XXXI Symposium on
Bioinformatics and
Computer-Aided Drug
Discovery (BCADD-2025)

IN SILICO REVERSE FRAGMENT BASED DRUG DISCOVERY APPROACH (R-FBDD): CORE IDEAS, CURRENT STATUS AND FUTURE DIRECTIONS

***Dmitry Shulga**, Nikita Ivanov, Arslan Shaimardanov, Vasily Myazin, Daniil Chernov, Ilyas Zhanataev, Vladimir Palyulin*

Department of Chemistry at Moscow State University, Moscow, Russia

Email: shulga@qsar.chem.msu.ru

New useful in silico approach in a historic perspective

In silico **Reverse Fragment Based Drug Discovery** (R-FBDD)[1,2]

1. Core ideas
   a. FBDD, its pros and cons
   b. **R-FBDD** as a solution
      i. implementation
      ii. core scenarios
2. On the concept of fragment-in-molecule-in-a-position
3. Recent advancements and future directions

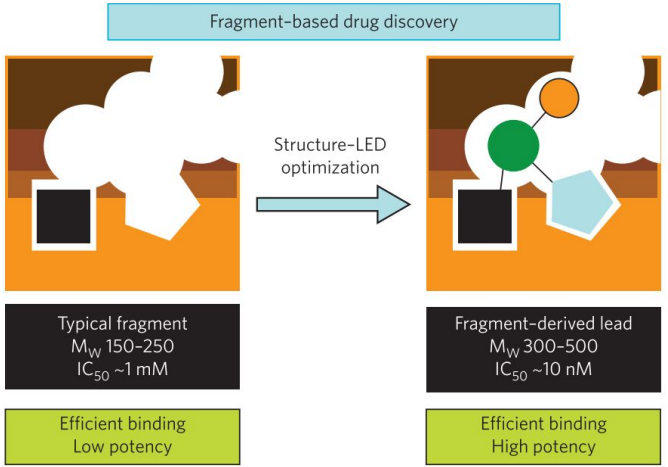
5 years history

1. Shulga D. A., Ivanov N. N., & Palyulin V. A. (2021) // *Mend. Comm.*, *31*(3), 291-293, doi: 10.1016/j.mencom.2021.04.004
2. Shulga D. A., Ivanov N. N., & Palyulin V. A. (2022) // *Molecules*, *27*(6), 1985., doi: 10.3390/molecules27061985

FBDD advantages

Fragment Based Drug Discovery/Design - FBDD

1. better chemical space coverage with smaller libraries
   a. less risks to find a hit
   b. fewer compounds to screen
2. highly efficient binding patterns (in terms of LE)
   a. higher chances to get a drug
   b. shorter optimization path
3. phys-chem properties control - less risks of lead attrition in clinical Phases
4. direct link to SBDD and subsequent rational design
5. alternative scaffolds for the same target - lower risks of 'bad' series
6. proven source of lead and approved drug compounds by 2020-ies



*taken from [1]

FBDD ~ 7% of lead compounds

1. Murray, C. W., & Rees, D. C. (2009) // *Nature chemistry*, *1*(3), 187-192., doi: 10.1038/nchem.217
2. Murray, C. W., Verdonk, M. L., & Rees, D. C. (2012). // *Trends in pharmacol. sci.*, *33*(5), 224-232., doi: 10.1016/j.tips.2012.02.006
3. Joseph-McCarthy, D., Campbell, A. J., Kern, G., & Moustakas, D. (2014) // *JCIM*, 54(3), 693-704., doi: 10.1021/ci400731w
4. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016) // *Nat. Rev. Drug Disc.*, *15*(9), 605-619., doi: 10.1038/nrd.2016.109

FBDD difficulties

Fragment Based Drug Discovery/Design - FBDD

1. specialized methods are required to measure weak binding of fragment hits
2. specialized methods are required to establish binding modes
3. working at the edge of the applicability domain of many tools
4. requires highly experienced experts in different fields to collaborate
5. … and finally…

1. Murray, C. W., & Rees, D. C. (2009) // *Nature chemistry*, *1*(3), 187-192., doi: 10.1038/nchem.217
2. Murray, C. W., Verdonk, M. L., & Rees, D. C. (2012). // *Trends in pharmacol. sci.*, *33*(5), 224-232., doi: 10.1016/j.tips.2012.02.006
3. Joseph-McCarthy, D., Campbell, A. J., Kern, G., & Moustakas, D. (2014) // *JCIM*, 54(3), 693-704., doi: 10.1021/ci400731w
4. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016) // *Nat. Rev. Drug Disc.*, *15*(9), 605-619., doi: 10.1038/nrd.2016.109

FBDD difficulties

Fragment Based Drug Discovery/Design - FBDD

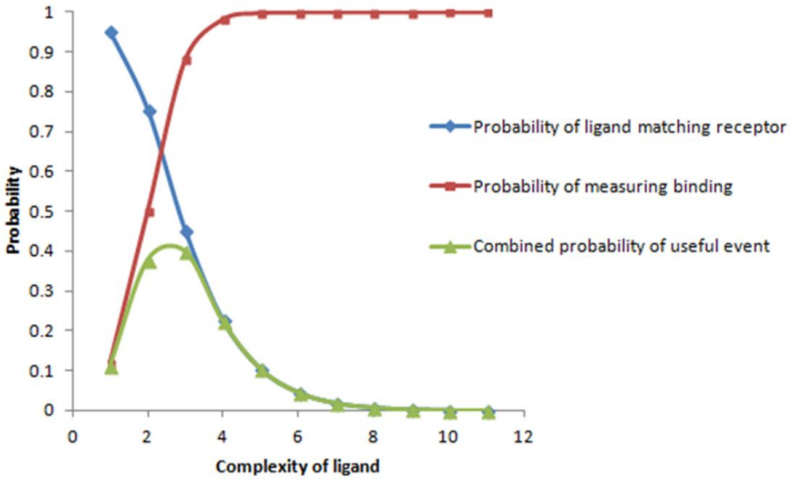
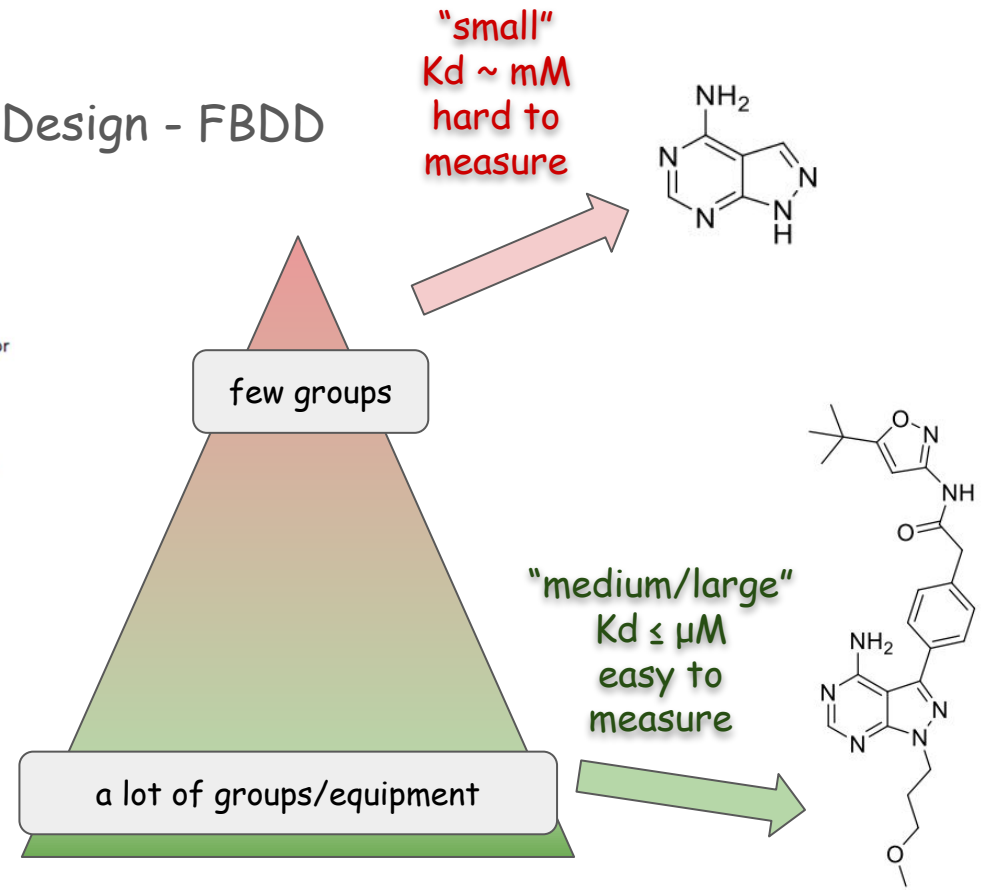


Figure taken from: Keseru, G. M., Erlanson, D. A., et al. (2016) // *JMC*, 59(18), 8189-8206, doi: 10.1021/acs.jmedchem.6b00197



FBDD difficulties… and new challenge

Fragment Based Drug Discovery/Design - FBDD

Democratize FBDD!

few groups

a lot of groups/equipment

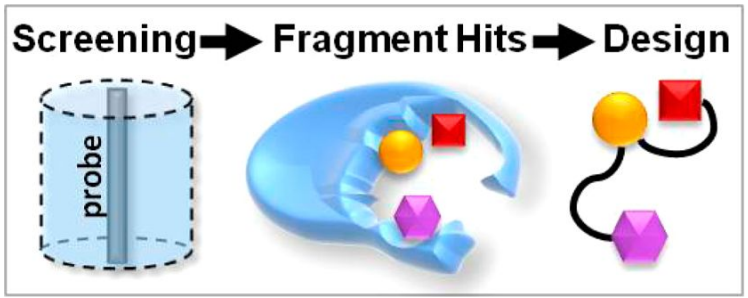
Mission: make benefits of FBDD accessible to many scientists!

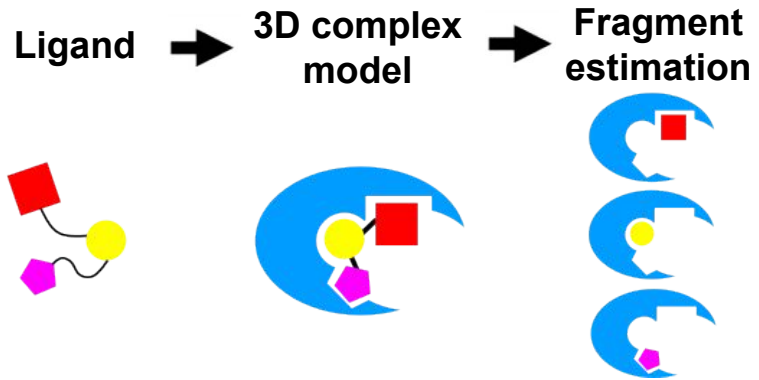
Corollary: large and medium sized molecules as input

Enter R-FBDD!

Fragment approach (FBDD): find the most active/affine fragments -> combine/grow into a new lead molecule.

***Reverse* fragment approach (R-FBDD)**: split a molecule into fragments -> estimate energetic contribution of each fragment





Joseph-McCarthy, D., Campbell, A. J., Kern, G., & Moustakas, D. (2014) // *JCIM*, 54(3), 693-704., doi: 10.1021/ci400731w

Shulga D. A., Ivanov N. N., & Palyulin V. A. (2021) // *Mend. Comm.*, *31*(3), 291-293, doi: 10.1016/j.mencom.2021.04.004

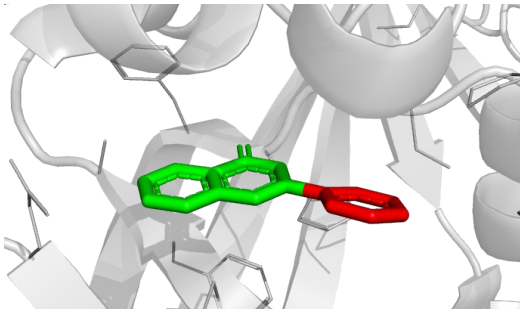
Enter R-FBDD! Basics

Input
- 3D ligand-receptor complex
- reference energy, ΔE_{mol}
- fragments of interest

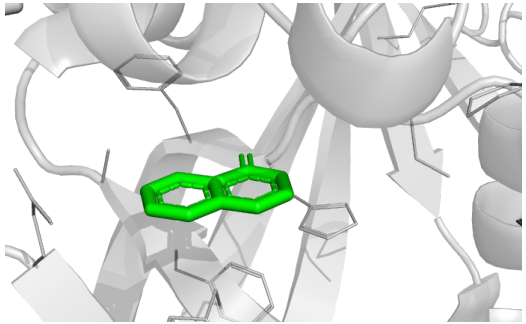
Means
- scoring function[1]

Output
- energy contribution for each defined fragment

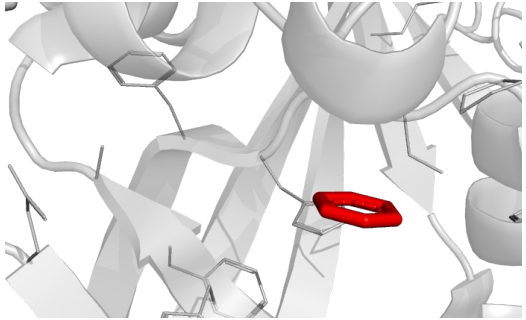
Entire ligand
@ receptor



Fragment#1
@ receptor



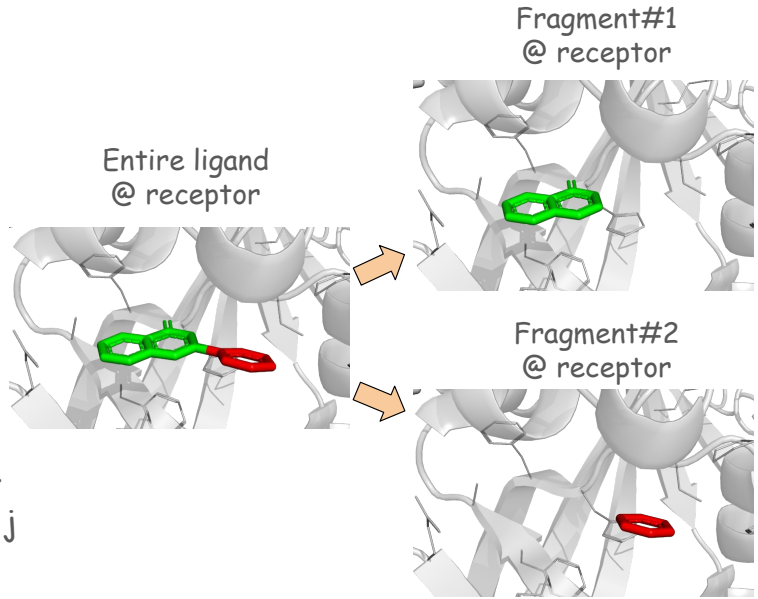
Fragment#2
@ receptor



1. fast and consistent initial implementation

Enter R-FBDD! Details

Input
- 3D ligand-receptor complex
- reference energy, ΔE_{mol}
- fragments of interest

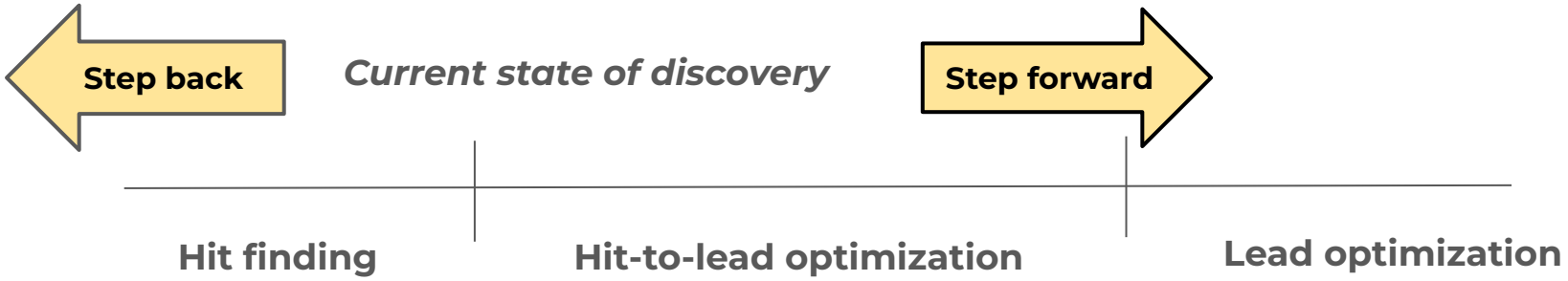


1. Estimate binding energy of each fragment, E_j
2. Stockholders' share $\omega_j = \frac{E_j}{\sum_{i=1}^{N} E_i}$
3. Scaled energy $E_j^{Scaled} = \omega_j \cdot \Delta E_{mol}$
4. Energy share of each fragment $Share_j = 100\% \cdot \omega_j$
5. Ligand efficiency of j-th fragment $\mathrm{LE(GE)} = E_j^{Scaled}/\mathrm{NH}$

NB $\sum_{i=1}^{N} E_i \neq \Delta E_{mol}$

R-FBDD: main use cases

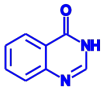
I. Consecutive hit-to-lead design: how to guess/choose a probable/fruitful scaffold binding mode ("**step forward**")?
II. Due diligence for a ligand: should we remove something to get a more promising molecule? ("**step back**")



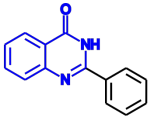
Shulga D. A., Ivanov N. N., & Palyulin V. A. (2021) // *Mend. Comm.*, *31*(3), 291-293, doi: 10.1016/j.mencom.2021.04.004

R-FBDD: use case I - "step forward"

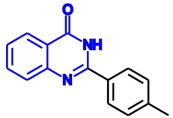
Retrospective design



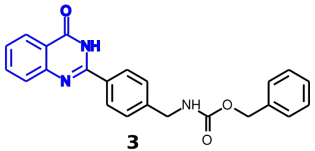
Scaffold of interest



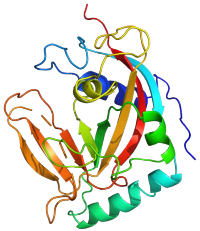
1
PDB ID: 4BU3
IC_{50}= 2200 nM

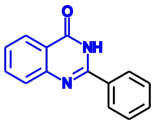


2
PDB ID: 4UHG
IC_{50}= 500 nM



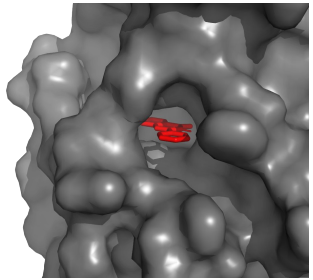
3
PDB ID: 4UI4
IC_{50}= 50 nM



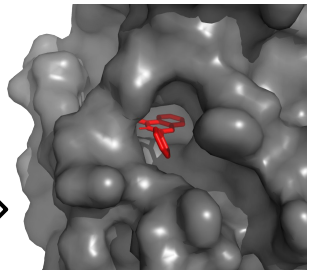


1

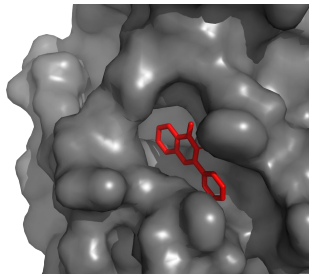
Docking



Mode 1



Mode 2



Mode 3

Human tankyrase 2, PARP5B, PDB 3U9Y

Energy contribution of the scaffold

Mode	Share, %	E, kcal/mol	LE, kcal*mol^{-1}*N^{-1}	E_{mol}, kcal/mol
1	69	-7.3	0.66	-10.6
2	64	-6.5	0.59	-10.2
3	60	-5.9	0.54	-9.8

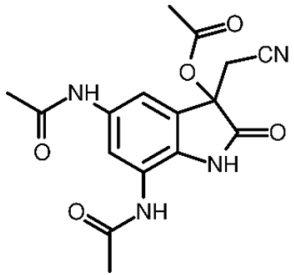
Shulga D. A., Ivanov N. N., & Palyulin V. A. (2021) // *Mend. Comm.*, *31*(3), 291-293, doi: 10.1016/j.mencom.2021.04.004

R-FBDD: use case I - “step forward”

Key takeaway for “**step forward**” use case:

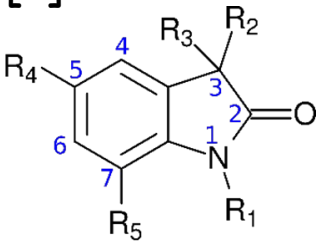
R-FBDD helped to select the anchoring binding modes of the scaffold in order to (retrospectively) streamline further development

R-FBDD: use case II - "step back"



4

[1]

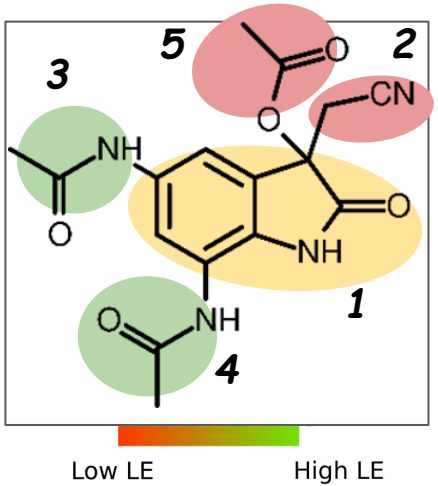


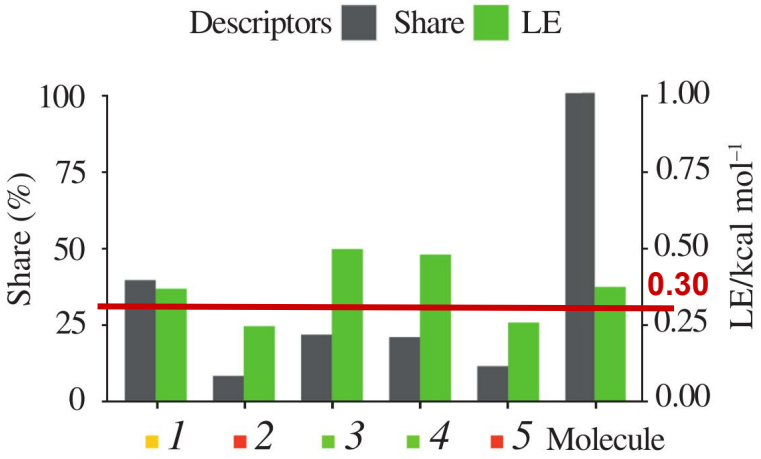
SBDD target:
QR2/MT3

Decide based on LE estimation for each group

Where to go next?!

[2] R-FBDD





1. Zaryanova, E. V., Lozinskaya, N. A., Beznos, O. V., Volkova, M. S., Chesnokova, N. B., & Zefirov, N. S. (2017) // *Bioorganic & med. chem. let.*, 27(16), 3787-3793, doi: 10.1016/j.bmcl.2017.06.065
2. Shulga D. A., Ivanov N. N., & Palyulin V. A. (2021) // *Mend. Comm.*, *31*(3), 291-293, doi: 10.1016/j.mencom.2021.04.004

R-FBDD: use case II - “step back”

Key takeaway for “**step back**” use case:

R-FBDD helped to select low LE substituents to simplify the structure and restart design with smaller entire compounds with higher LE values

What we have so far with R-FBDD beyond the method itself?

R-FBDD: the means for "scaffold-lean" discovery

Lean project management (~Agile)

- systematically deliver more value with less waste

Essence: minimize **risks** of an **entire** (complex) **project** by **iteratively** bringing the most **value** using **least resources** (~minimize **waste**)

Scaffold-lean discovery

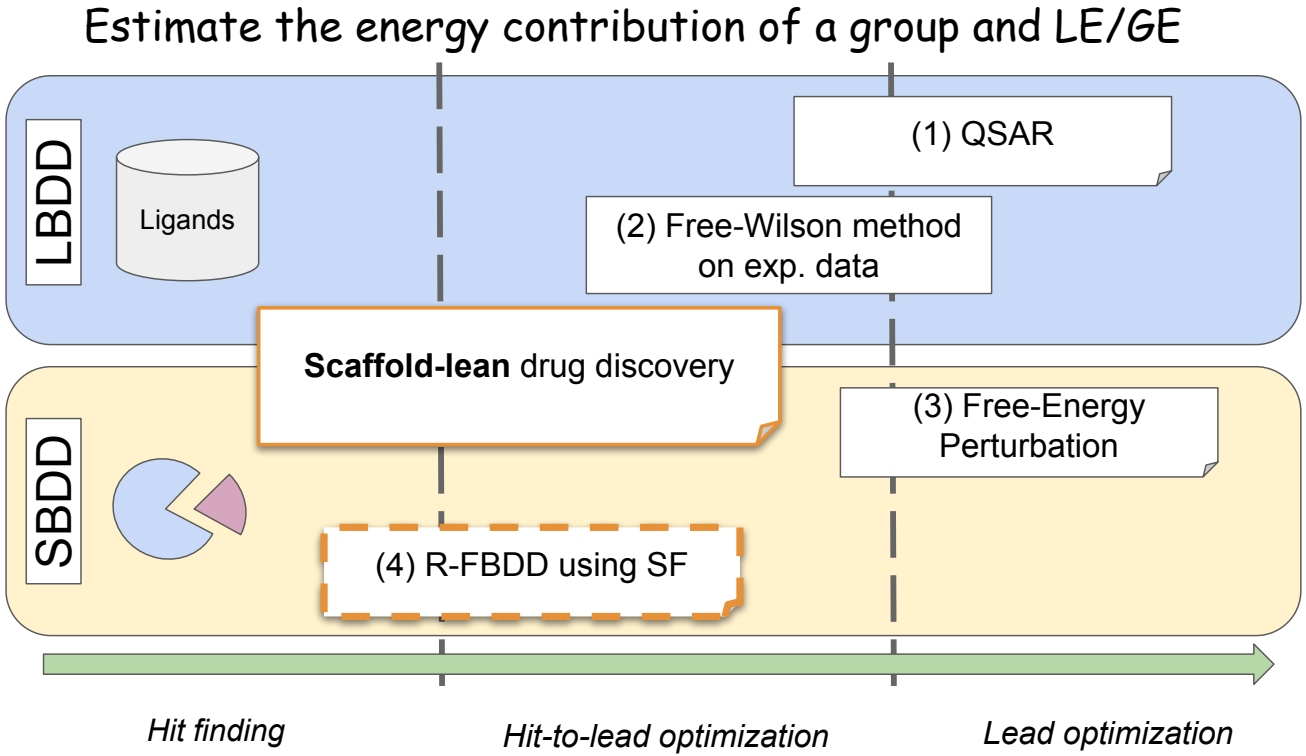
- systematically get the maximum value out of the scaffold
- scaffold - the basis for intellectual property

Essence: minimize **risks** of an **entire** (complex) **DD project** by **iteratively** bringing the most **value** using **pertinent scaffolds**

Beneficiaries:

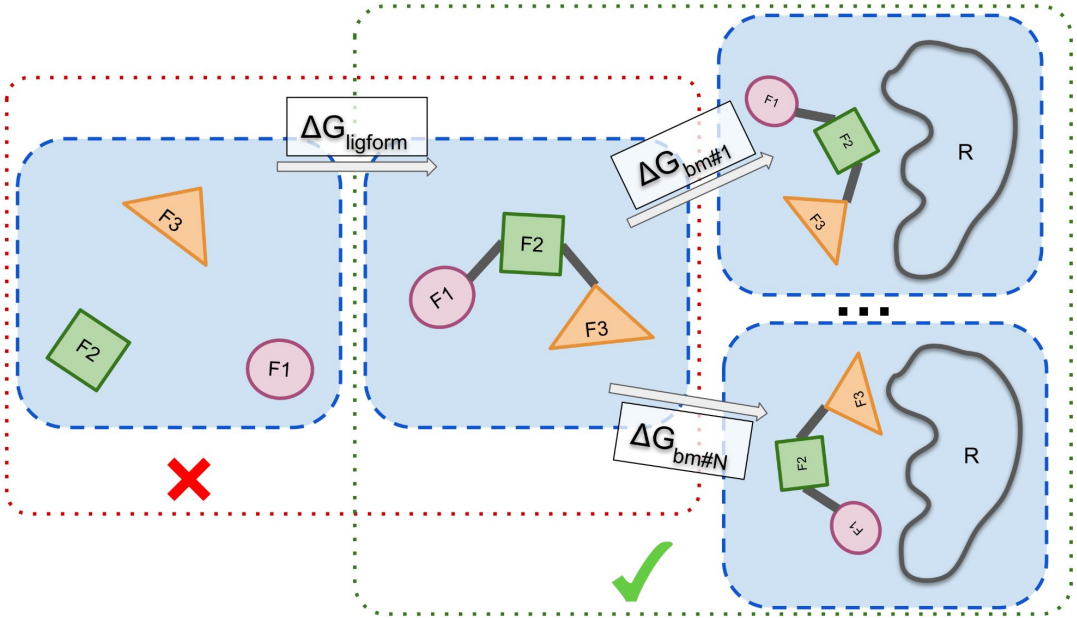
1. academic labs - scaffold is a core competence
2. CRO - the fewer scaffold switching, the faster and cheaper the projects are
3. Big Pharma labs - better IP planning

R-FBDD: positioning among other in silico approaches



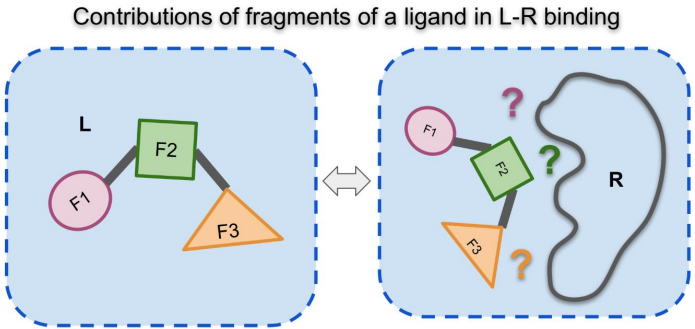
Shulga D. A., Ivanov N. N., & Palyulin V. A. (2022) // *Molecules*, *27*(6), 1985., doi: 10.3390/molecules27061985

R-FBDD: fragment-in-molecule-in-a-position



Main question: how each fragment interacts with the receptor:

A. within the molecule
B. in a specific arrangement
C. in a specific position of the molecule at a target?



$$P_L(binding|receptor = R, position = site_i)$$

Probabilistic view

Reported at BCADD2022

R-FBDD: current achievements

Outline

1. Estimation of LE for groups - group efficiencies (GE)
2. Use of different scoring function to partition energy to contributions
3. Splitting the ligands in a medicinal chemist meaningful way
   a. MedChemFrag
   b. Access hard cases of conjugated molecules
4. Statistical mechanical representation of ligand-receptor complex within R-FBDD (using MM/GBSA)
5. Prospective applied projects guided by R-FBDD… WIP

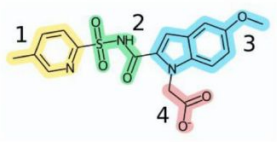
R-FBDD: current achievements

Hung, A. W., Silvestre, H. L., Wen, S., ... & Abell, C. (2016). // *ChemMedChem*, *11*(1), 38-42.: doi: 10.1002/cmdc.201500414

Shulga D. A., Ivanov N. N., & Palyulin V. A. (2022) // *Molecules*, *27*(6), 1985., doi: 10.3390/molecules27061985

Optimization of Inhibitors of *Mycobacterium tuberculosis* Pantothenate Synthetase Based on Group Efficiency Analysis

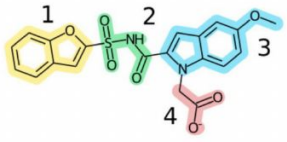
Our work



1

Fragments

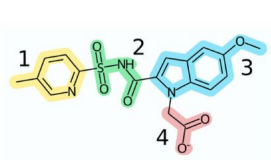
Compound	3	2	1	4	ΔG, kcal/mol
(1)	x				−4.0
(3)	x	x			−5.0
(4)	x	x	x		−6.2
1/(5)	x	x	x	x	−7.9
ΔΔG, kcal/mol	−4.0	−1.0	−1.2	−1.7	
NH	11	6	7	4	
GE	0.36	0.17	0.17	0.43	



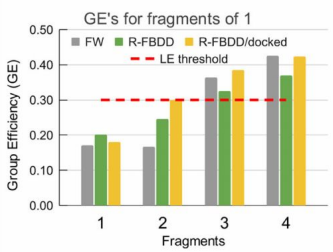
2

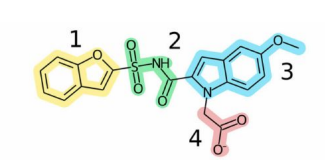
Fragments

Compound	3	2	1	4	ΔG, kcal/mol
(1)	x				−4.0
(3)	x	x			−5.0
(8a)	x	x	x		−6.4
2/(8)	x	x	x	x	−7.8
ΔΔG, kcal/mol	−4.0	−1.0	−1.4	−1.4	
NH	11	6	9	4	
GE	0.36	0.17	0.16	0.35	

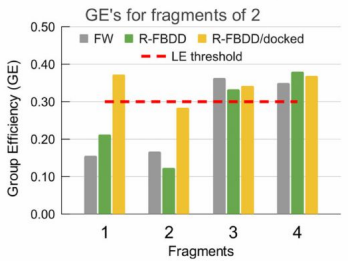


1





2



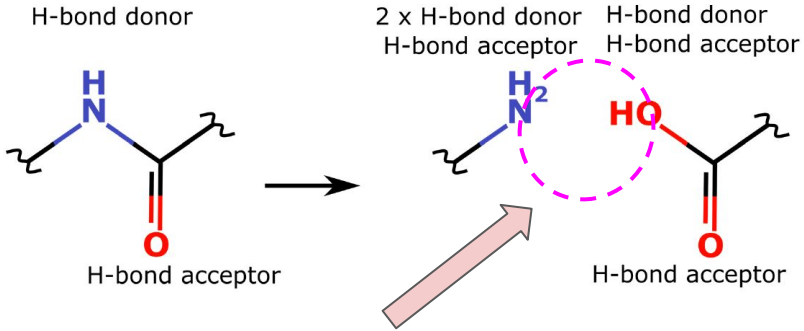
Experimental GE

GE estimated with R-FBDD

R-FBDD: current achievements

Goal: automatic proper breakdown into fragments which preserve interaction patterns as in entire ligands for R-FBDD

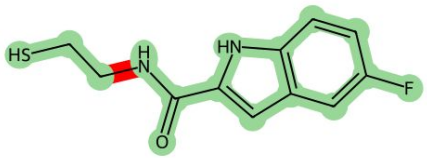
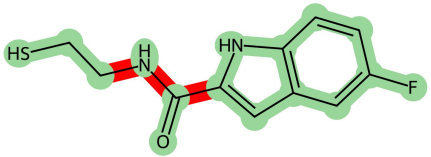
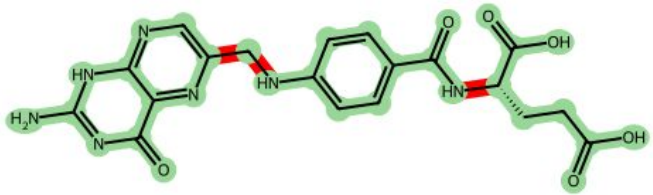
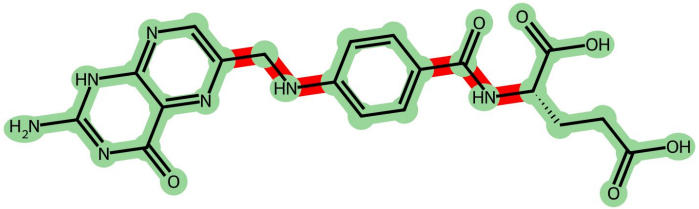
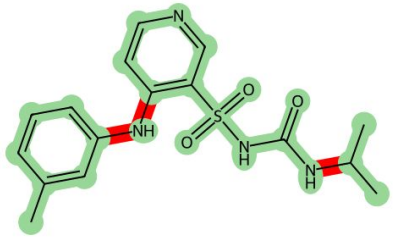
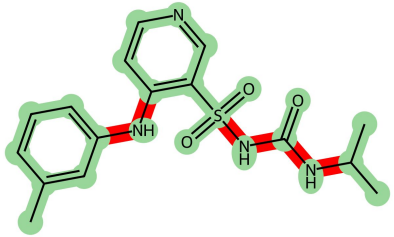
But: most existing molecular splitters are based on retrosynthetic schemes (e.g. BRICS and RECAP)!



new interaction pattern

1. Ivanov, N. N., Shulga, D. A., & Palyulin, V. A. (2023) // Biophysica, 3(2), 362-372., doi: 10.3390/biophysica3020024

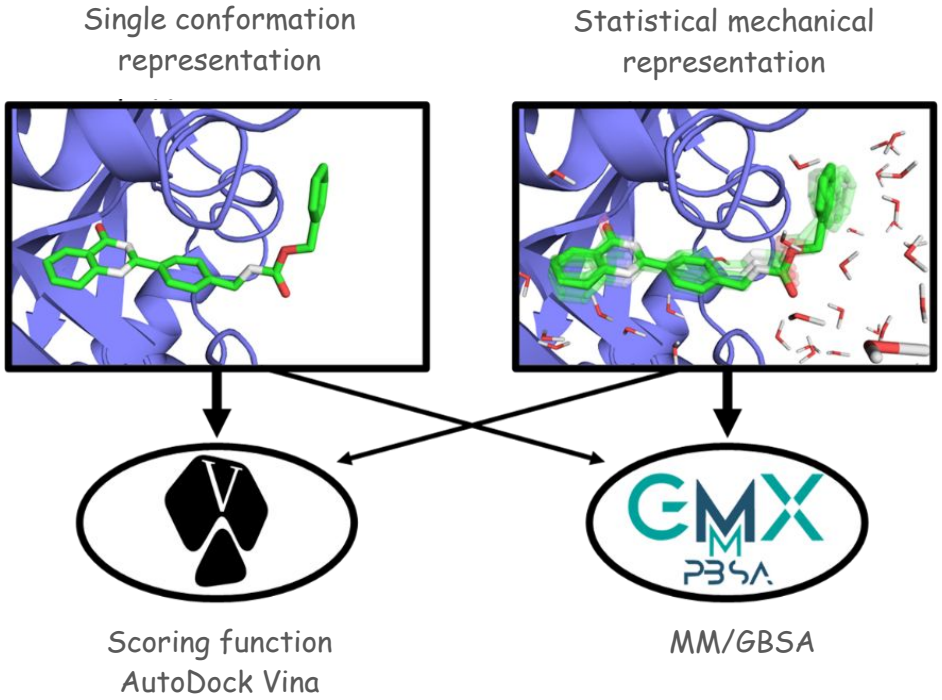
R-FBDD: current achievements

	MedChemFrag	BRICS
DB03384		
DB00158		
DB00214		

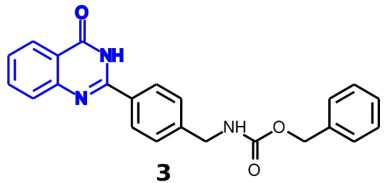
1. Ivanov, N. N., Shulga, D. A., & Palyulin, V. A. (2023) // Biophysica, 3(2), 362-372., doi: 10.3390/biophysica3020024

R-FBDD: current achievements

Statistical mechanical representation - R-FBDD within MM/GBSA context



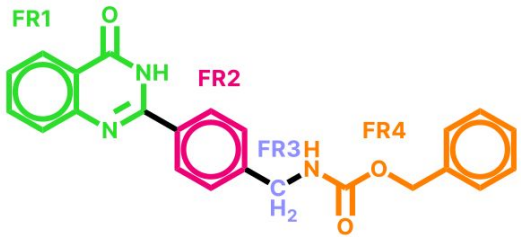
Human tankyrase 2, PARP5B, PDB 3U9Y



3

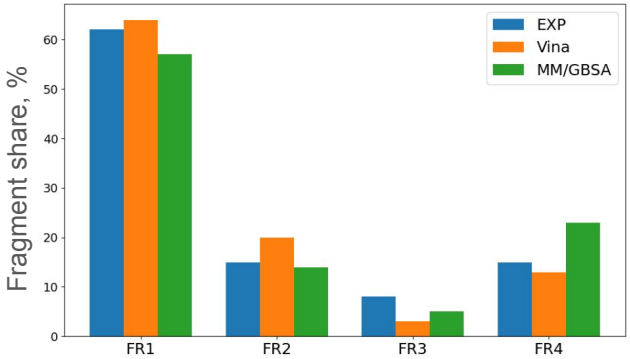
PDB ID: 4UI4

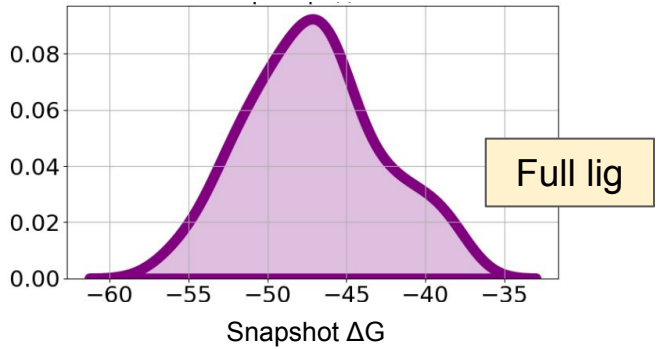
IC_{50}= 50 nM



R-FBDD: current achievements

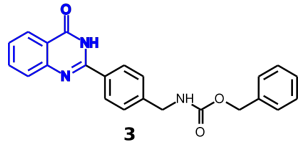
Statistical mechanical representation - R-FBDD within MM/GBSA context



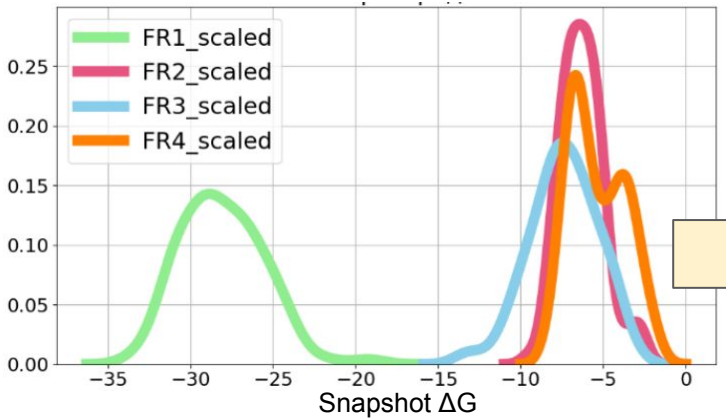


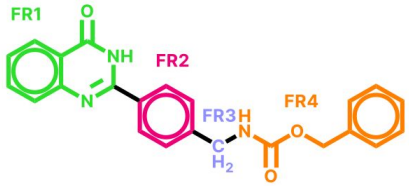
Full lig

Human tankyrase 2,
PARP5B, PDB 3U9Y



PDB ID: 4UI4
IC_{50}= 50 nM





Fragments

R-FBDD: prospects

1. **Scaffold-lean** drug discovery - we introduce a new term
    a. for academic groups
    b. for CRO
    c. for Big Pharma

R-FBDD: prospects

2. MedChemFrag
    a. medchem relevant splitting preserving interaction pattern
    b. useful for fragment interaction analysis
    c. useful for numerous AI/ML/DL methods
3. R-FBDD is excellent to use within more complex and specific pipelines
    a. we already use it for several applied (prospective) projects, wait for the papers…
4. R-FBDD for
    a. hit finding
    b. hit-to-lead optimization

R-FBDD: repositories

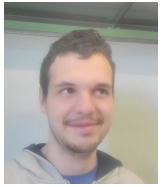
1. R-FBDD
   - https://molmodel.com/hg/CalcFragsContribution
2. MedChemFrag
   - https://molmodel.com/hg/medchem_fragment_splitter/
3. R-FBDD for MM/GBSA
   - to be published soon…

Acknowledgements

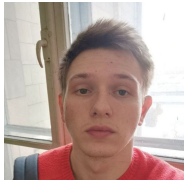
Nikita Ivanov - the main author/developer - PhD to be defended soon…

Arslan Shaimardanov - SF/electrostatic interaction contribution - PhD to be defended soon…

Vladimir Palyulin - general supervision, scientific expertise



Vasily Myazin - graduate student - use of R-FBDD in hit finding



Mikhail Rybkin - graduate student - use of R-FBDD in hit-to-lead opt

Daniil Chernov - MS 2025 - R-FBDD within MM/GBSA

Ilyas Zhanataev - MS 2025 - MEP based splitting of conjugated fragments

“Modes” of collaboration

academic

Dmitry A. Shulga,
PhD, MBA

applied/contract

basement

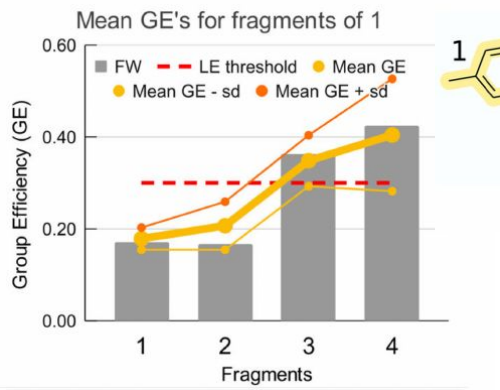
Table 1. Pros, cons and applicability domain of different fragment contribution methods.

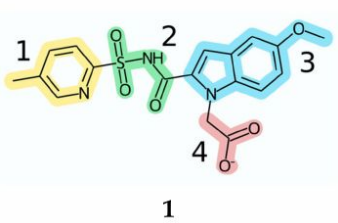
Method	(1) G-QSAR, QSAR Interpretation Methods	(2) Experimental Affinity Free-Wilson Partitioning	(3) FEP Estimation of Fragment Addition	(4) Fragment Contribution Based on Scoring Function
Essence	Extract contribution of each fragment via statistical analysis of its influence on the predicted activity	Extract fragment contribution from a minimal set of structurally nested ligands with known experimental activity	The simulated free energy differences between a pair of ligands are used to estimate the contribution brought by the changed group	Extract fragment contribution from a scoring function estimation of each fragment contribution, provided a relevant ligand-receptor complex geometry is available
Pros	The most robust and statistically significant Non-additive dependencies can be studied Ligand-based, no need to rely on receptor structure Fragment contributions can be calculated prospectively for new structures but within the applicability domain of the underlying QSAR models	The least possible efforts can be made in terms of synthesis and activity measurement of a small series of structural analogs The activities of the structures are valuable by themselves and can be rationalized even without building a model—by a qualitative SAR analysis	Potentially takes into account many dynamic and statistical effects associated with binding Only physics-based models (force fields or QM/MM) are used—no reliance on any additional mathematical models Additional statistical metrics from the MD trajectory regarding ligand-receptor interactions No reliance on any chemical class	"Cold start"—no need to synthetize and measure the experimental activity of any structure Several feasible binding modes can be considered for further ligand optimization Often clear interpretation of the nature of the fragment-receptor interactions No dependence on congeneric series of ligand—any chemical structure suits
Cons	An extended series of ligands with established activity is needed to ensure statistically significant results New structures, for which partitioning into fragment contribution is made, should be within the applicability domain of the underlying QSAR models	Partitioning of affinity in fragment contribution is done *retrospectively*, i.e., only *after* the structures are synthetized and their activity is measured	A steep learning curve for beginners in the field due to the inherent complexity Dedicated computational resources, often requiring high end CPU/GPU	The relevant geometry of the target needs to be known Reliance on the geometry of ligand-receptor complex Moderate accuracy of scoring functions Dependence of activity estimation on structural choices: account of crystallization waters, protonation states of receptor's residues and a ligand, etc.
Applicability	The most applicable to lead optimization, when an extended series of different structures with known experimental activity is known	The most applicable to the hit-to-lead stage, where a fragment-sized hit is expanded to a drug-sized molecule. The FW scheme should be used to plan the synthesis of a minimal series of analogs	The most applicable to the lead optimization stage, where the chemical modifications are generally small and the binding mode is well established. Can be applied at hit-to-lead stage also by the experienced teams	The most applicable at early stages of hit finding and/or hit-to-lead, when a reliable/relevant structure/conformation of the target is known.
Principle	Ligand-based drug design (LBDD)	Ligand-based drug design (LBDD)	Structure-based drug design (SBDD)	Structure-based drug design (SBDD)

https://doi.org/10.3390/molecules27061985

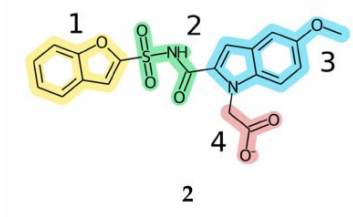
GE with different scoring functions

Scoring Function	Source	Classification	References
AutoDock4	AutoDock v4.2.6	Physics-based function	[38]
AutoDock4	AutoDock Vina v1.2.3	Physics-based function	[38,42]
AutoDock Vina	AutoDock Vina v1.1.2	Empirical scoring function	[41]
AutoDock Vina	AutoDock Vina v1.2.3	Empirical scoring function	[42]
Vinardo	AutoDock Vina v1.2.3	Empirical scoring function	[42,43]
DSX	DrugScoreX v0.9	Knowledge-based potential	[44]
$\Delta_{Vina}RF_{20}$	https://github.com/chengwang88/deltavina (accessed on 19 February 2022)	Machine-learning model	[45]





1



2