Alexey Sulimov, Danil Kutov, Ivan Ilin,
Anna Gribkova, Anna Taschilova, Vladimir Sulimov

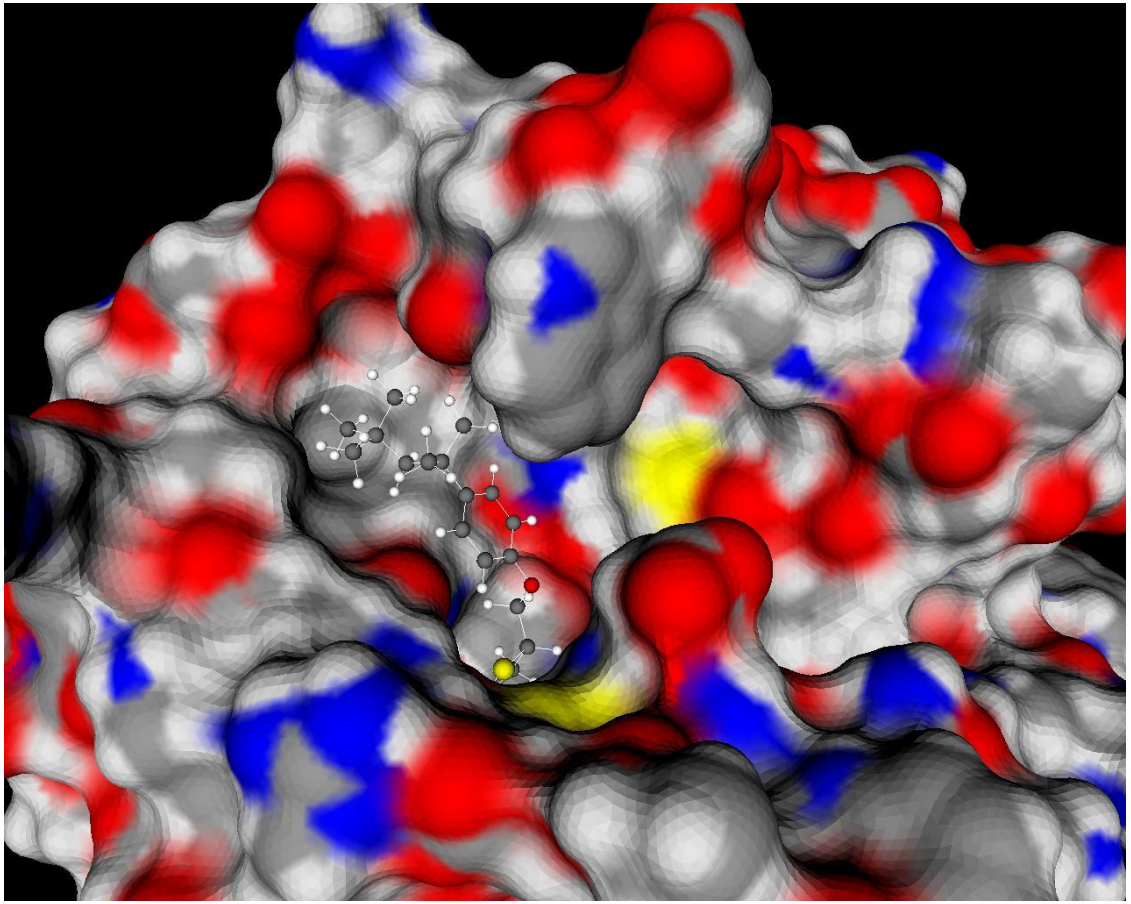
SEARCH FOR APPROACHES TO SUPERCOMPUTER QUANTUM-CHEMICAL DOCKING

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***

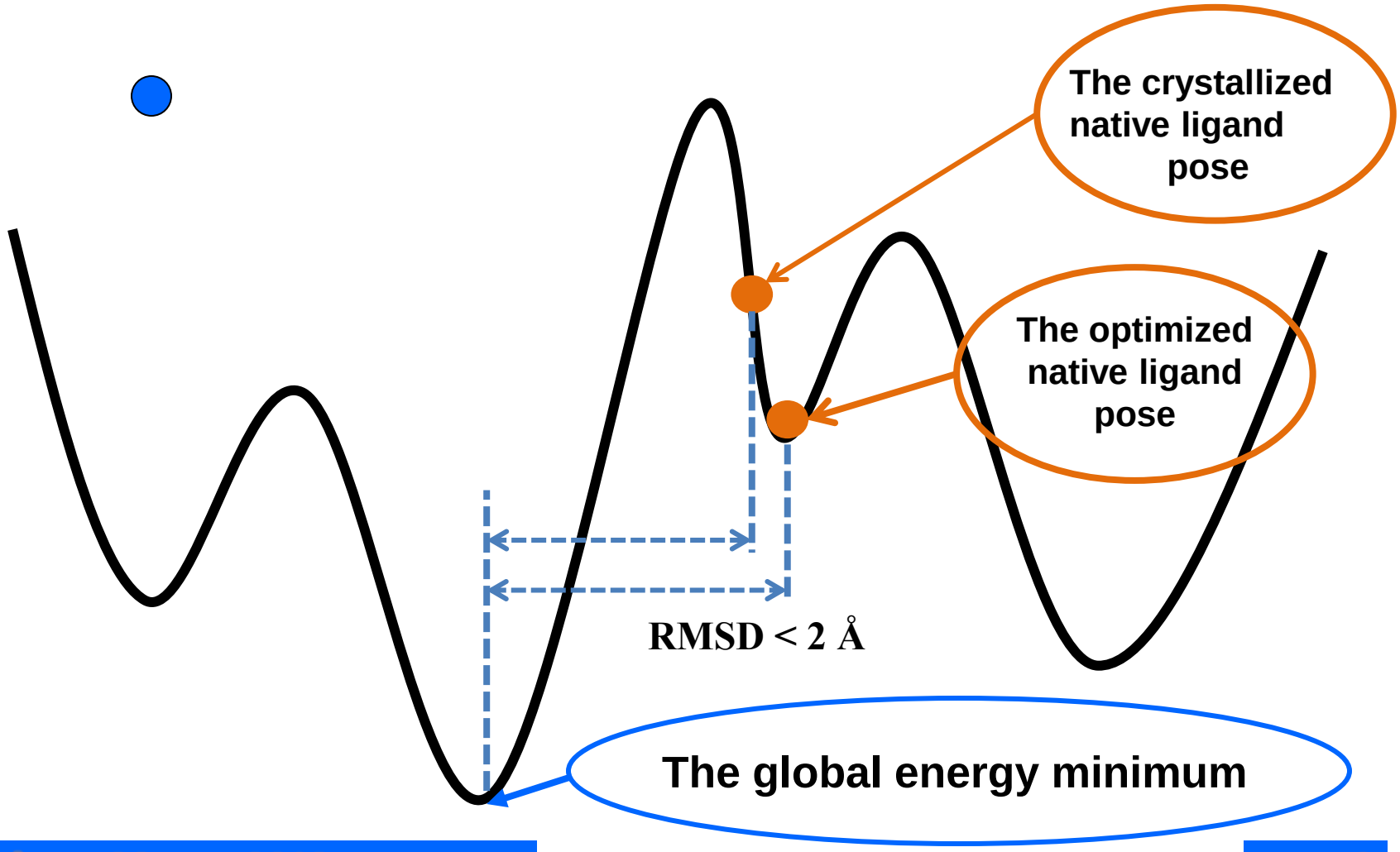
Docking accuracy → **Drug discovery efficiency**

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 is the search for the global minimum of the energy of the protein-ligand complex

Many local minima on the protein-ligand multi-dimention energy surface



Docking accuracy depends on:

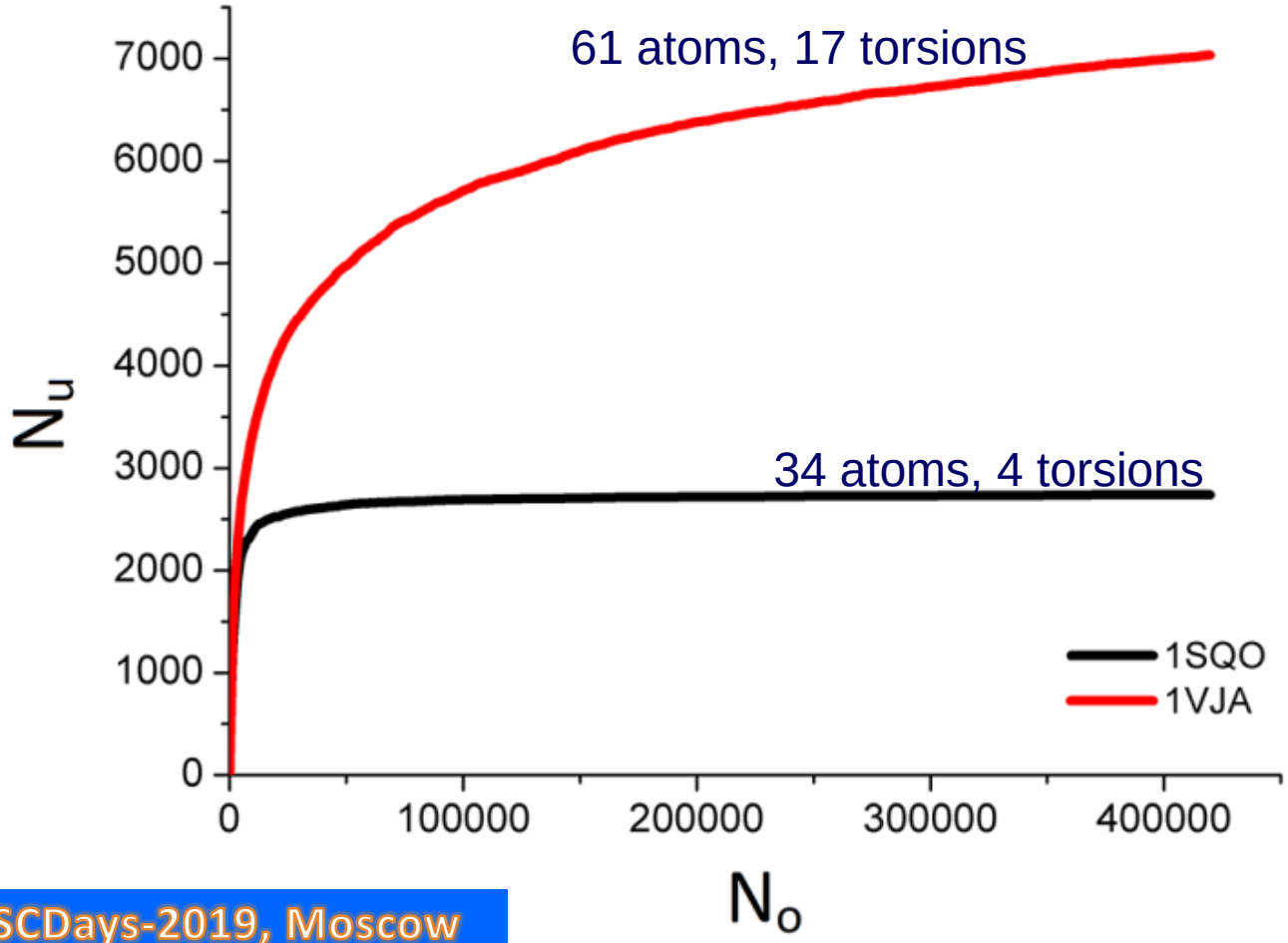
- Models of the target protein and ligands
- Modeling of inter- and intra-molecular interactions – the Force Field choice
- Solvent (water) model
- The docking algorithm
- The free energy calculation method and approximations
- Computer resources for docking of 1 ligand

FLM – the supercomputer docking program of the new generation

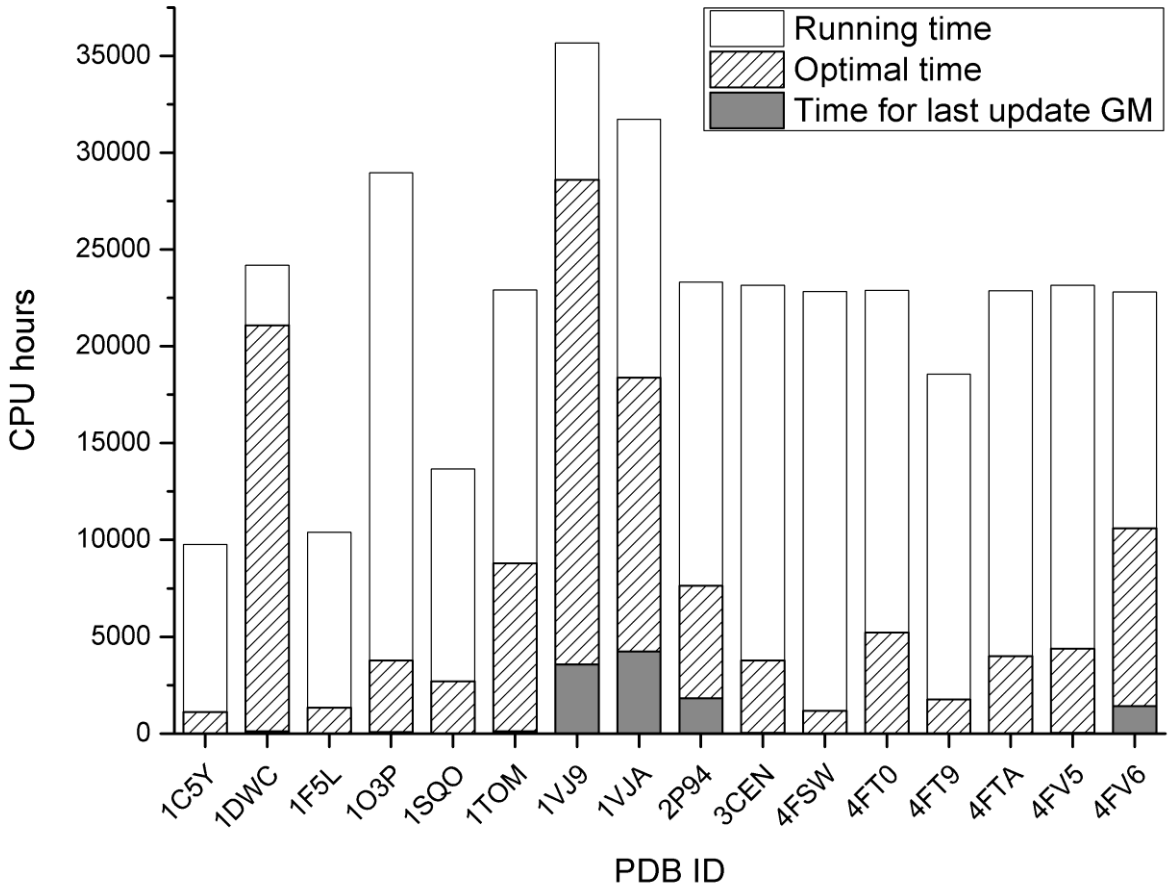
I.V. Oferkin et al. Advances in Bioinformatics, vol. 2015, Article ID 126858

- The **FLM** (**F**ind **L**ocal **M**inima) 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**
- **FLM-0.05**: the MMFF94 force field **in vacuum**
- **FLM-0.10**: the MMFF94 force field **in water** – the **PCM** 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**

25 test protein-ligand complexes

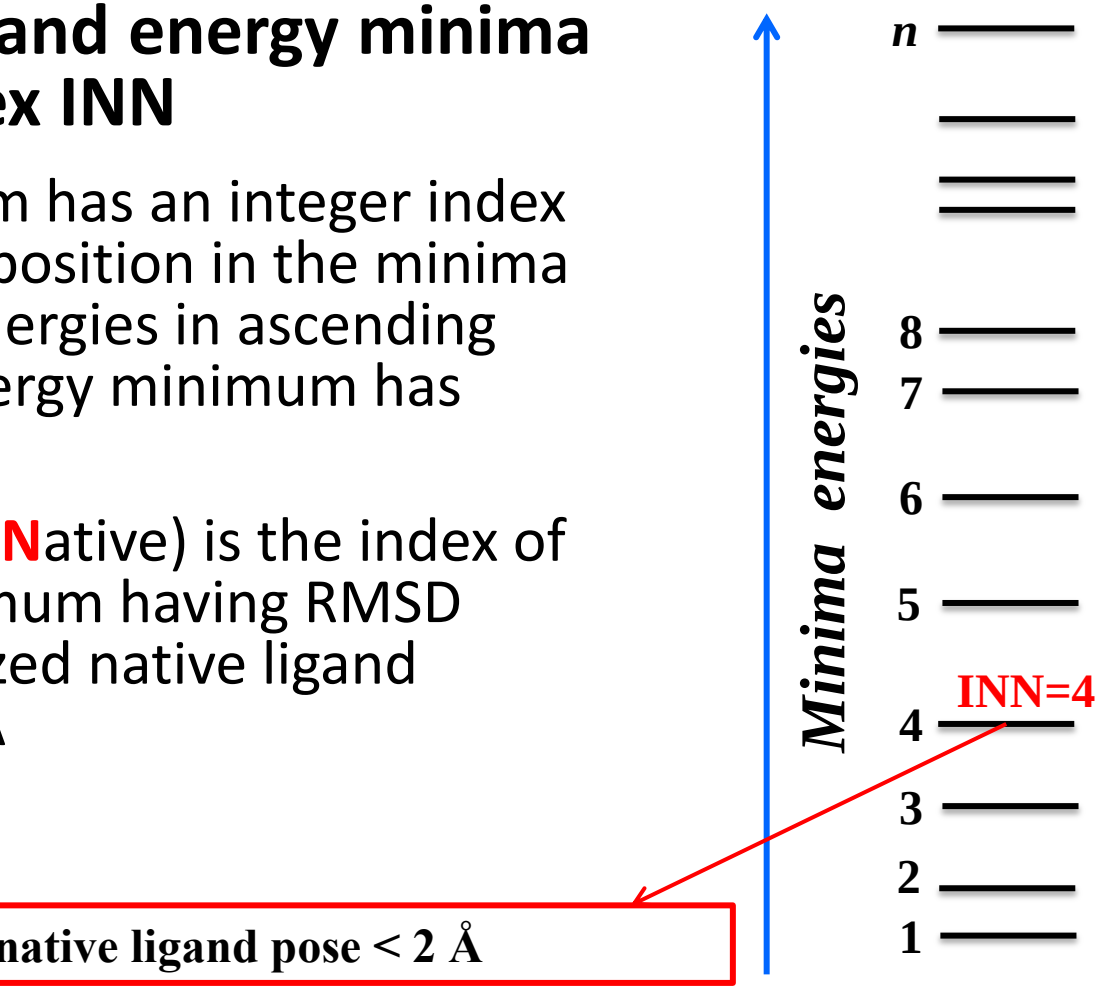
PDB ID	*RES*, Å	N_P	N_L	N_{tor}
1C5P	1,43	3220	18	1
1HW5	1,82	3284	33	1
1J84	2,02	2642	87	10
1K1J	2,2	3220	68	10
1MRX	2	3140	74	11
1MSM	2	3138	78	12
2IKO	1,9	5144	46	5
2PYM	1,9	3100	86	12
2PYN	1,85	3116	86	12
2XII	1,8	7042	51	4
2Z8D	1,85	5897	51	6
2Z8E	1,99	5897	51	6
2ZDM	1,93	3220	59	9

PDB ID	*RES*, Å	N_P	N_L	N_{tor}
2ZDN	1,98	3220	58	9
2ZFS	1,51	3220	64	9
3KDB	1,66	3138	86	13
3KIV	1,8	1206	20	5
3NU3	1,02	3134	70	13
4CRC	1,6	3711	60	11
4CRD	2,1	3692	57	11
4LL3	1,95	3134	75	13
4MR5	1,63	1860	42	3
4MR6	1,67	1860	49	6
4P8V	1,64	5741	57	8
5CSD	1,45	2407	53	14

The local energy minima nomenclature

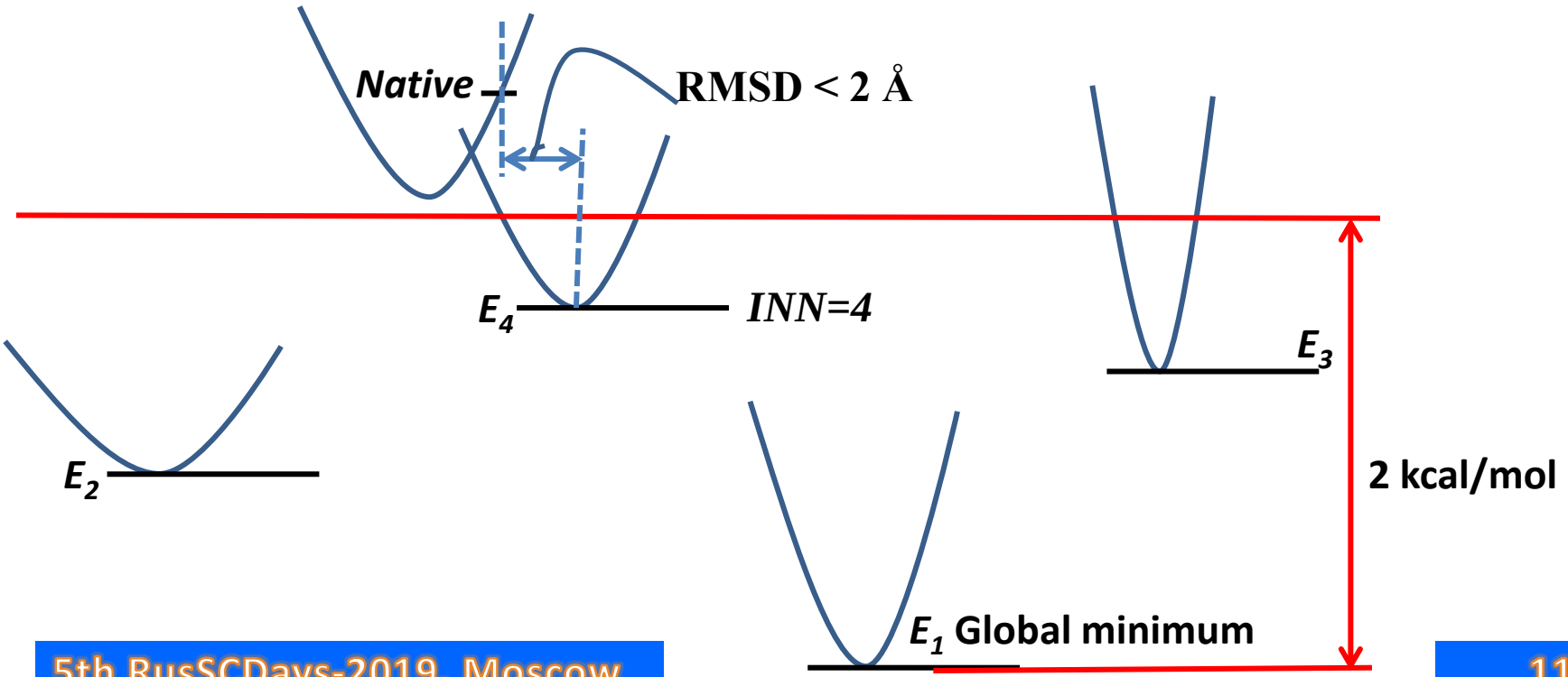
Indices of protein-ligand energy minima index INN

- 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



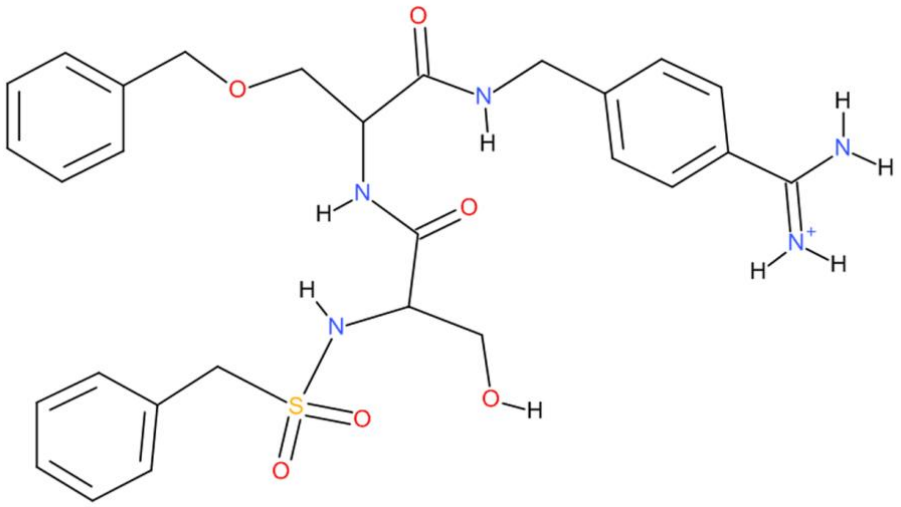
INN = 1 – the docking paradigm is satisfied

- 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



Positioning accuracy

- **Check for feasibility of the docking paradigm**
- A set of 25 protein-ligand complexes from Protein Data Bank, RMSD ≤ 2 Å, different ligand sizes and flexibility with known inhibition constants K_i , *binding enthalpy* ΔH



The number of ligand atoms: 18-86; the number of ligand torsions: 1-14

Quantum chemical quasi-docking

- The search and storing a large number of low energy minima (**8192**) in the frame of the MMFF94 force field in vacuum (no solvent)
- Recalculation of all **8192** minima by the quantum chemical semiempirical PM7 method (the MOPAC package) taking into account water using the implicit COSMO model: no optimization – 1SCF – **PM7+COSMO 1SCF**
- Local energy optimization with PM7(MOPAC) with variations of ligand atoms positions and the energy recalculations with **PM7+COSMO – taking into account solvent in the implicit model**

PM7 – the new quantum-chemical semiempirical method NDDO (Neglect of Diatomic Differential Overlap)

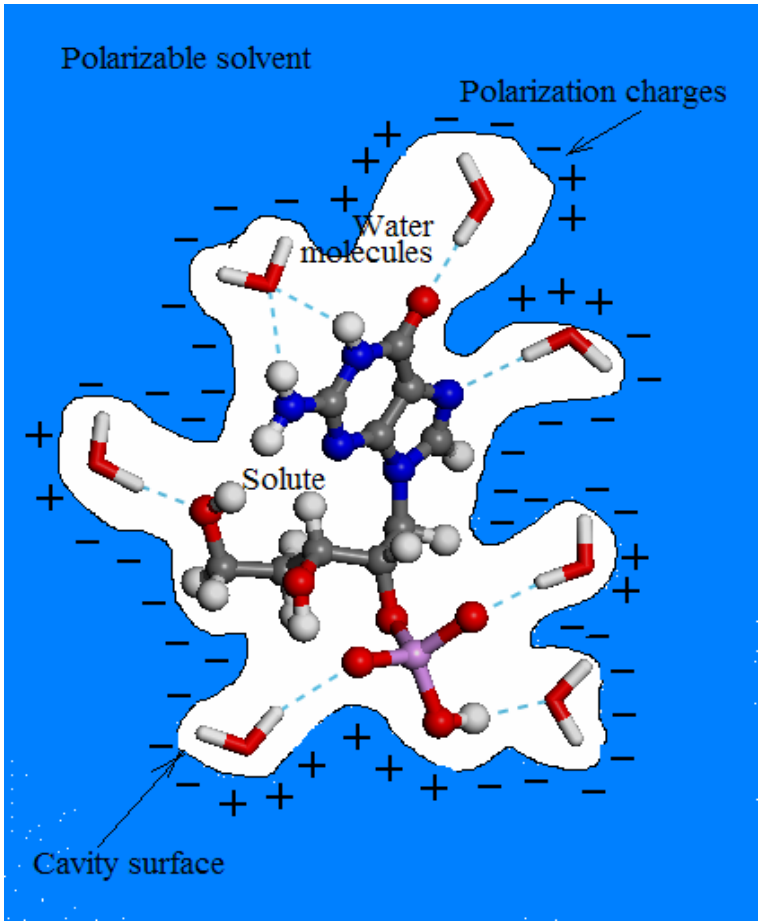
- **Improved dispersion interactions**
- **Improved Hydrogen Bonds description**

J. J. P. Stewart J. of Molecular Modeling, 2013, vol. 19, 1–32

Water important for protein-ligand binding
Dielectric constant of water ε = 78.5 at 300K

Implicit solvent models

- Solvent is a homogeneous continuum surrounding a molecule: solute-solvent interactions:
  - polar
  - non-polar
- Poisson's equation can be converted into the equation for polarization charges on the surface separating the molecule and solvent.
- COSMO model: ε = ∞, polarization charges are found on the surrounding metal surface



MOPAC the package for semiempirical quantum-chemical calculations

1. PM7 method
2. The MOZYME module allows to perform calculations of proteins: thousands of atoms
3. COSMO was implemented into MOPAC at 1993, parameterized for AM1 to reproduce experimental hydration energies of small molecules; only polar term
4. COSMO2: re-parameterized in 2019 for PM7 and PM6 on a large set of molecules with experimentally measured hydration energies; polar+non-polar terms
5. Non-polar term: $E_{np}^{S}(X) = \xi \times S(X), \xi$ – surface tension parameter
6. Fitting calculated hydration energies to experimental hydration energies: variations of model parameters: VdW radii and ξ

RMSD calculated hydration energy against experimental values, kcal/mol

Methods	MNSol*	Mobley266	SAMPL1	SAMPL4
PM6/COSMO2	2.65	2.81	5.08	2.44
PM6/COSMO	4.31	3.72	9.07	4.01
PM6/PCM	6.35	2.24	3.93	1.92
PM6/SMD	4.80	3.03	4.91	2.59
PM7/COSMO2	2.62	2.54	3.73	1.92
PM7/COSMO	3.96	3.44	6.07	3.21
AM1/COSMO	4.80	2.26	8.89	2.26
DFTB3/C-PCM	9.92	6.40	12.87	8.96
DFTB3/C-PCM′	6.16	2.31	3.66	2.84
DFTB3/SMD	4.86	2.65	4.85	3.15
DFT/COSMO-RS	2.54	1.08	1.88	1.59

INN and RMSD deviation for different energy functions

PDB ID	MMFF94		PM7 COSMO		PM7 COSMO2	
	INN	RMSD	INN	RMSD	INN	RMSD
1C5P	1	0,62	1	0,43	1	0,43
1HW5	1	0,54	1	0,48	1	0,48
1J84	538	5,43	1	1,97	3	2,12
1K1J	1	1,96	1	0,33	7	9,43
1MRX	1	0,62	1	0,47	96	3,93
1MSM	1	0,95	1	1,87	2	8,34
2KIO	1	0,55	1	0,49	1	0,49
2PYM	2	9,57	9 (1)	1,66	5 (1)	1,66
2PYN	2	9,61	1	1,22	1	1,22
2XII	3	6,59	1	0,58	1	0,58
2Z8D	1	0,73	2	3,67	1	0,68
2Z8E	1	1,09	1	1,11	1	1,11
2ZDM	1	0,96	1	0,91	1	1,11

PDB ID	MMFF94		PM7 COSMO		PM7 COSMO2	
	INN	RMSD	INN	RMSD	INN	RMSD
2ZDN	1	0,59	1	0,68	1	0,68
2ZFS	1	1,56	2	2,67	3	2,67
3KDB	2	12,18	1	0,96	2	6,43
3KIV	1	1,25	1	0,75	1	0,75
3NU3	1	1,34	1	0,44	2	4,16
4CRC	928	9,79	2	2,67	2	2,67
4CRD	5237	5,46	1	1	1	1
4LL3	1	1,04	4	8,38	1	1,89
4MR5	1	0,45	5	8,05	3	8,05
4MR6	9	8,84	1	1,16	1	1,16
4P8V	232	8,1	1	0,58	2	4,13
5CSD	14997	11,65	993	10,33	1	0,93

Protein-ligand binding enthalpy

- The ligand position in the target protein correspond to the global energy minimum of the protein-ligand complex PL
- Complexes for which the docking paradigm is satisfied INN=1
- COSMO $\Delta H_{bind} = E_1(PL) - E_1(P) - E_1(L)$
- COSMO2 $\Delta H_{bind} = E_1(PL) - E_1(P) - E_1(L) + E_{np}^{DS}$

$$E_{np}^{DS} = \xi \times \Delta S \qquad \Delta S = S(PL) - S(P) - S(L)$$

$$\xi = 0.042 \text{ kcal/mol/Å}^2$$

$E_1(P)$ – energy of protein structure taken from Protein Data Bank

$E_1(L)$ – unbound ligand energy found in the quasi-docking

ΔH_{exp} and ΔH_{bind} binding enthalpies (kcal/mol)

The ligand pose is obtained by		Quasi-docking	
PDB ID	ΔH_{exp}	ΔH_{bind} COSMO	ΔH_{bind} COSMO2
1C5P	-4.52	-54.75	-44.7
1HW5	-0.97	-54.74	-52.12
1K1J	-9.46	-82.71	—
1MRX	-2.1	-54.89	—
1MSM	-7.6	-67.86	—
2IKO	-9.5	-81.19	-78.78
2XII	-9.8	-92.09	-86.41
2ZDM	-7.24	-82.2	-81.2
2ZDN	-5.09	-85.08	-79.37
2ZFS	-4.52	—	—
3KDB	-1.55	-54.68	—
3NU3	-7.3	-54.46	—
4LL3	-16.4	—	-52.87
4MR6	-4.04	-47.42	-54.41
5CSD	-14.2	—	-67.5
The coefficient of correlation R		0.74	0.20

Conclusions

- Docking. High accuracy of ligand **positioning**: the quasi-docking with the PM7 quantum-chemical semiempirical method and the COSMO solvent model is demonstrated.
- The high correlation (0.74) of measured and calculated values of the **binding enthalpy** for the PM7 method and the COSMO.
- The performance of the new COSMO2 parameterization [18] is worse in positioning and in binding enthalpy calculations
- The PM7 method with COSMO is appropriate for the realization supercomputer quantum-chemical docking program.

Спасибо за внимание

… 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 was financially supported by the Russian Science Foundation, Agreement no. 15-11-00025-П