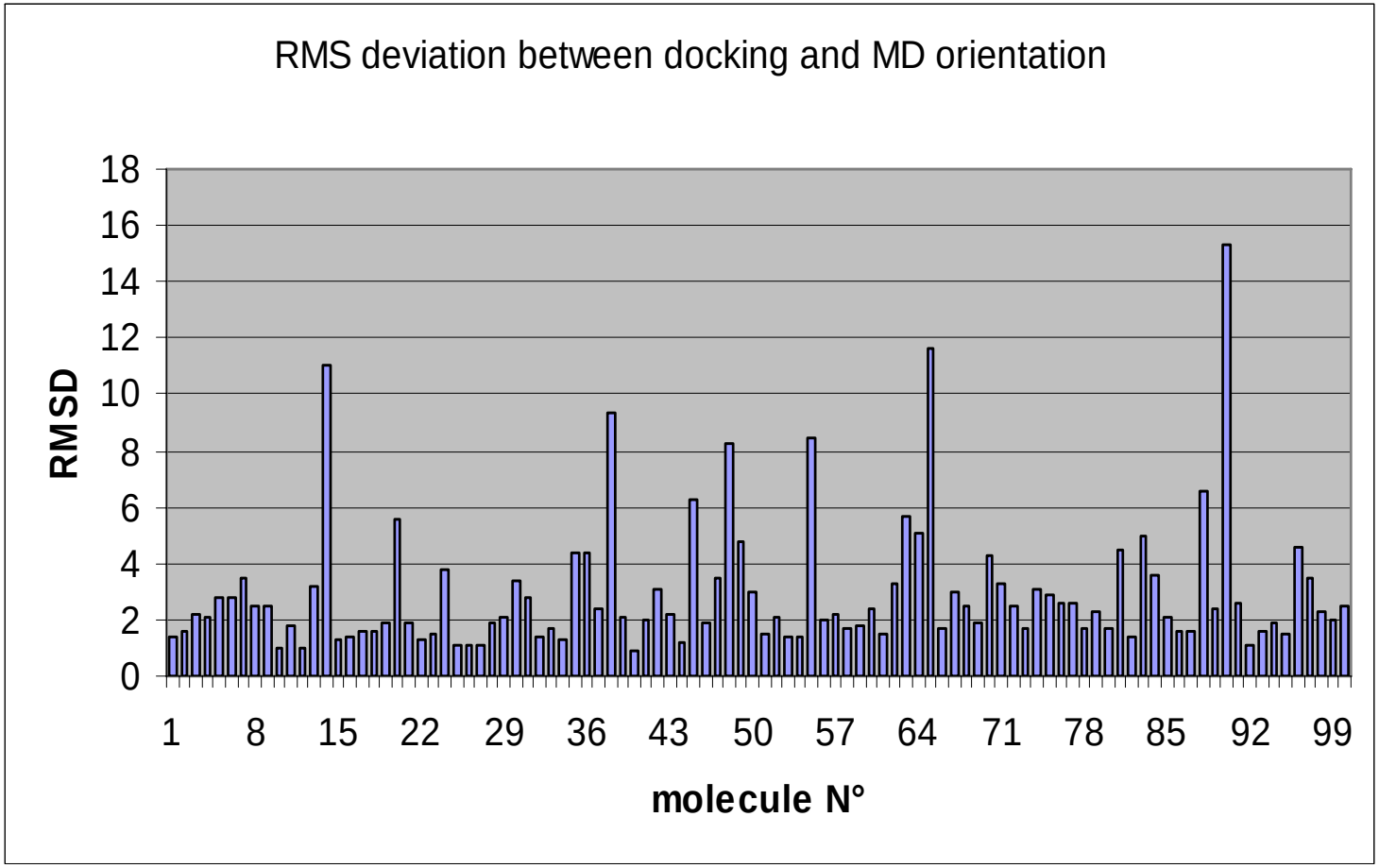
- Electrostatic Solvation parameters (MD)
- A work flow which continues on re-ranking
- Manual interventions are required at each and every step (Automation of the workflow)

MOLECULAR DYNAMICS

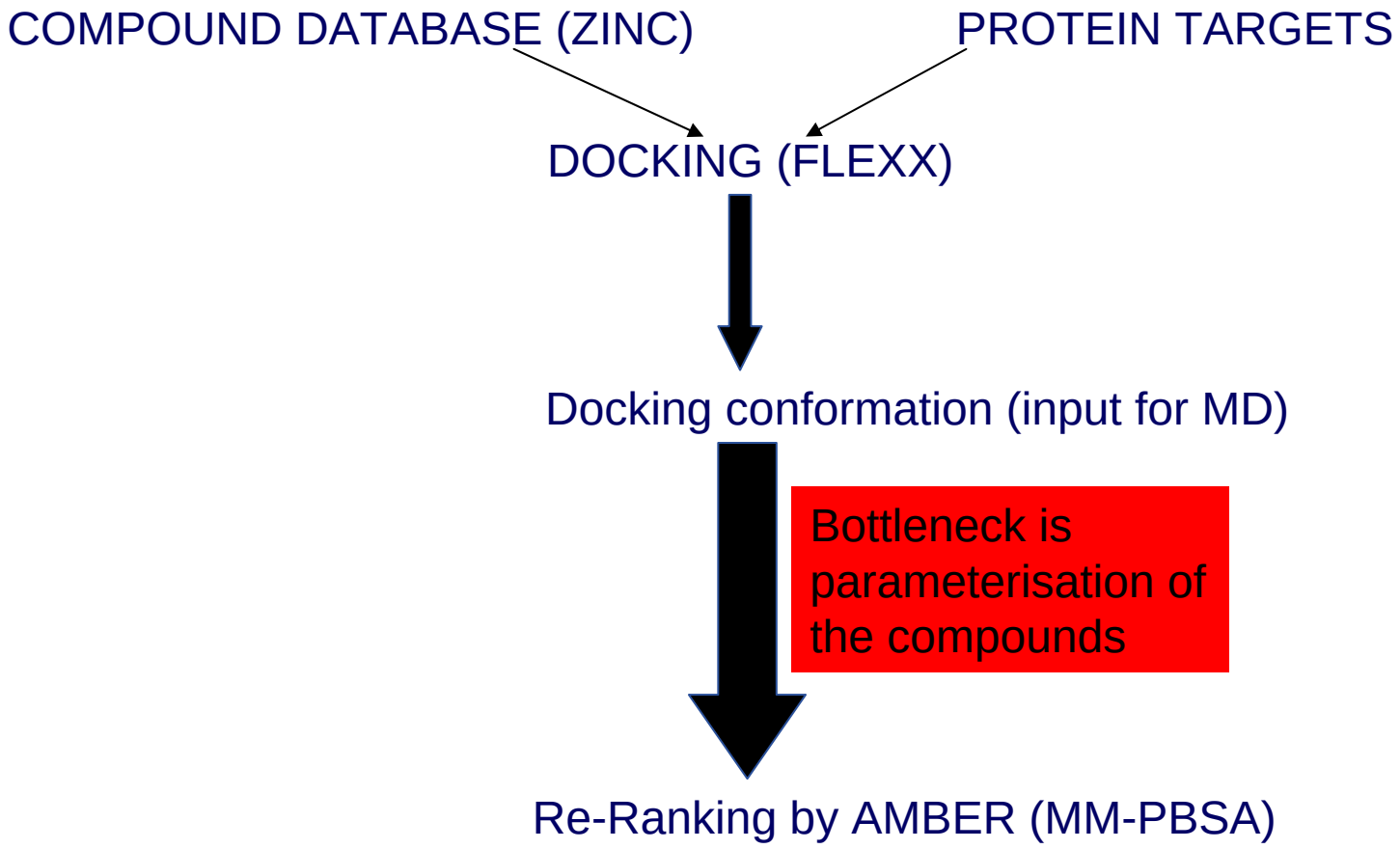
- MD calculations are computationally expensive
- Several steps are involved
- Output of each step serves as input for the subsequent steps, so can be deployed on GRID (potential bottleneck: WMS)
- MM-PBSA method developed by Giulio Rastelli
- This method is faster and well validated for free energy calculations (10-15 mins per complex)

J.Med.Chem. 2005, 48, 4040-4048. Bernd Kuhn, Paul Gerber, Tanja Schulz-Gasch, Martin Stahl

Re-ranking of 100 best hits by MD



Complete Workflow



Perspectives and timescale

- Fall 2006: re-ranking of WISDOM-I results using MD on EGEE
  - Top 1% corresponding to 5000 compounds using 2 different MD approaches
- End of 2006: 100 best hits for in vitro testing
- January 2007: Testing MD giving full flexibility to the protein (A computationally intensive approach of MD)

Acknowledgement

SCAI-Fraunhofer

Marc Zimmermann

Astrid Maass

Horst schwichtenberg

Antje Wolf

Martin Hofmann

CNRS-IN2P3

Nicolas Jacq

Jean Salzemann

Yannick Legre

Matthieu Reichstadt

Florence Jacq

Vincent Breton

UNIMORE-ITALY

Giulio Rastelli

Anna Ferrari