Parallel supercomputer docking program of the new generation: finding low energy minima spectrum

A.V. Sulimov, D.C. Kutov, V.B. Sulimov

Dimonta, Ltd, Moscow / Research Computing Center at Lomonosov Moscow State University

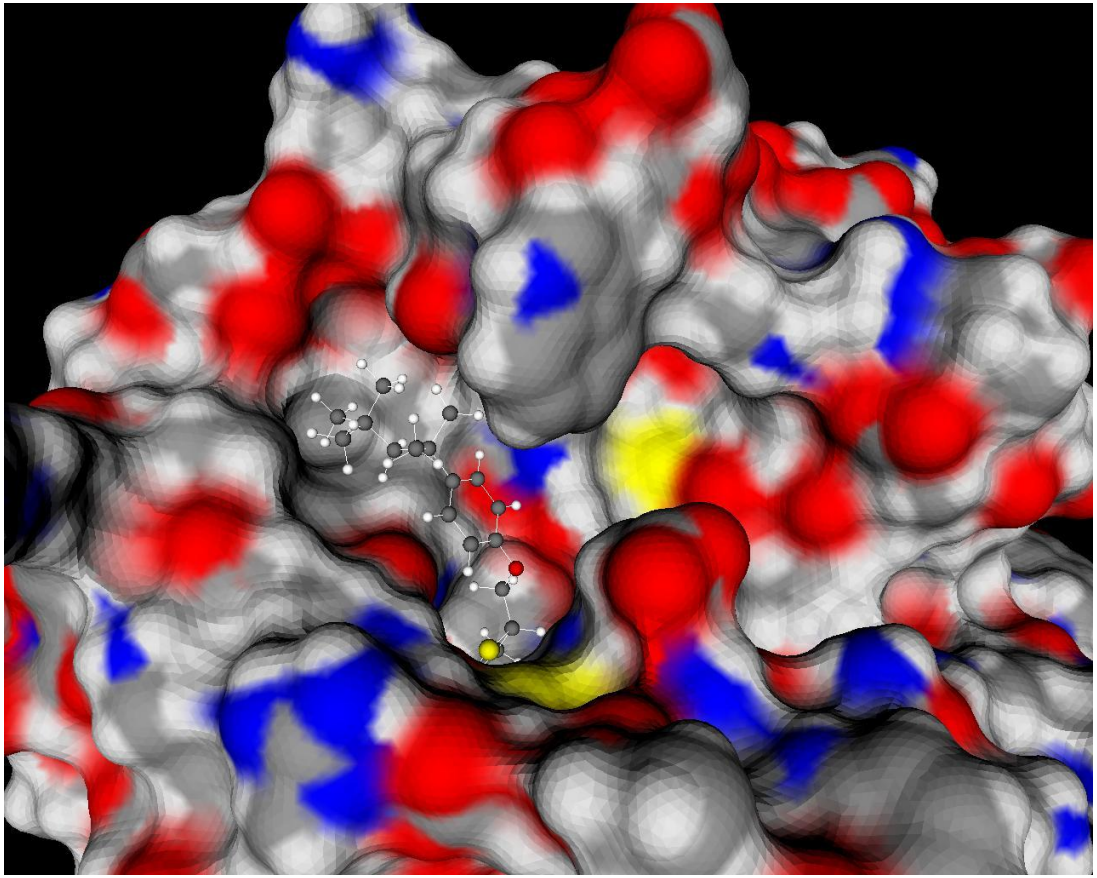
Docking is a popular software used for the drug development

- **Docking:**
  - Ligand positioning in the target protein
  - Computing the protein-ligand binding energy ΔG_{bind}

- **Is it possible to increase docking accuracy?**

  - Positioning accuracy – ***satisfactory***
  - Accuracy of the calculations of the protein-ligand binding energy ΔG_{bind} – ***bad:***
  - *Small errors in positioning result* ***in large errors of*** ΔG_{bind}

Docking paradigm: the ligand binds in the active site of the target protein in close proximity of the global energy minimum of the protein-ligand complex

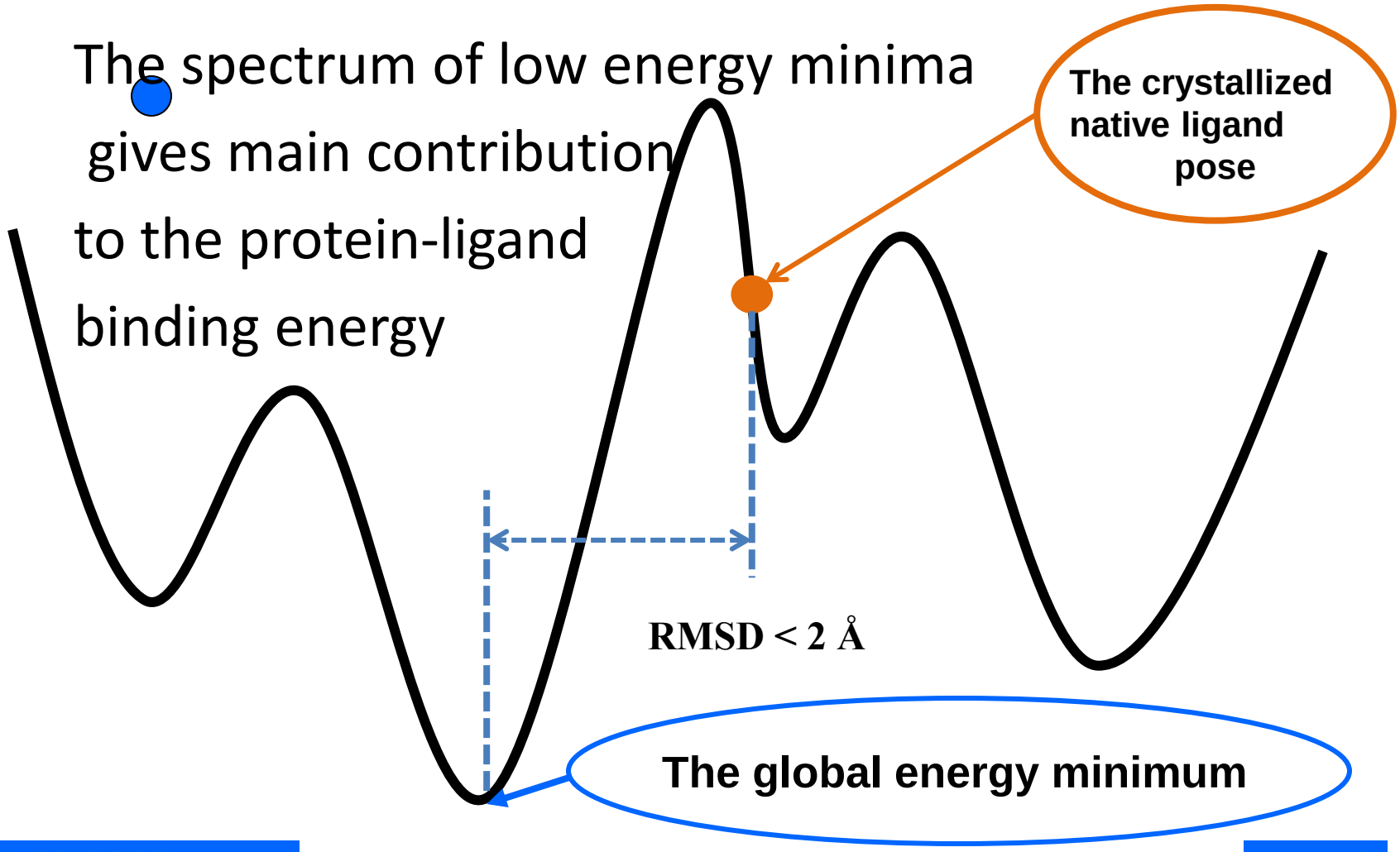


Docking: the search for *low energy minima* of the protein-ligand complex

THE GLOBAL MINIMUM
+ all lowest minima

Thermal motion of the ligand in the active site of the target protein

The spectrum of low energy minima gives main contribution to the protein-ligand binding energy



The test set of 16 protein-ligand complexes

Protein name	PDB ID	Numbers of ligand		Number of Protein atoms	Crystall resolution, Å
		atoms	torsions		
CHK1 (checkpoint kinase 1)	4FT0	42	3	4255	2.3
	4FT9	32	5	4394	2.2
	4FSW	26	0	4342	2.3
	4FTA	35	6	4336	2.4
ERK2 (extracellular signal-regulated kinase 2)	4FV5	52	8	5414	2.4
	4FV6	57	12	5449	2.5
Thrombin	1DWC	71	12	4494	3.0
	1TOM	64	10	4455	1.8
Urokinase	1C5Y	20	2	3869	1.65
	1F5L	24	6	3823	2.1
	1O3P	46	6	3839	1.81
	1SQO	34	4	3823	1.84
	1VJ9	74	19	3859	2.4
	1VJA	61	17	3858	2.0
Factor Xa	2P94	60	7	3676	1.8
	3CEN	50	7	3676	1.6

FLM – the supercomputer docking program of the new generation

I.V. Oferkin et al. Advances in Bioinformatics, vol. 2015, Article ID 126858

Almost **50 "classical" docking programs** and about **10 Internet docking resources** have been developed until now. They were developed within the ***"Faster, even faster" mantra.***

- **All of them** use the preliminary calculated grid of potentials of interaction of ligand probe atom with the target protein – to accelerate calculations
- **Most of them** do not take into account the water solvent or use very simplified models – to accelerate calculations
- **They** use many simplifications for description of inter- and intra-molecular interactions - to accelerate calculations
- **Many of them** are not based on the docking paradigm – they use intuitive (biological, chemical and medical!!!) considerations on positioning of the ligand in the protein
- They use fitting parameters – to "increase" accuracy using experimental training data

Nowadays more efforts are spent to move docking programs to the **"High Accuracy" mode**.

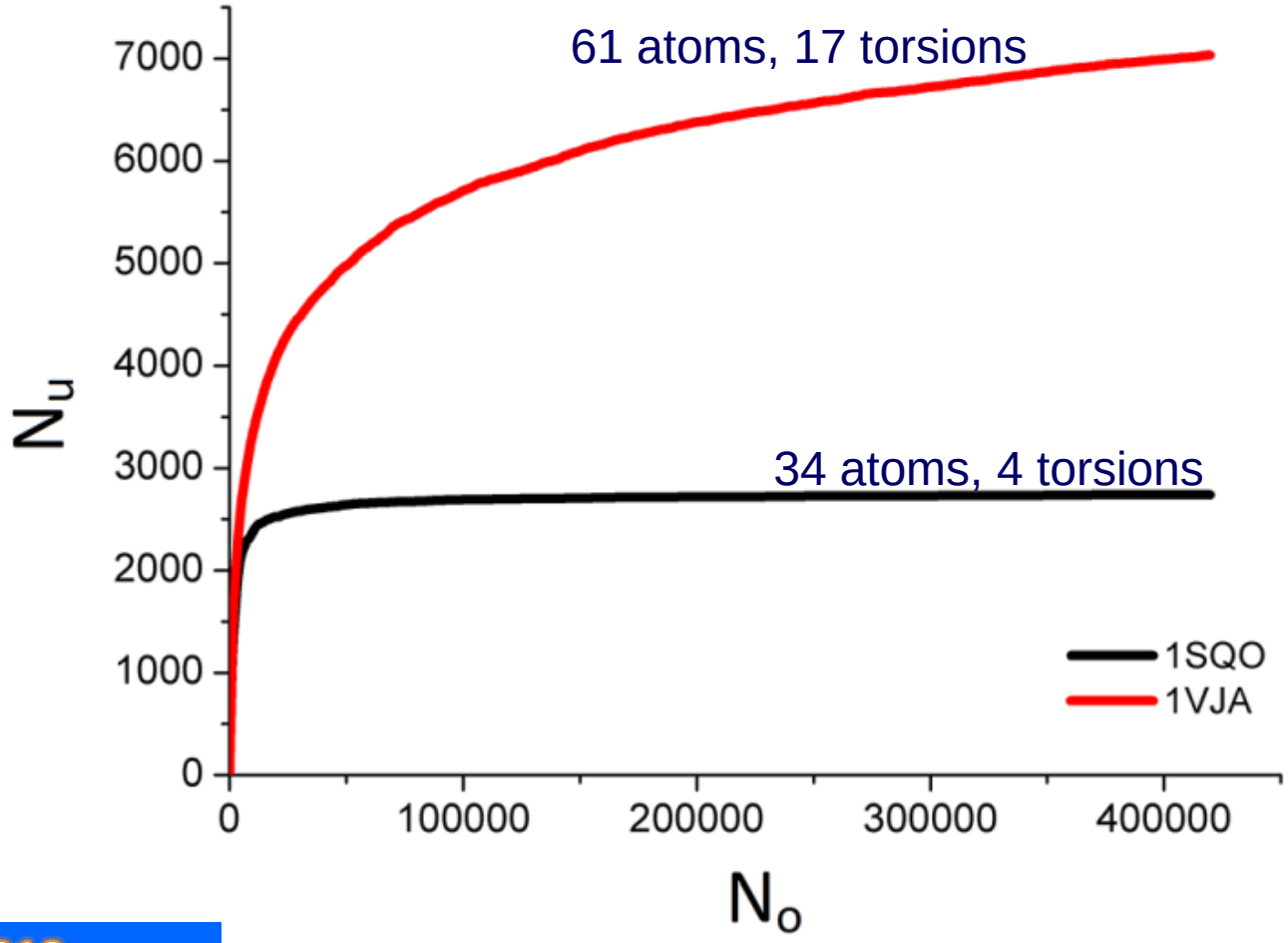
FLM (**F**ind **L**ocal **M**inima) computes the energy in **the MMFF94 force field** for any configuration of a protein-ligand complex without simplifications and approximations.

FLM – the supercomputer docking program of the new generation

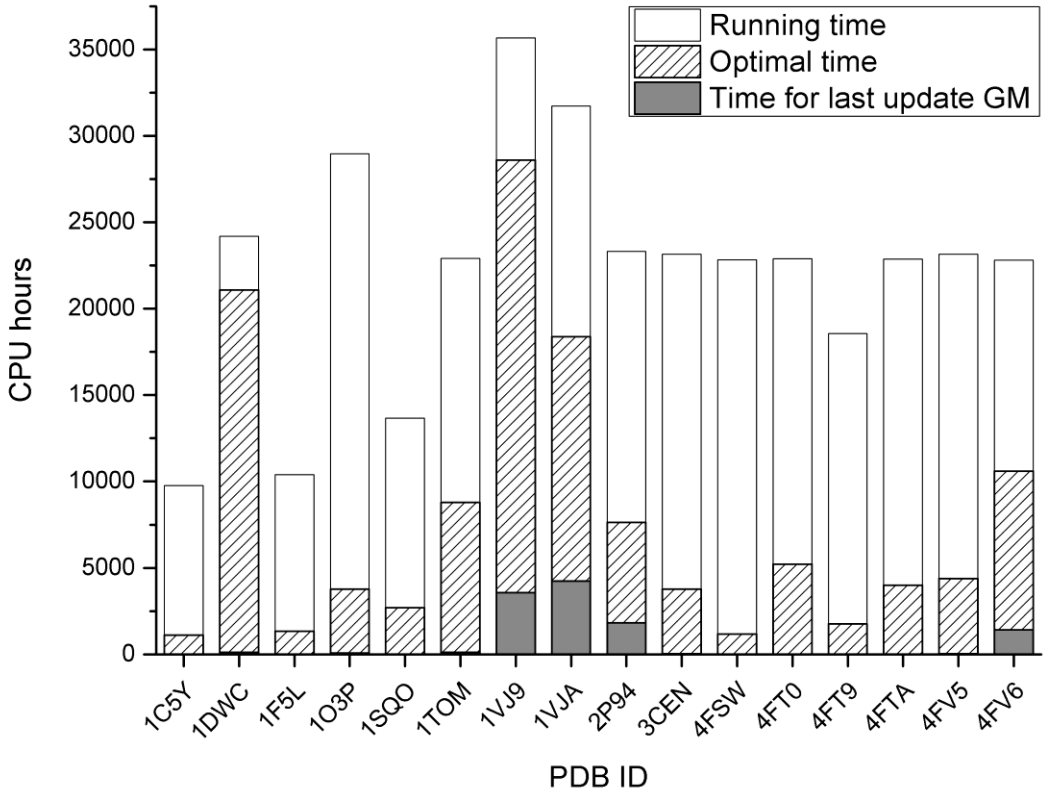
I.V. Oferkin et al. Advances in Bioinformatics, vol. 2015, Article ID 126858

- The **FLM** program searches for low energy minima of a protein-ligand complex in **the MMFF94 force field**
- A **flexible ligand** and a non-flexible protein
- **FLM** does not use the **preliminary calculated grid of potentials** of interaction of ligand probe atom with the target protein
- Algorithm: **massive local energy optimizations from random initial ligand poses, the variation of Cartesian coordinates of all ligand atoms**
- Special attention was paid to the filtration of only **unique minima** in the pool of low energy minima
- FLM finds and saves a pool of a given number of unique low energy minima. This pool consists of the global minimum and every successive minimum above it.
- **FLM-0.05**: the MMFF94 force field **in vacuum**
- **FLM-0.10**: the MMFF94 force field **in water** – the **PCM** or **S-GB** solvent model
- Supercomputer Lomonosov: **8192 computing cores** several hours; FLM performs as long as possible until the pool of low energy minima stops to get renewed.

The number of updates of the pool of low energy minima (*Nu*) as a function of the number of local optimizations (*No*) for the 1SQO complex (black line) and the 1VJA complex (red line). Energy is calculated with MMFF94 in vacuum



CPU time of FLM performance



1. **CPU time depends the number of ligand atoms and torsions**
2. **The global minimum is found much faster than the whole pool of low energy minima**

Energies of global minima found by docking programs of the new generation, FLM and SOL-P, relative to the energy of the global minimum found by the local optimization of best ligand poses determined by the SOL "classical" docking program. SOL-P – uses the tensor train global optimization algorithm. SOL-P much faster than FLM

PDB ID	ΔE_{GM} **FLM**, kcal/mol	ΔE_{GM} **SOL-P**, kcal/mol	PDB ID	ΔE_{GM} **FLM**, kcal/mol	ΔE_{GM} **SOL-P**, kcal/mol
1C5Y	**0.00**	**-0.08**	**2P94**	**-2.75**	**-0.92**
1DWC	**-67.56**	**-67.56**	**3CEN**	**-2.19**	**5.59**
1F5L	**-1.16**	**-1.16**	**4FSW**	**-12.04**	**-12.04**
1O3P	**-5.13**	**-5.11**	**4FT0**	**-24.80**	**-26.57**
1SQO	**-0.11**	**-0.11**	**4FT9**	**-18.31**	**-18.31**
1TOM	**-5.18**	**-1.13**	**4FTA**	**-17.53**	**-17.53**
1VJ9	**-5.55**	**-3.04**	**4FV5**	**-36.65**	**-21.03**
1VJA	**-5.33**	**-2.93**	**4FV6**	**-16.93**	**-7.39**

Quasi-docking: the comparison of different methods of energy calculations. Which method is better for docking?

Alexey V. Sulimov, et al. Journal of Molecular Graphics and Modelling, 2017, 78, 139-147

- FLM-0.10, **MMFF94** + **PCM** solvent, pools of **8192** low energy minima are found for 16 test protein-ligand complexes
- The energy of every minimum is recalculated with different methods:

Force fields

- **CHARMM**,
- **AMBER**,

Quantum-chemical semiempirical methods

- **PM7**
- **PM6-D3H4X**

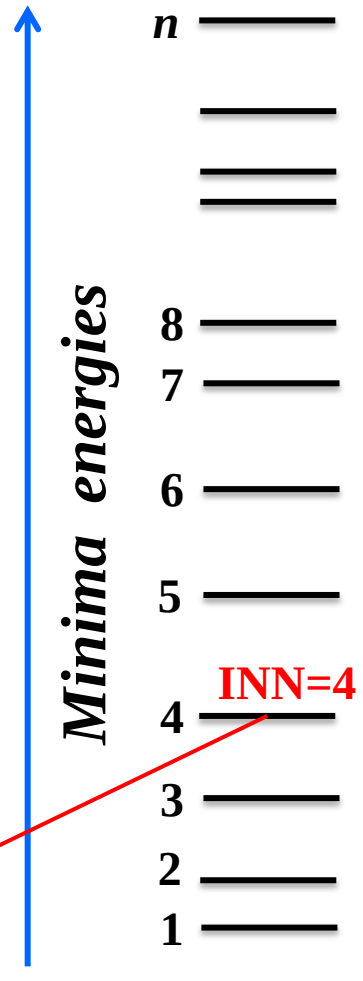
In Vacuum or with implicit solvent models

- Analysis of the feasibility of the docking paradigm is performed: How close in space to the crystallized native ligand pose is the ligand pose corresponding to the global energy minimum?
- The best energy calculation methods for docking are revealed

The local energy minima nomenclature

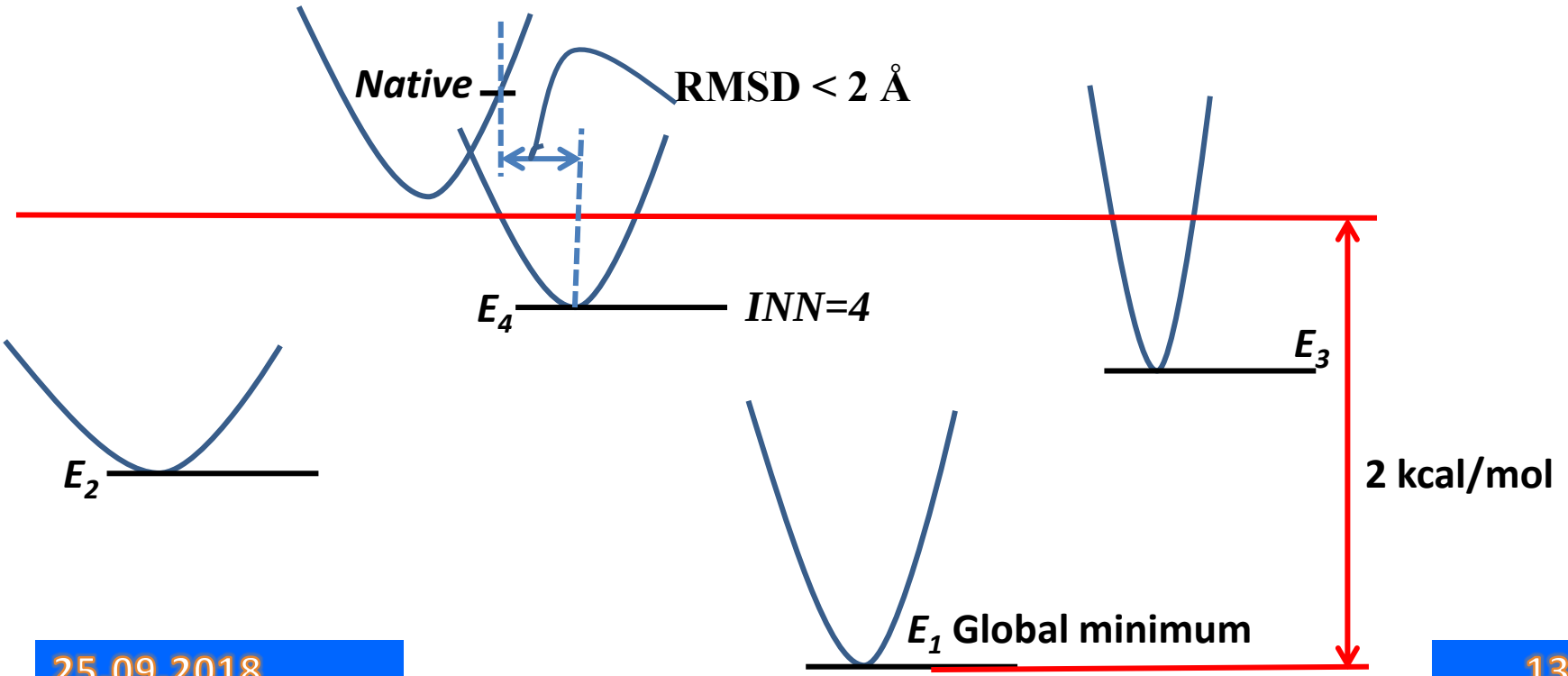
Indices of protein-ligand energy minima index INN

- Each energy minimum has an integer index corresponding to its position in the minima list sorted by their energies in ascending order. The lowest energy minimum has index equal to 1.
- **INN** – (**I**ndex of **N**ear **N**ative) is the index of the low energy minimum having RMSD from the non-optimized native ligand position less than 2 Å

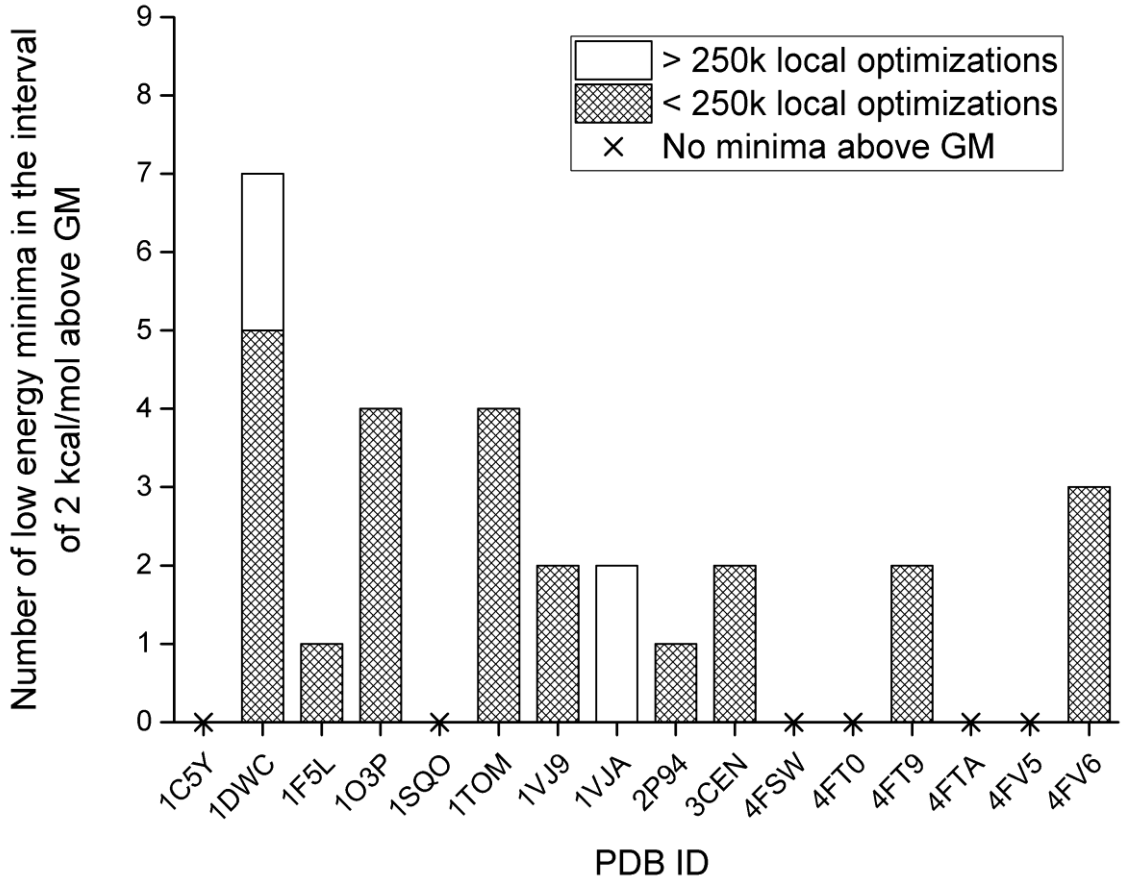


INN=1 – the docking paradigm is satisfied

- The minimum # 4 with the energy E_4 is near the native ligand pose: index ***INN=4***
- **The native** ligand pose – the ligand crystallized in the protein – 3D-structure in Protein Data Bank



How many minima are in the 2 kcal/mol interval above the global minimum?

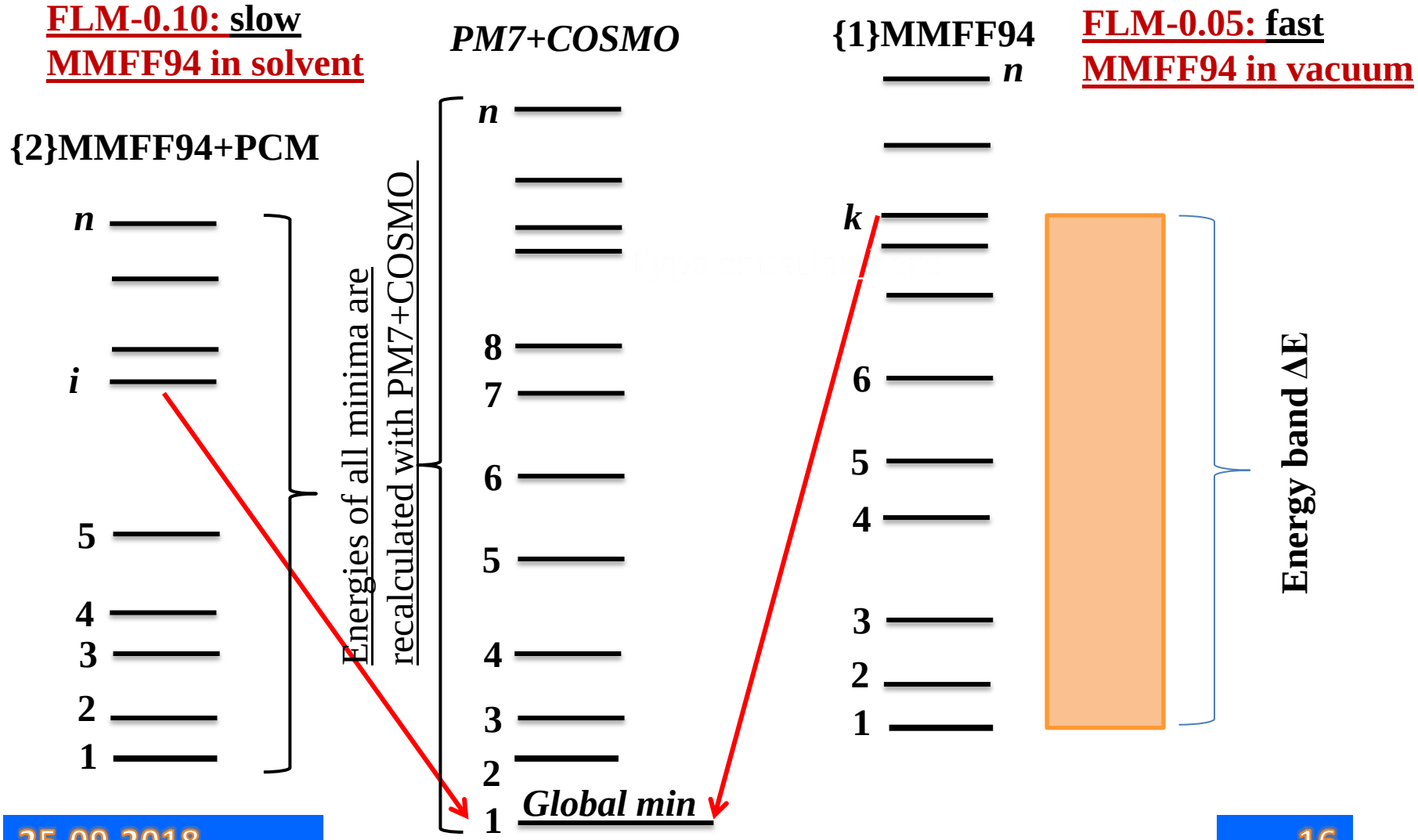


There are 0, 1, … , 7 energy minima in 2 kcal/mol above the global minimum
Energies are calculated with MMFF94 in vacuum

Index INN

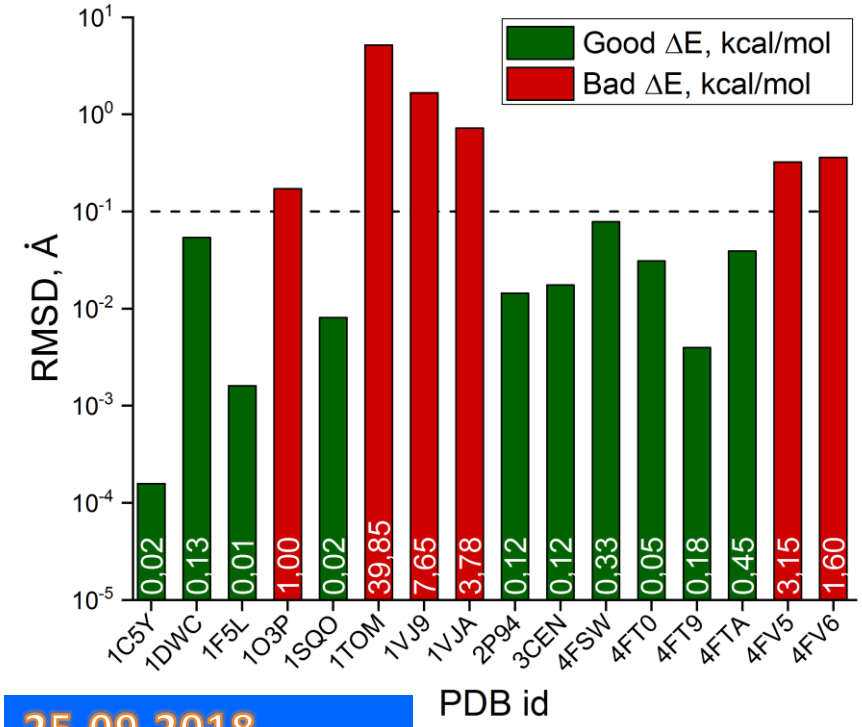
PDB ID	MMFF+SGB	CHARMM +GBSW	AMBER+GB	PM7+COSMO		1SCF PM6-D3H4X+COSMO
				1SCF	OPT	
4FT0	6	76	1	1	1	1
4FT9	1	7	1	1	1	1
4FSW	3	1	1	1	80	1
4FTA	97	8	7	1	1	1
4FV5	5	1	1	1	1	5
4FV6	24	1	1	1	1	3
1DWC	8	1	8	2	4	8
1TOM	1	1	2	1	1	1
1C5Y	1	1	1	1	1	1
1F5L	1	1	1	2	39	2
1O3P	2	393	1	1	7	6
1SQO	1	1	1	1	1	1
1VJ9	14	1	74	14	74	61
1VJA	1	1	40	2	4	6
2P94	1	1	1	1	1	1
3CEN	1	1	5	1	1	4

Is it possible to perform quasi-docking on the base of MMFF94 in vacuum energy minima?



Is it possible to perform quasi-docking on the base of MMFF94 in vacuum energy minima?

- FLM-0.10: **{2}MMFF94+PCM** – low energy minima ***in solvent***
  Quasi-docking: the recalculation of energy of minima with PM7+COSMO,
  The determination of the PM7+COSMO global minimum
- FLM-0.05: **{1}MMFF94** – low energy minima ***in vacuum***

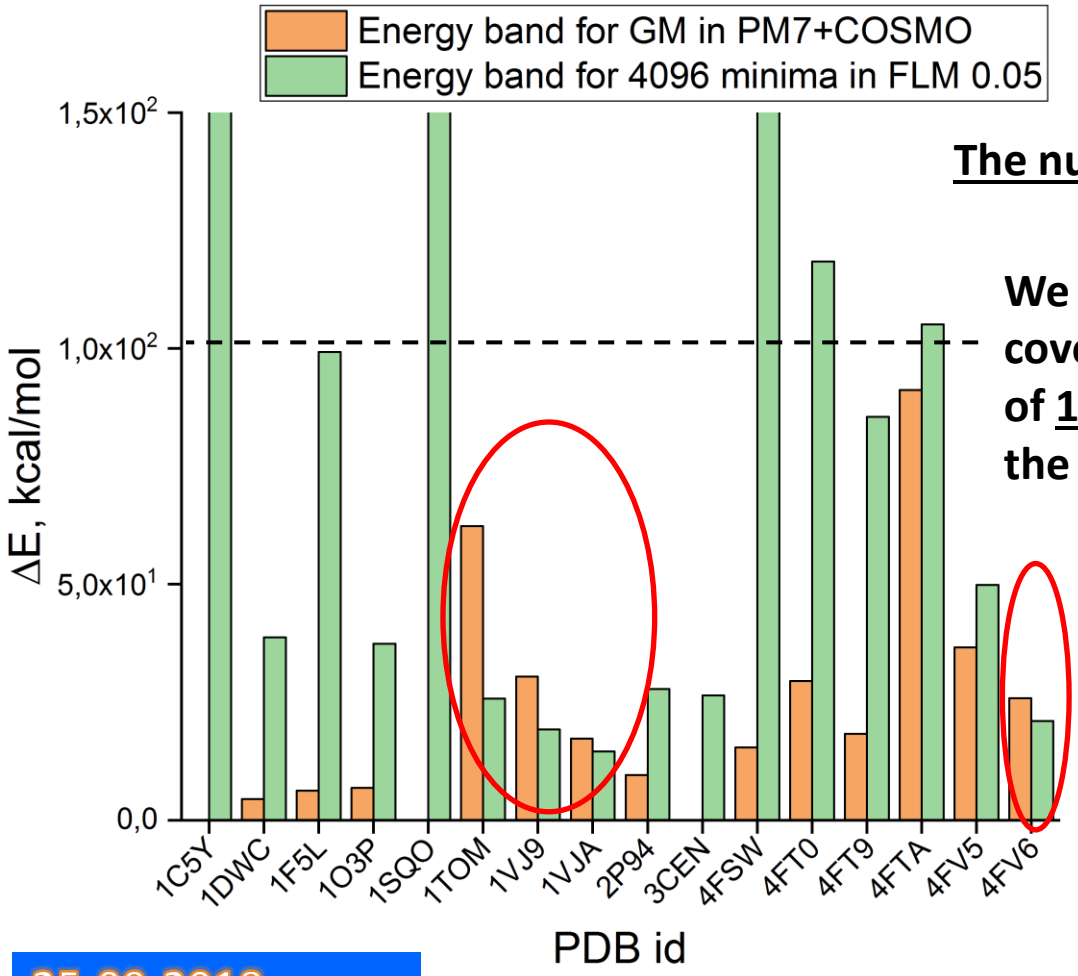


Is it possible to find the ligand pose corresponding to <u>the global minimum</u> among {1}MMFF94 minima?

YES

<u>Bad!</u> <u>What is the reason?</u>

How many low energy minima calculated with *MMFF94 in vacuum* should we save to identify the global minimum for PM7+COSMO?



The number of minima in {1}MMFF94?

We should save all minima spectrum covering the MMFF94 energy band of 100 kcal/mol width from the global minimum

Conclusions

- The docking paradigm is satisfied **for some protein-ligand complexes** when the energy is calculated in vacuum with force fields MMFF94, CHARMM, AMBER and with quantum-chemical semiempirical methods PM7, PM6-D3H4X
- **Including water with implicit solvent models increases the number of test complexes for which the docking paradigm is satisfied for all these methods of energy calculations**
- **The CHARMM force field with solvent is better than MMFF94 and AMBER with solvent**.
- **The PM7 quantum-chemical method with the COSMO solvent is better than force fields: MMFF94, CHARMM and AMBER with solvent models**
- The PM6-D3H4X quantum-chemical method with the COSMO solvent is somewhat worse than PM7 with COSMO

MAIN CONCLUSION: Docking should be made with PM7+COSMO energy

Thank you

… Surely every medicine is an innovation; and he that will not apply new remedies, must expect new evils …

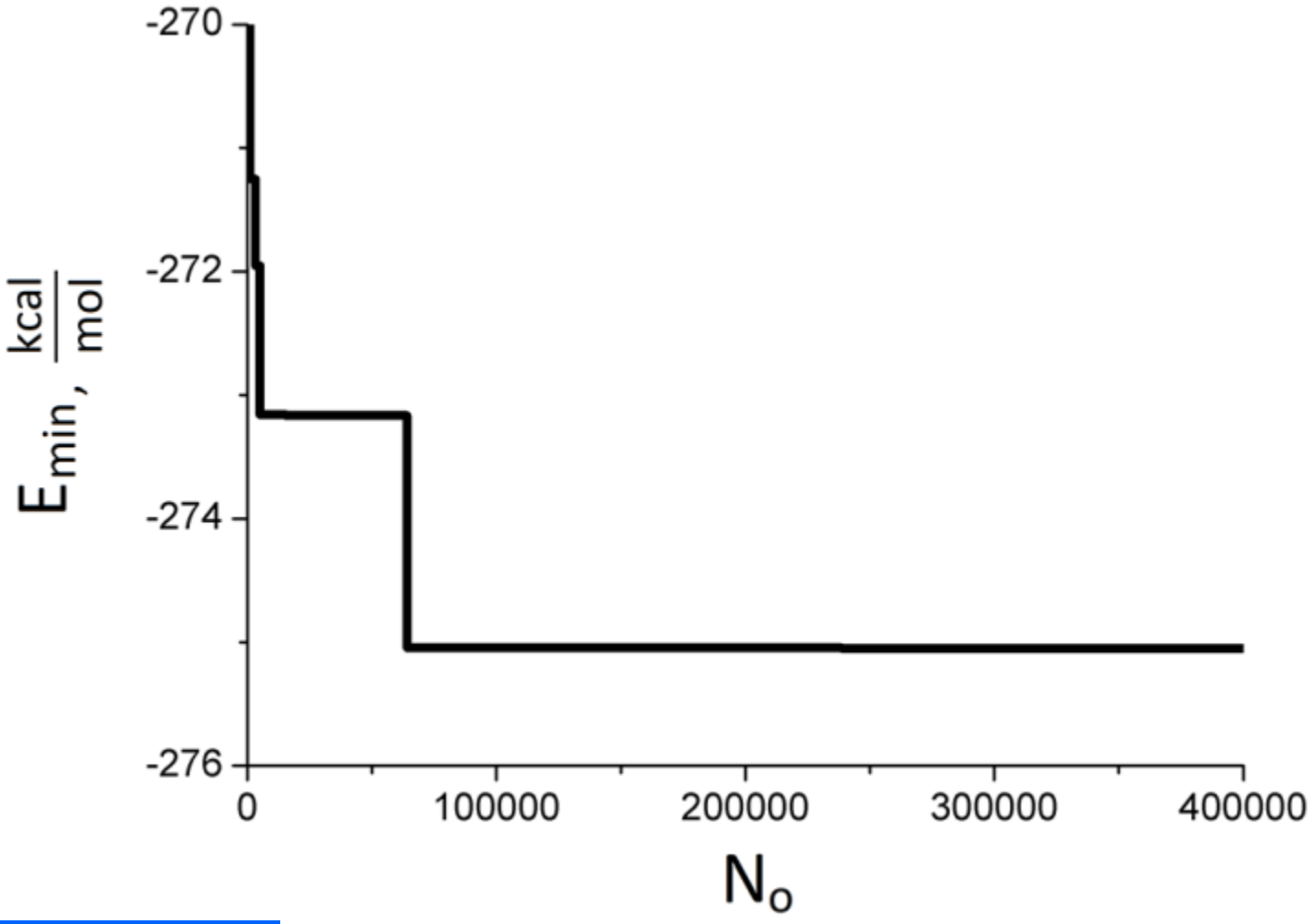
Francis Bacon
(1561-1626)
OF INNOVATIONS

The work is financially supported by the Russian Science Foundation, Agreement no. 15-11-00025-II
The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011

Docking accuracy depends on:

- Modeling of inter- and intra-molecular interactions – **the Force Field choice**
- Which **force field or quantum method** is better for docking?
- Which is better for docking: **vacuum or water calculations**
- 3D models of the protein and ligands
- The Docking algorithm – the Global optimization algorithm
- The Binding free energy calculation: methods and approximations

E_{min} – the global energy minimum
N_o – the number of optimizations, the 1VJA complex



Index INN: vacuum - solvent

Complex PDB ID	MMFF vacuum	MMFF PCM	MMFF SGB	CHARMM vacuum	CHARMM GBSW	AMBER vacuum	AMBER GB	PM7 vacuum	PM7 COSMO
4FT0	99	159	6	1	76	219	1	75	1
4FT9	125	1	1	221	7	283	1	22	1
4FSW	102	140	3	38	1	31	1	413	1
4FTA	Inf	187	97	2675	8	289	7	Inf	1
4FV5	134	3	5	14	1	1138	1	279	1
4FV6	289	68	24	1	1	1362	1	11	1
1DWC	114	35	8	102	1	141	8	106	2
1TOM	Inf	4	1	4010	1	245	2	877	1
1C5Y	1	1	1	1	1	1	1	1	1
1F5L	1	1	1	1	1	29	1	21	2
1O3P	62	1	2	20	393	298	1	2	1
1SQO	1	1	1	1	1	4	1	1	1
1VJ9	1	18	14	1	1	4838	74	17	14
1VJA	49	2	1	1	1	5798	40	6	2
2P94	1	1	1	1	1	204	1	7	1
3CEN	1	1	1	10	1	990	5	1	1