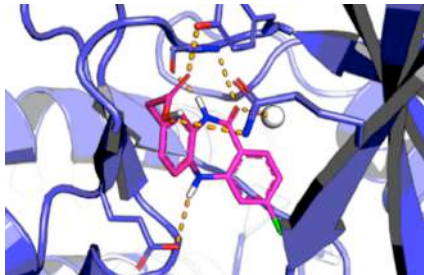
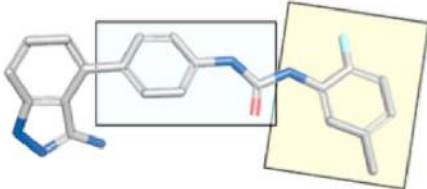



ХИМИЧЕСКАЯ ИНФОРМАТИКА: ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

Лекции 1-2

Подготовка баз данных, дизайн библиотек соединений

COMPUTER-AIDED DRUG DESIGN (CADD)

	Лиганды известны	Лиганды неизвестны
Структура белка известна	<p>Structure-based drug design (SBDD)</p>  <p>A 3D molecular model showing a protein structure in blue and grey, with a pink and red ligand molecule docked into its binding pocket. Dotted lines indicate interactions between the ligand and the protein.</p>	<p><i>De novo</i> design</p>  <p>A diagram illustrating the de novo design process. It shows a target molecule (a benzene ring with a blue substituent) being broken down into fragments (a benzene ring and a blue substituent) which are then recombined to form the target molecule.</p>
Структура белка неизвестна	<p>Ligand-based drug design (LBDD)</p>  <p>A 3D molecular model of a ligand molecule (a benzene ring with a red substituent) surrounded by a semi-transparent grey surface representing its molecular surface. The surface is colored with red and blue, indicating different chemical environments.</p>	<p>Experimental data required</p>

ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И
ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА:
“FAIL EARLY, FAIL FAST, FAIL CHEAP”

ТЕРМИНОЛОГИЯ

Соединение-хит – соединение, обнаруженное в ходе тестирования на биологическую активность

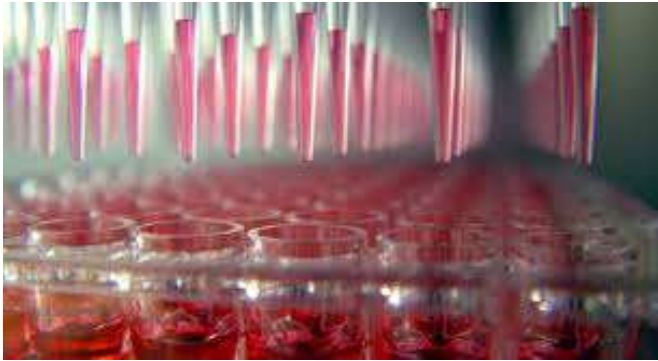
Соединение-лидер – соединение, проявляющее желаемую биологическую активность, отобранное среди соединений хитов, структурный прототип будущего лекарства

Библиотека соединений – коллекция соединений

Химическое пространство данных – совокупность химических соединений, заданное при помощи дескрипторного описания

Комбинаторная химия – генерация большого числа химических соединений систематическим комбинированием отдельных подструктур, часто при помощи библиотек фрагментов

ВЫСОКОПРОИЗВОДИТЕЛЬНЫЙ СКРИНИНГ



Возможность быстрого скрининга (до 100 тысяч соединений ежедневно)

Надёжность получаемых результатов

Высокая эффективность при корректной подготовке библиотек

Труднодоступность для академических исследователей и небольших фармакомпаний (высокая стоимость)

Разработка автоматизированных методик не всегда возможна;

Не все мишени доступны для подобного скрининга

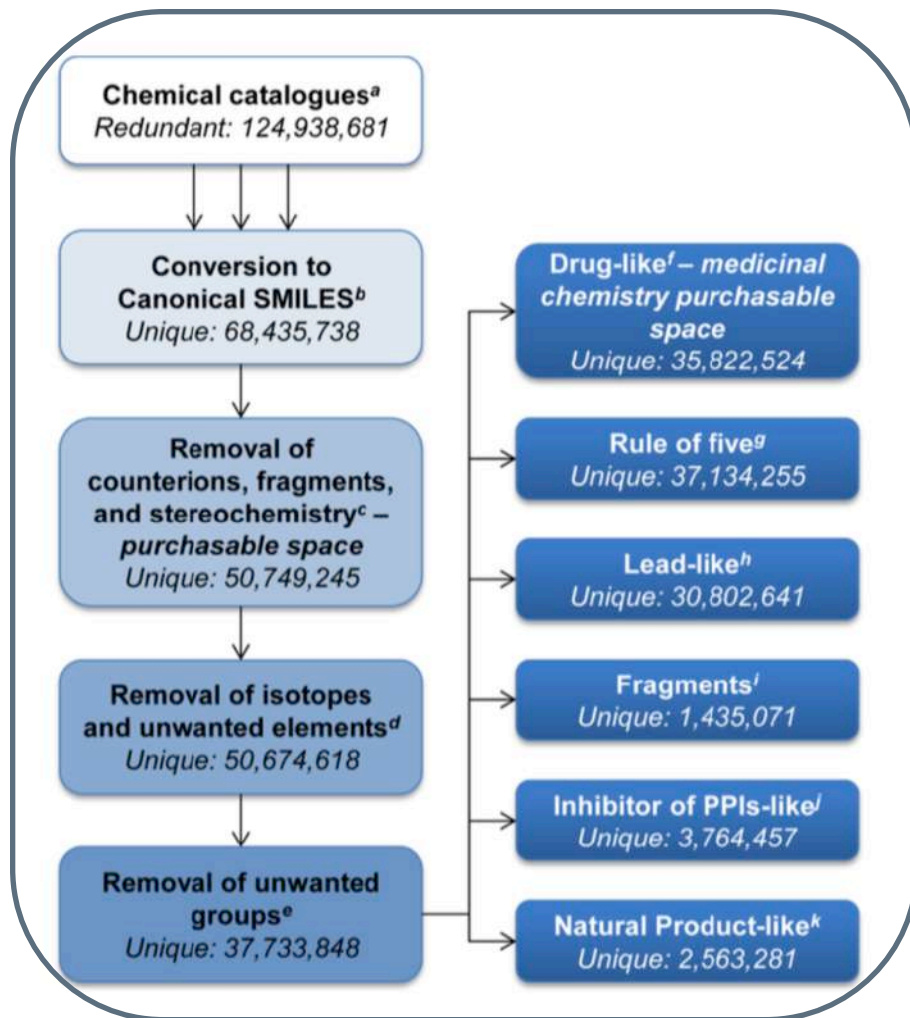
Нерационально использование для поискового скрининга

Низкая вероятность обнаружения активных соединений в «слепом» варианте метода (без предварительной подготовки выборки).

АНАЛИЗ ХИМИЧЕСКОГО ПРОСТРАНСТВА СОЕДИНЕНИЙ

Number of compounds	Limitations			Method
	Size	Composition	Other	
$6,2 \times 10^{13}$	≤ 40 atoms*	C, H	Acyclic alkanes without stereoisomers	Exhaustive enumeration
$1,3 \times 10^{15}$	≤ 38 atoms*	C, H	Acyclic stereoisomeric alkanes	Exhaustive enumeration
10^{21}	$< 7 \text{ \AA}$	40 functional groups	Neurological drugs	Combinatorial enumeration
10^{23}	≤ 36 atoms	C, N, O, S, P, Se, Si, Hal	Scaffold with 2 or 3 attachment points	Combinatorial estimation
10^{26}	≤ 50 atoms	C, N, O, S, Cl	–	Combinatorial enumeration
10^{33}	≤ 750 Da	C, N, O, F	Heptanes and hexanes including stereoisomers	Combinatorial enumeration
10^{33}	≤ 36 atoms, ≤ 500 Da	C, N, O, S, Hal	Stable compounds (stereoisomers are not taken into account)	Learning of exhaustively enumerated structures from GDB-17
10^{60}	≤ 30 atoms	C, N, O, S	–	Combinatorial enumeration
10^{100}	N/A	N/A	N/A	No clear explanations
10^{180}	≤ 1000 Da	C, N, O, P, S, Hal	With stereoisomers counted	No clear explanations

THE PURCHASABLE CHEMICAL SPACE: A DETAILED PICTURE

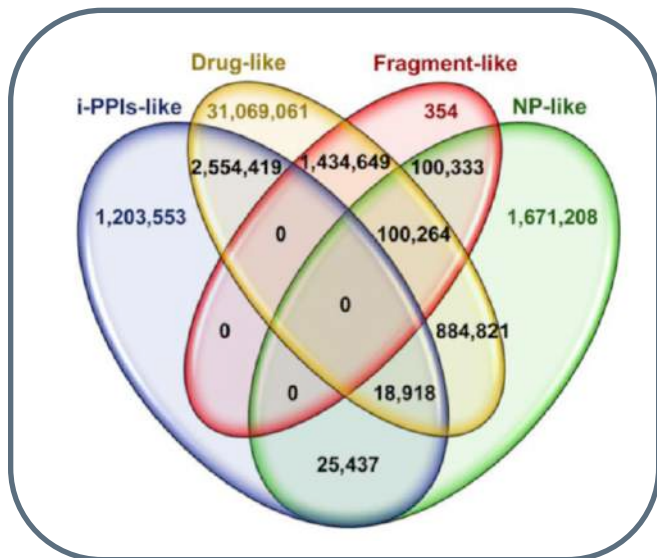


The Purchasable Chemical Space: A Detailed Picture

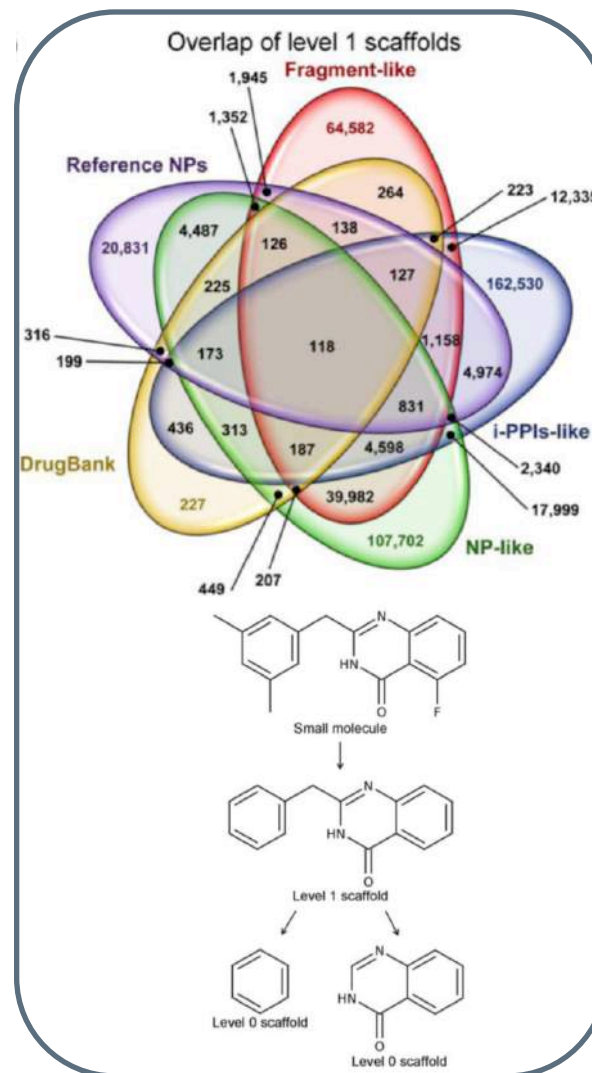
X. Lucas et al *Journal of Chemical Information and Modeling* 2015 55 (5), 915-924 DOI: 10.1021/acs.jcim.5b00116

THE PURCHASABLE CHEMICAL SPACE: A DETAILED PICTURE

Уникальные и перекрывающиеся соединения



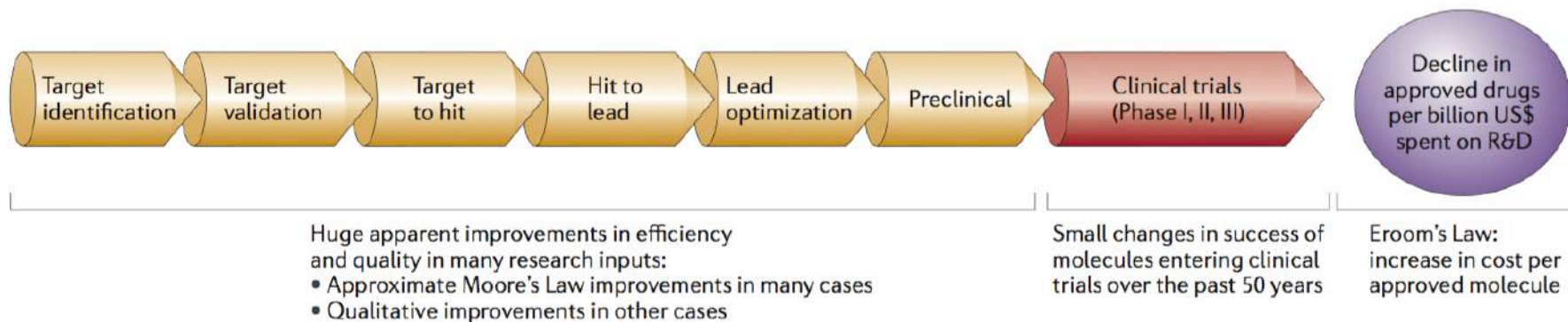
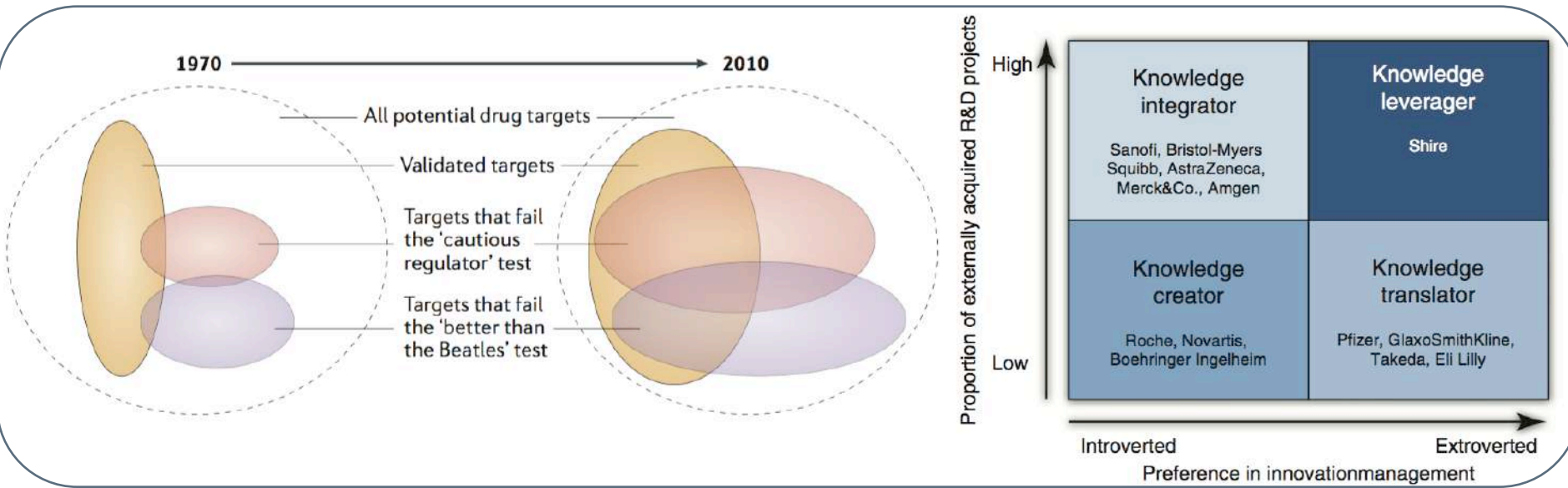
Молекулярные скелеты



The Purchasable Chemical Space: A Detailed Picture

X. Lucas et al *Journal of Chemical Information and Modeling* 2015 55 (5), 915-924 DOI: 10.1021/acs.jcim.5b00116

АНАЛИЗ ХИМИЧЕСКОГО ПРОСТРАНСТВА СОЕДИНЕНИЙ



Scannell et al Diagnosing the decline in pharmaceutical R&D efficiency Nature Reviews Drug Discovery 11, 191-200 (March 2012)

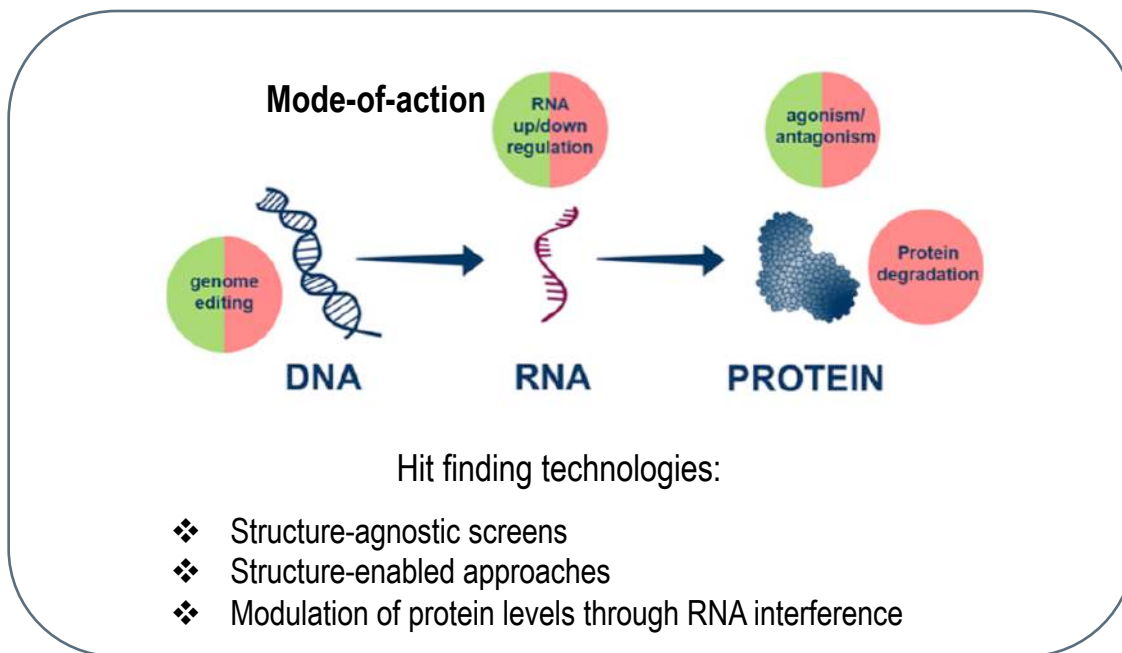
Schuhmacher et al Models for open innovation in the pharmaceutical industry Drug Discovery Today (2013) 18, 1133

АНАЛИЗ ХИМИЧЕСКОГО ПРОСТРАНСТВА СОЕДИНЕНИЙ



- Поиск среди мишеней прежде считавшихся "undruggable"
- Использование технологий нового поколения

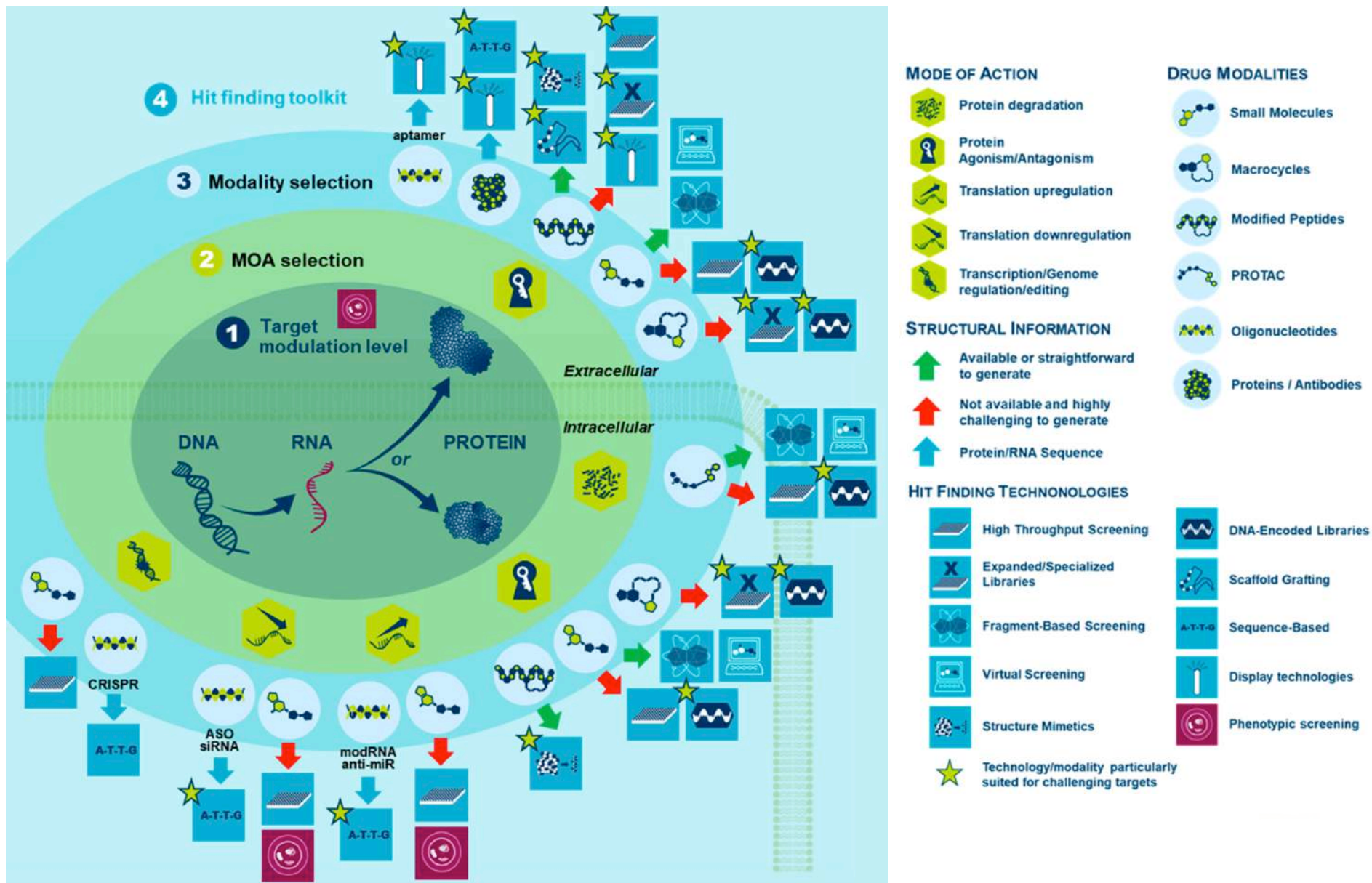
МЕХАНИЗМ ДЕЙСТВИЯ (MODE-OF-ACTION) КАК ОСНОВА ОПРЕДЕЛЕНИЯ СТРАТЕГИИ СКРИНИНГА НОВЫХ СОЕДИНЕНИЙ



Examples of possible cheminformatics support:

- ❖ Aid new technologies (e.g. DNA-encoded libraries DEL)
- ❖ Data sharing and analysis
- ❖ Integrating new machine learning technologies (Deep learning, one-shot-learning, machine learning for dynamic systems, etc), rational use of already known approaches

MODE-OF-ACTION CHOICE AS A STARTING POINT FOR SCREENING



E. Valeur et al **New Modalities, Technologies, and Partnerships in Probe and Lead Generation: Enabling a Mode-of-Action Centric Paradigm**

J Med Chem 2018, 61 (20), pp 9004–9029

ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА



ПОНЯТИЯ DRUGLIKENESS И LEADLIKENESS

"Правило пяти" Липинского

Биологическая активность вещества (при приеме внутрь) **менее вероятна, если:**

- ❖ Молекулярный вес более 500.
- ❖ $\log P$ более 5.
- ❖ Число доноров водорода водородной связи более 5 (определяют по сумме OH- и NH-групп).
- ❖ Число акцепторов водорода водородной связи более 10 (определяют по сумме атомов O и N).
- ❖ У 70% веществ, имеющих признаки биологической активности: 0-2 донора водородной связи, 2-9 акцепторов водородной связи, 2-8 конформационных связей, 1-4 цикла.
- ❖ Площадь полярной поверхности более 150 \AA^2
- ❖ Число конформационных связей более 10

Подобие соединению-лидеру («Lead-Likeness»)

- ❖ Молекулярный вес менее 350.
- ❖ $\log P$ менее 3.
- ❖ Число доноров водорода водородной связи менее 3 (определяют по сумме OH- и NH-групп).
- ❖ Число акцепторов водорода водородной связи менее 8 (определяют по сумме атомов O и N).
- ❖ Площадь полярной поверхности менее 120 \AA^2
- ❖ Число конформационных связей менее 8

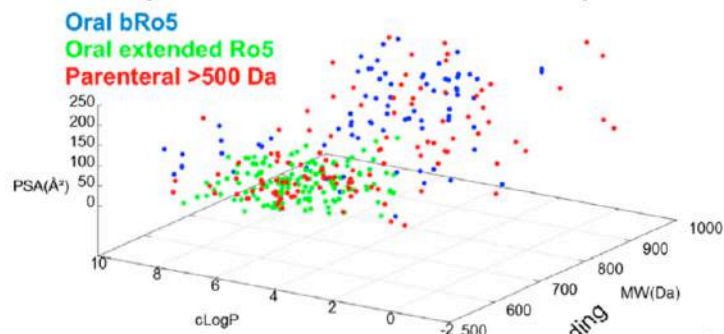
Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates

Physicochemical property distribution of chemical classes

Class	Delivery	N	MW (Da)	cLogP	HBD	HBA	PSA (Å ²)	NRotB	Fsp ³
all	Oral	226	642(621-662)	4.1(3.8-4.3)	2.6(2.4-2.9)	7.7(7.3-8.1)	137(130-145)	10.4(9.8-11.1)	0.49(0.46-0.52)
	Parenteral	259	1059(983-1136)	-1.4(-2.3- -0.5)	10.5(9.2-11.8)	15.6(14.2-17.0)	357(320-394)	21.9(19.3-24.6)	0.52(0.50-0.54)
peptides & mimetics	Oral	25	815(700-929)	3.2(2.6-3.9)	4.5(3.5-5.5)	8.7(7.6-9.9)	199(169-229)	15.1(12.5-17.7)	0.56(0.50-0.62)
	Parenteral	100	1473(1336-1610)	-6.9(-8.2- -5.6)	19.3(17.1-21.5)	22.1(19.7-24.5)	582(518-645)	35.6(30.2-40.9)	0.56(0.54-0.58)
natural products	Oral	51	713(670-756)	3.7(3.0-4.3)	3.3(2.8-3.8)	10.2(9.2-11.2)	168(153-182)	9.2(8.0-10.5)	0.68(0.62-0.73)
	Parenteral	73	883(765-1002)	0.7(-0.6-2.0)	6.8(5.1-8.4)	14.3(11.8-16.9)	270(219-322)	13.6(10.1-17.2)	0.59(0.54-0.63)
de novo designed	Oral	136	590(574-606)	4.4(4.1-4.7)	2.1(1.8-2.3)	6.6(6.2-6.9)	113(107-120)	9.3(8.8-9.9)	0.40(0.37-0.43)
	Parenteral	75	733(662-803)	3.0(2.1-3.9)	3.7(2.8-4.6)	9.4(7.8-11.0)	176(146-206)	12.2(10.5-14.0)	0.39(0.34-0.43)
prodrugs	Oral	14	572(534-611)	3.8(2.2-5.4)	2.1(1.6-2.7)	7.6(6.6-8.6)	149(135-164)	17.1(13.6-20.6)	0.62(0.49-0.74)
	Parenteral	11	699(558-840)	4.6(1.6-7.6)	2.2(0.3-4.0)	7.2(5.2-9.2)	126(85-168)	19.6(14.9-24.2)	0.60(0.53-0.66)
CNS	Oral	6	601(474-727)	4.4(2.1-6.7)	0.8(-0.6-2.2)	5.0(3.9-6.2)	81(56-106)	8.8(5.6-12.1)	0.42(0.36-0.49)

Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates

A Expanded oral bRo5 chemical space



"possible to be oral" chemical space ~

MW ≤ 1000 Da, -2 ≤ cLogP ≤ 10,

HBD ≤ 6, HBA ≤ 15,

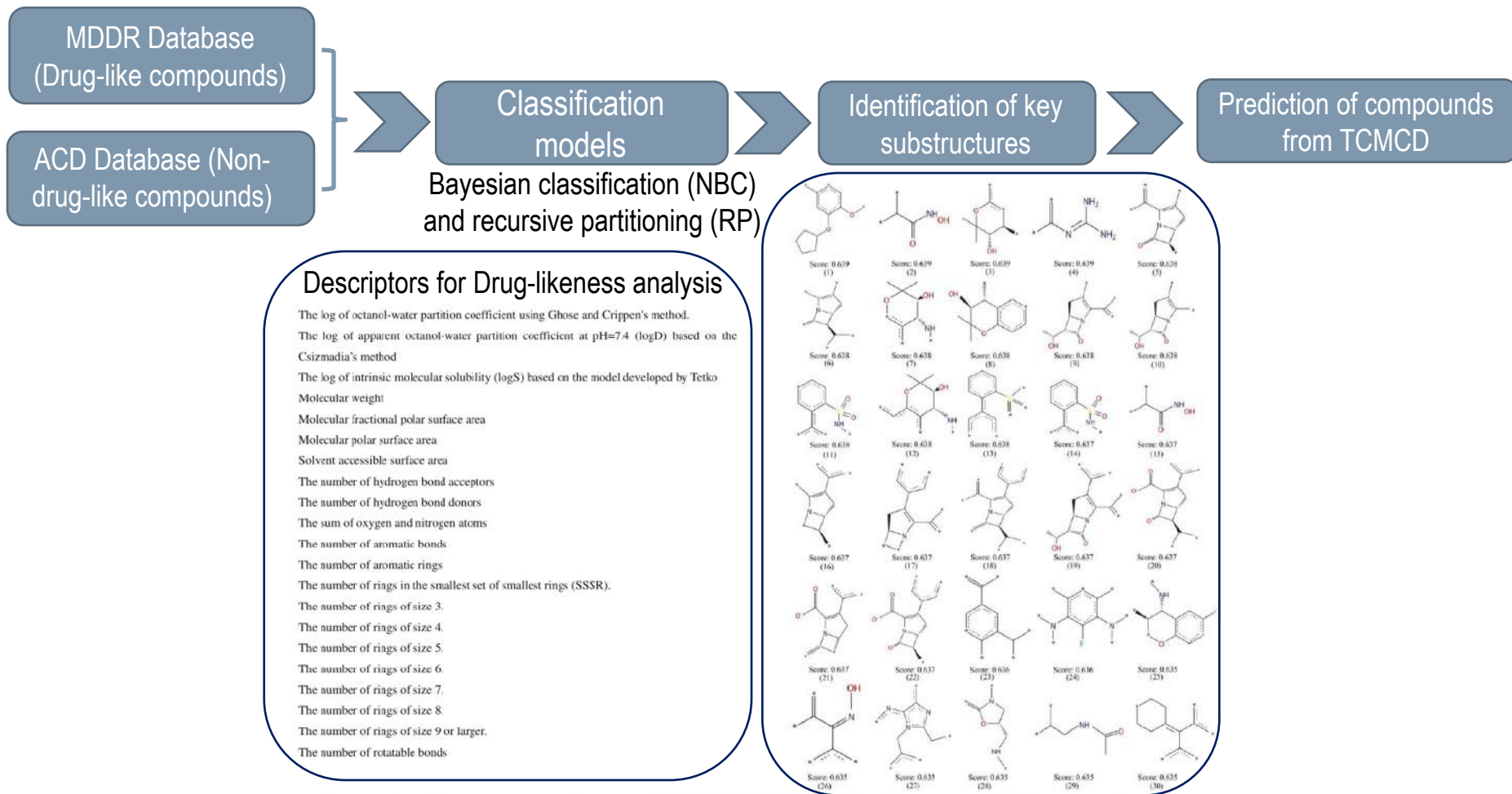
PSA ≤ 250 Å², NRotB ≤ 20

Intramolecular hydrogen bonding
macrocyclic members
high dose
transporter efflux
formulation
transporter distribution effect

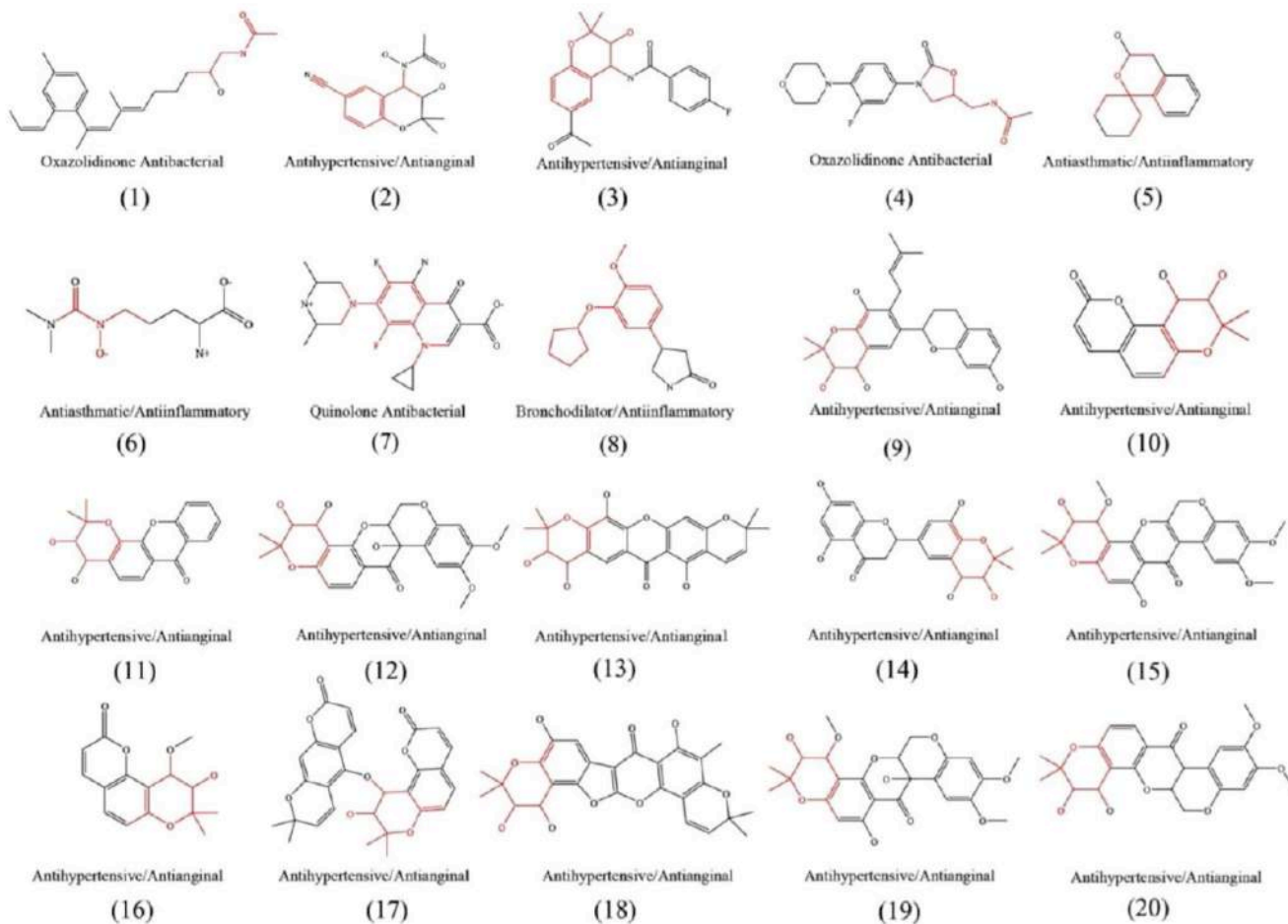
B

Major oral bRo5 Class	Intramolecular hydrogen bonding	macrocyclic members	high dose	transporter efflux	formulation	Origin
Erythronolides	✓	✓	✓	✓		natural product
Leucomycins	✓	✓	✓	✓		natural product
Rifamycins	✓	✓	✓	✓	✓	natural product
HCV NS3/4A protease inhibitors	✓	✓	✓	✓	✓	peptide
HCV NS5A inhibitors	✓			✓		HTS
HIV-1 protease inhibitors	✓		✓	✓		peptide
Ascomycins and Rapamycins		✓		✓	✓	natural product
Cyclosporins	✓	✓	✓	✓	✓	natural product
Azoles			✓	✓	✓	<i>de novo</i>
Taxanes			✓	✓	✓	natural product
Cardiac glycosides				✓		natural product
Lipid mimetic prodrugs					✓	<i>de novo</i>




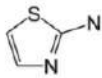
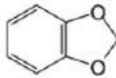
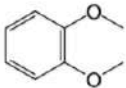






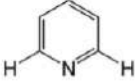
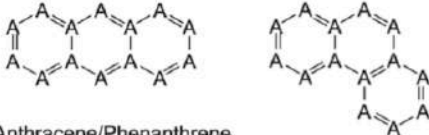
DRUG-LIKENESS ANALYSIS OF TRADITIONAL CHINESE MEDICINES: PREDICTION OF DRUG-LIKENESS USING MACHINE LEARNING APPROACHES



DRUG-LIKENESS ANALYSIS OF TRADITIONAL CHINESE MEDICINES: PREDICTION OF DRUG-LIKENESS USING MACHINE LEARNING APPROACHES



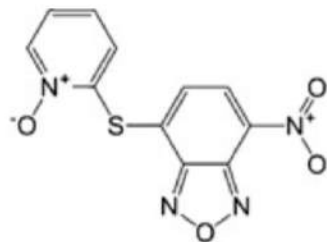
ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА: ПОДСТРУКТУРНЫЕ ФИЛЬТРЫ

 1,2 Dicarbonyl <chem>[#6](=O)[#6](=O)</chem>	 $\alpha\beta$ -Unsaturated Carbonyl <chem>C=CC(=O)</chem>	 Alkene <chem>C=&l@C</chem>	 Aminothiazole <chem>c1sc(nc1)N</chem>	 Methylenedioxy <chem>C1OCCO1</chem>
 1,2 Dimethoxy <chem>COCCOC</chem>	 1,4 Dimethoxy <chem>COCCOC</chem>	 Acylhydrazide <chem>C(=O)NN</chem>	 Acetal <chem>OCO</chem>	 Thiourea <chem>NC(=S)N</chem>
 Nitro <chem>[N:\$(N(=O)~[O;H])]</chem>	 Aliphatic Ketone <chem>CC(=O)C</chem>	 Unflanked Pyridyl <chem>c1(c(nc(c1)[H]))[H]</chem>	 Anthracene/Phenanthrene <chem>[a:\$(aa[R3](a)a[R3](a)a),(aa[R3](a)aa[R3](a)a)]</chem>	

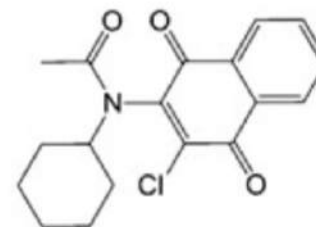
- ❖ 1,2-dicarbonyl: metabolically unstable/potential toxicity due to mutagenicity
- ❖ 1,2-dimethoxy: prone to oxidation yielding reactive quinones.
- ❖ 1,4-dimethoxy: very prone to oxidation yielding reactive quinones
- ❖ $\alpha\beta$ -unsaturated carbonyl: prone to reactivity by acting as a Michael acceptor
- ❖ acetal: metabolically unstable due to acetal hydrolysis acylhydrazide: metabolically unstable due to acyl hydrolysis aliphatic ketone: metabolically unstable due to nucleophilic attack
- ❖ alkene: metabolically unstable due to epoxidation aminothiazole: potential toxicity anthracene/phenanthrene-like: known DNA intercalation nitro group: prone to reduction yielding reactive species/ potential hepatocarcinogens
- ❖ methylenedioxy: metabolically unstable due to acetal hydrolysis/prone to oxidation yielding reactive quinones
- ❖ thiourea: metabolically unstable due to flavin oxidation/ potential on specific protein binding
- ❖ unflanked pyridyl: potential interference with cytochrome P450s due to metal ion coordination

ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА: СОЕДИНЕНИЯ С НЕСПЕЦИФИЧЕСКИМ ВЗАИМОДЕЙСТВИЕМ

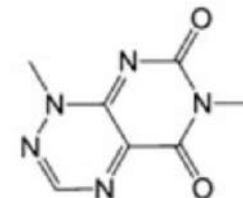
# targets with reported activity	# active compounds	Ratio
1	68,809	49.11%
2	29,022	20.71%
3	16,272	11.61%
4	9810	7.00%
5	5566	3.97%
6-10	8596	6.14%
11-15	1473	1.05%
16-20	404	0.29%
21-25	121	0.09%
26-30	32	0.02%
31-35	6	0.004%
36-40	1	0.001%
SUM	140,112	100%



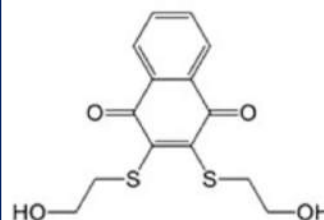
*Tested 71
Active 40*



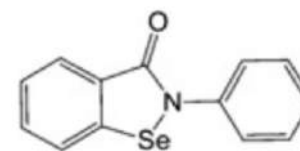
*Tested 81
Active 35*



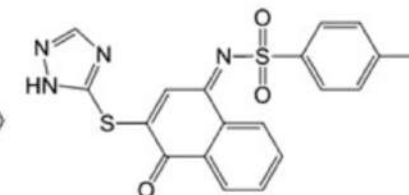
*Tested 85
Active 32*



*Tested 74
Active 34*



*Tested 70
Active 29*



*Tested 65
Active 31*

ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА: ПОДСТРУКТУРНЫЕ ФИЛЬТРЫ

WORST OFFENDERS

Pan-assay interference compounds (PAINS) fall into hundreds of chemical classes, but some groups occur much more frequently than others. Among the most insidious are the eight shown here (reactive portions shown in red and purple). These and related compounds should set off alarm bells if they show up as 'hits' in drug screens.

TOXOFLAVIN
Redox cyler: can produce hydrogen peroxide, which can activate or inactivate different proteins.

ISOTHIAZOLONES
Covalent modifier: reacts chemically with proteins in non-specific, non-drug-like ways.

CURCUMIN
Covalent modifier, membrane disruptor: muddles response of membrane receptors.

HYDROXYPHENYL HYDRAZONES
Covalent modifier, metal complexer: sequesters metal ions that inactivate proteins.

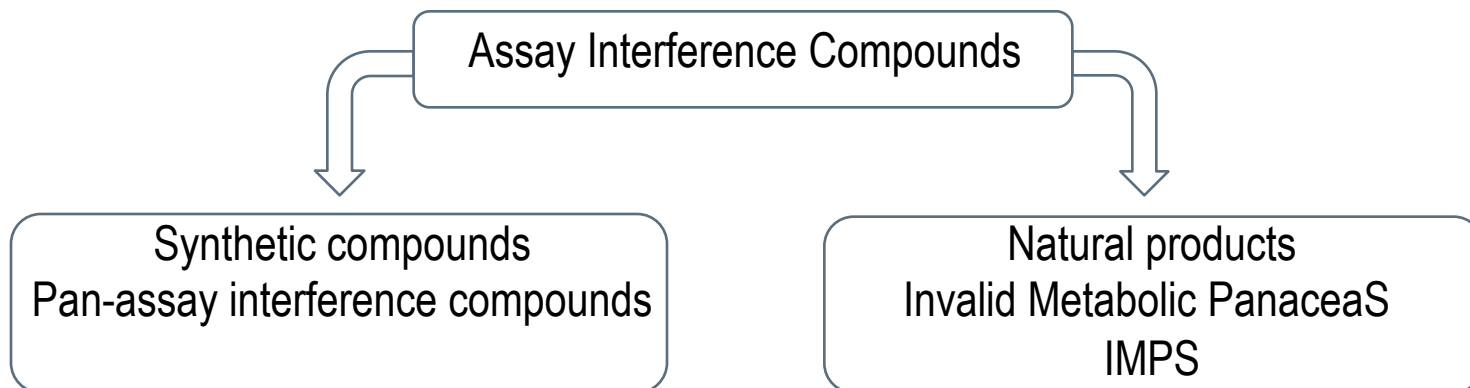
ENE-RHODANINE
Covalent modifier, metal complexer.

PHENOL-SULPHONAMIDES
Redox cyler, covalent modifier, unstable compound: breaks down into molecules that give false signals.

ENONES
Covalent modifier.

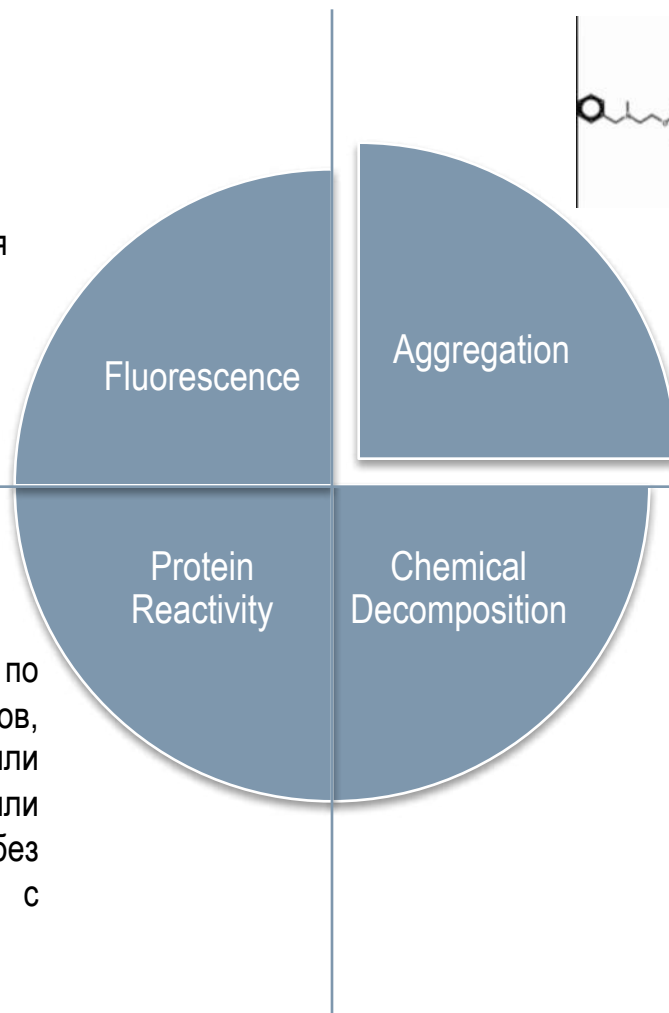
QUINONES AND CATECHOLS
Redox cyler, metal complexer, covalent modifier.

ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА: СОЕДИНЕНИЯ С НЕСПЕЦИФИЧЕСКИМ ВЗАИМОДЕЙСТВИЕМ

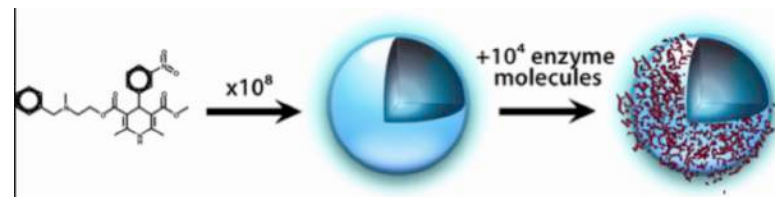


ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА: СОЕДИНЕНИЯ С НЕСПЕЦИФИЧЕСКИМ ВЗАИМОДЕЙСТВИЕМ

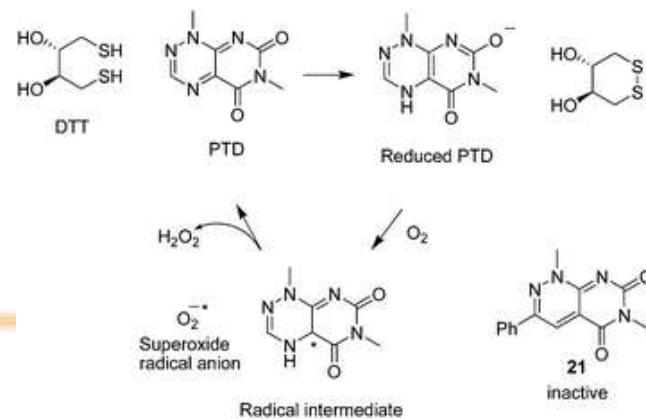
Выраженный сигнал определяется даже в отсутствие белка



Хелатные свойства соединений по отношению к ряду металлов, использованных при синтезе или использованных как реагенты или воздействие через поверхность без специфического взаимодействия с конкретным сайтом связывания



Образование пероксида водорода инактивирует мишень, что соответствует действию ингибитора, при этом сам лиганд может не связываться с мишенью



ПОДГОТОВКА БАЗ ДАННЫХ ДЛЯ ВИРТУАЛЬНОГО И ЭКСПЕРИМЕНТАЛЬНОГО СКРИНИНГА: ПОДСТРУКТУРНЫЕ ФИЛЬТРЫ ИДЕНТИФИКАЦИИ СОЕДИНЕНИЙ С НЕСПЕЦИФИЧЕСКИМ ВЗАИМОДЕЙСТВИЕМ

Общедоступные фильтры

<http://zinc15.docking.org/patterns/home>

<http://www.cbligand.org/PAINS>

<http://fafdrugs3.mti.univ-paris-diderot.fr/>

<http://fafdrugs3.mti.univ-paris-diderot.fr/>,
<http://advisor.docking.org>

АНАЛИЗ 29 БАЗ ДАННЫХ НА ПОДОБИЕ ЛЕКАРСТВУ (“DRUGLIKENESS”)

supplier	library size	3 of 4 Lipinski		4 of 4 Lipinski		Veber		absence of toxic/ reactive fragments		total filtered drug-like	
		No. of compounds passed	%	No. of compounds passed	%	No. of compounds passed	%	No. of compounds passed	%	No. of compounds passed	%
Albany Molecular Research	196 064	168 512	86%	104 334	53%	171 617	88%	186 901	95%	66 780	34%
ART-CHEM	110 873	101 959	92%	73 540	66%	94 653	85%	69 628	63%	35 040	32%
Asinex	457 842	443 550	97%	350 843	77%	399 447	87%	378 697	83%	212 472	46%
Asis Chem	32 749	31 834	97%	24 827	76%	30 118	92%	24 491	75%	14 031	43%
ChemBridge	741 176	728 794	98%	600 477	81%	678 668	92%	607 613	82%	386 566	52%
ChemDiv	785 740	747 052	95%	540 565	69%	720 162	92%	650 781	83%	330 144	42%
ChemStar	28 946	26 030	90%	16 336	56%	24 833	86%	13 288	46%	4977	17%
Enamine	1221 957	1160 859	95%	1 002 004	82%	1 063 102	87%	1 026 443	84%	672 076	55%
FluoroChem	23 498	23 040	98%	14 150	60%	22 027	94%	15 744	67%	7435	32%
InterBioScreen	466 671	444 365	95%	317 429	68%	405 300	87%	324 547	70%	144 850	31%
IVK Laboratories	46 515	45 144	97%	34 059	73%	41 926	90%	28 966	62%	13 688	29%
Key Organics	47 656	46 301	97%	33 383	70%	44 393	93%	33 065	69%	17 984	38%
Life Chemicals	426 135	401 386	94%	269 897	63%	372 876	88%	350 943	82%	153 850	36%
Maybridge	69 862	68 088	97%	48 988	70%	63 683	91%	46 238	66%	25 225	36%
Nanosyn	62 597	59 158	95%	42 241	67%	54 502	87%	37 063	59%	17 924	29%
Oakwood Chemicals	12 621	12 308	98%	6715	53%	12 046	95%	8500	67%	3796	30%
Otava Chemicals	173 941	159 692	92%	103 736	60%	150 082	86%	122 534	70%	49 281	28%
Peakdale	14 576	14 072	97%	10 246	70%	14 009	96%	13 903	95%	7209	49%
Pharmeks	155 800	138 626	89%	94 023	60%	127 982	82%	95 591	61%	34 811	22%
Princeton Biomolecular Research	380 424	352 728	93%	240 877	63%	335 530	88%	246 944	65%	107 968	28%
SALOR	48 693	44 053	90%	24 386	50%	40 715	84%	21 846	45%	8300	17%
Specs	223 630	207 769	93%	140 776	63%	201 236	90%	159 595	71%	71 882	32%
Spectrum	8497	7997	94%	5607	66%	7769	91%	7083	83%	3590	42%
TimTec	674 773	609 034	90%	431 965	64%	579 690	86%	390 235	58%	151 164	22%
TOSLab	26 713	22 748	85%	13 505	51%	23 045	86%	17 236	65%	5212	20%
Tripos	154 604	150 949	98%	115 517	75%	141 993	92%	129 010	83%	72 350	47%
Ufark	28 881	26 975	93%	14 984	52%	25 636	89%	15 813	55%	7650	26%
UORSY	794 997	683 318	86%	596 784	75%	591 882	74%	532 054	67%	277 088	35%
Vitas-M Lab	442 971	419 064	95%	302 404	68%	382 768	86%	327 505	74%	143 146	32%
unique	5 183 506	4 909 717	95%	3 716 537	72%	4 545 593	88%	4 010 470	77%	2 020 027	39%
open NCI database compounds	231 458	210 856	91%	161 543	70%	182 841	79%	140 826	61%	95 588	41%

LIGAND EFFICIENCY METRICS

Ligand efficiency – merit of compound's propensity to use its structural features for binding to target

Ligand Efficiency:

$$LE = (-2.303(RT/HA)) \times \log K_d \quad \text{or} \quad LE = (1.37/HA) \times pIC_{50} \quad \text{or} \quad LEI = pIC_{50}/HA$$

Free energy of binding:

$$\Delta G^0 = -RT \times \ln(K_d/C^0) = -2.303RT \times \log(K_d/C^0)$$

$$LE = \Delta G^0/HA = -(2.303RT/HA) \times \log(K_d/C^0) \approx -(1.37/HA) \times \log(K_d) = (1.37/HA) \times pK_d$$

pIC50 – half-maximal inhibitory concentration, p*K_i* – inhibition constant, p*K_d* – dissociation constant

Lipophilic Ligand Efficiency:

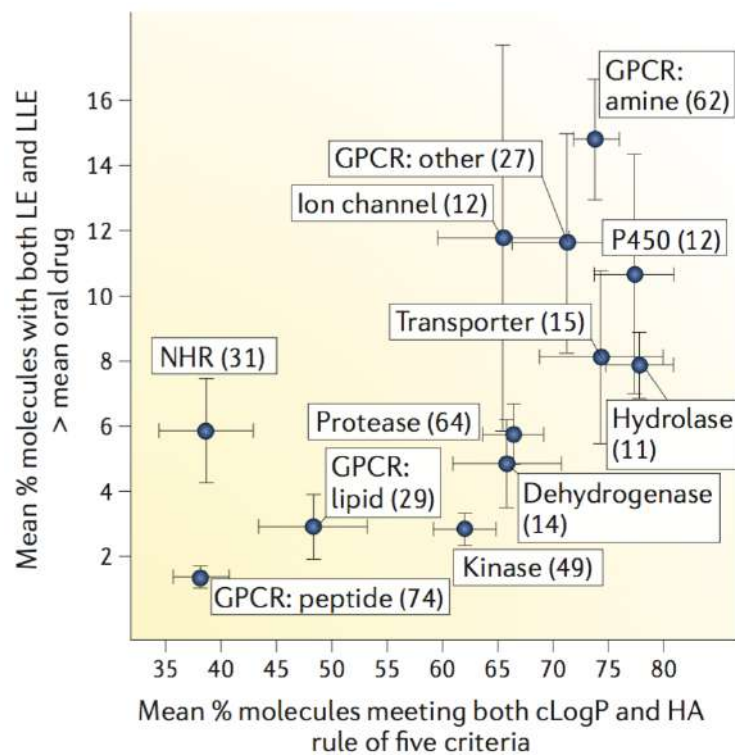
$$LLE = pIC_{50} - c\text{LogP} \quad \text{or} \quad LLE_{AT} = 0.111 + 1.37(LLE/HA) \quad \text{or} \quad LLEP = c\text{LogP}/LE$$

Size Independent Ligand Efficiency:

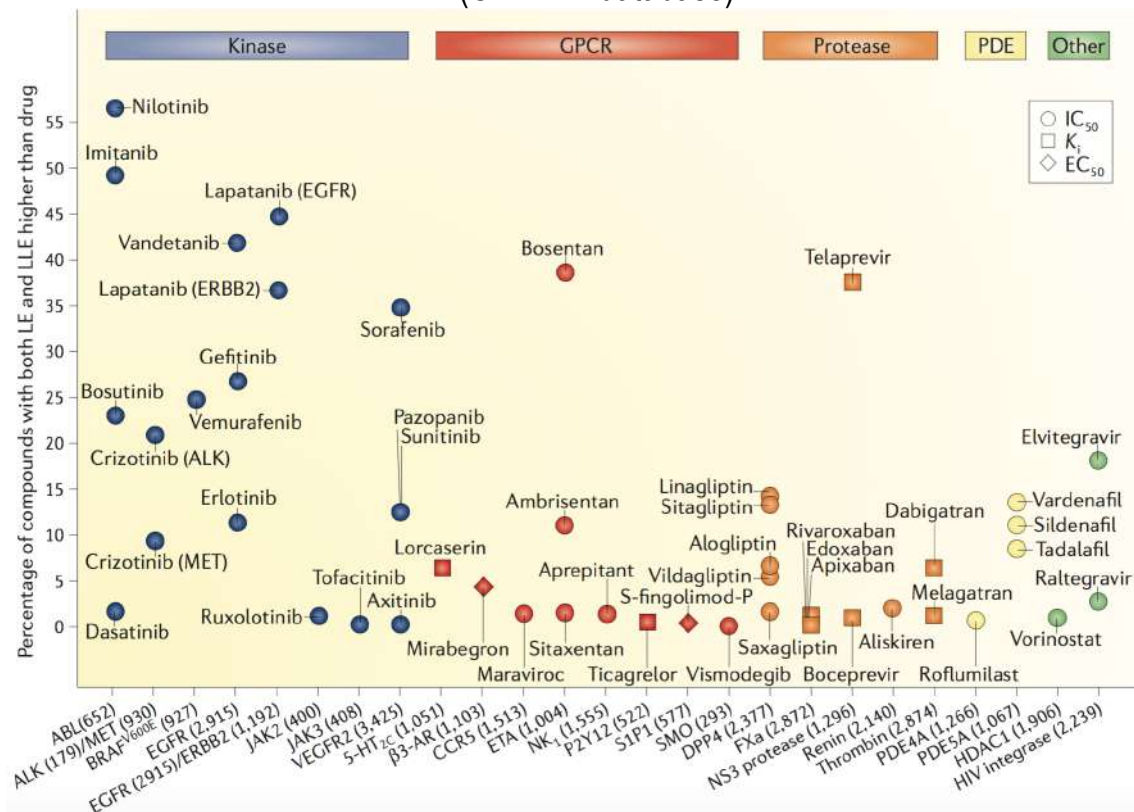
$$SILE = pIC \text{ or } pK \div HA^{0.3}$$

LIGAND EFFICIENCY METRICS: DRUGGABILITY OVER TARGET FAMILIES

480 target-assay pairs (GVK BIO Database)



Relative ligand efficiencies of 46 oral drugs acting at 25 targets (ChEMBL database)



ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Internal Structure and Preferential Protein Binding of Colloidal Aggregates

D. Duan et al *ACS Chemical Biology* **2017** 12 (1), 282-290 DOI: 10.1021/acscchembio.6b00791

The Ecstasy and Agony of Assay Interference Compounds

C. Aldrich et al *Journal of Medicinal Chemistry* **2017** 60 (6), 2165-2168 DOI: 10.1021/acs.jmedchem.7b00229

Highly Promiscuous Small Molecules from Biological Screening Assays Include Many Pan-Assay Interference Compounds but Also Candidates for Polypharmacology

E. Gilberg et al *Journal of Medicinal Chemistry* **2016** 59 (22), 10285-10290 DOI: 10.1021/acs.jmedchem.6b01314

Dependence of Molecular Properties on Proteomic Family for Marketed Oral Drugs

M. Vieth et al *Journal of Medicinal Chemistry* **2006** 49 (12), 3451-3453 DOI: 10.1021/jm0603825

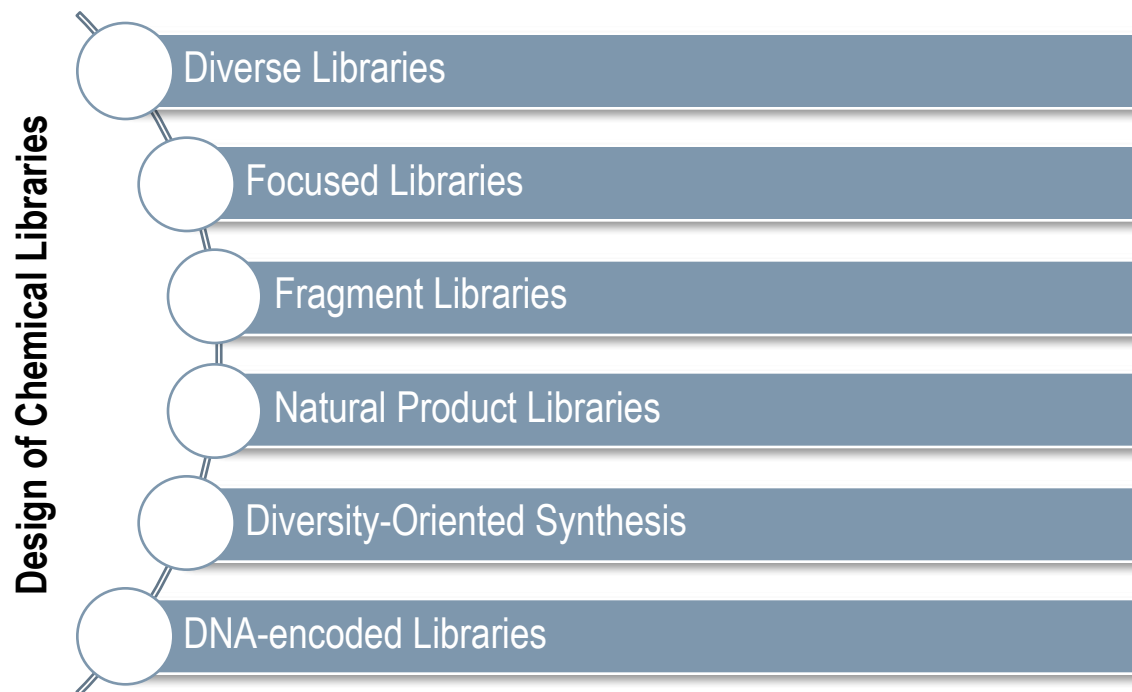
Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay Interference Compounds

Capuzzi et al *Journal of Chemical Information and Modeling* **2017** 57 (3), 417-427

The role of ligand efficiency metrics in drug discovery

A. L. Hopkins et al *Nature Reviews Drug Discovery* 13, 105–121 (2014)

ДИЗАЙН БИБЛИОТЕК СОЕДИНЕНИЙ



Дизайн диверсифицированных и сфокусированных библиотек соединений

Диверсифицированные библиотеки (Diverse Libraries)

Библиотеки соединений, содержащие максимально разнородные лиганды, взаимодействующие с широким диапазоном мишеней

- ❖ Поиск соединений при ограниченной информации о мишени
- ❖ Отбор соединений для взаимодействия с несколькими мишенями
- ❖ Пополнение базы данных
- ❖ Предварительный этап для высокопроизводительного скрининга

Сфокусированные библиотеки (Focused Libraries)

Библиотеки соединений, созданные на основе отобранных молекулярных скелетов, ориентированных на взаимодействие со специфической биологической мишенью

- ❖ Поиск соединений, сходных с соединением-лидером или отобранных согласно QSAR модели
- ❖ Отбор соединений, взаимодействующих с одной конкретной мишенью или семейством взаимосвязанных

МЕТОДЫ ДИЗАЙНА ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК, ОСНОВАННЫЕ НА ДЕСКРИПТОРНОМ ОПИСАНИИ ДАННЫХ

Пошаговый отбор
соединений на основе
разнородности

Кластеризация

Методы с заданной
топологией

Использование
оптимизационных
алгоритмов

МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ: ОТБОР СОЕДИНЕНИЙ НА ОСНОВЕ РАЗНОРОДНОСТИ

Общая последовательность:

- ❖ Случайный выбор
- ❖ Наиболее репрезентативное
- ❖ Наиболее несходное

Выбор первого соединения

Расчет разнородности между всеми соединениями в базе данных и уже отобранными

$$(Dissimilarity)_{ij} = 1 - (Similarity)_{ij}$$

Метод исключения сферы

Выбор следующего соединения, наиболее отличающегося от предыдущих

Наибольшая разнородность с уже отобранными

$$score_i = \sum_{j=1}^m D_{i,j}$$

Наибольшее расстояние до ближайших соседей среди уже отобранных:

$$score_i = \min(D_{i,j}; j=1, m)$$

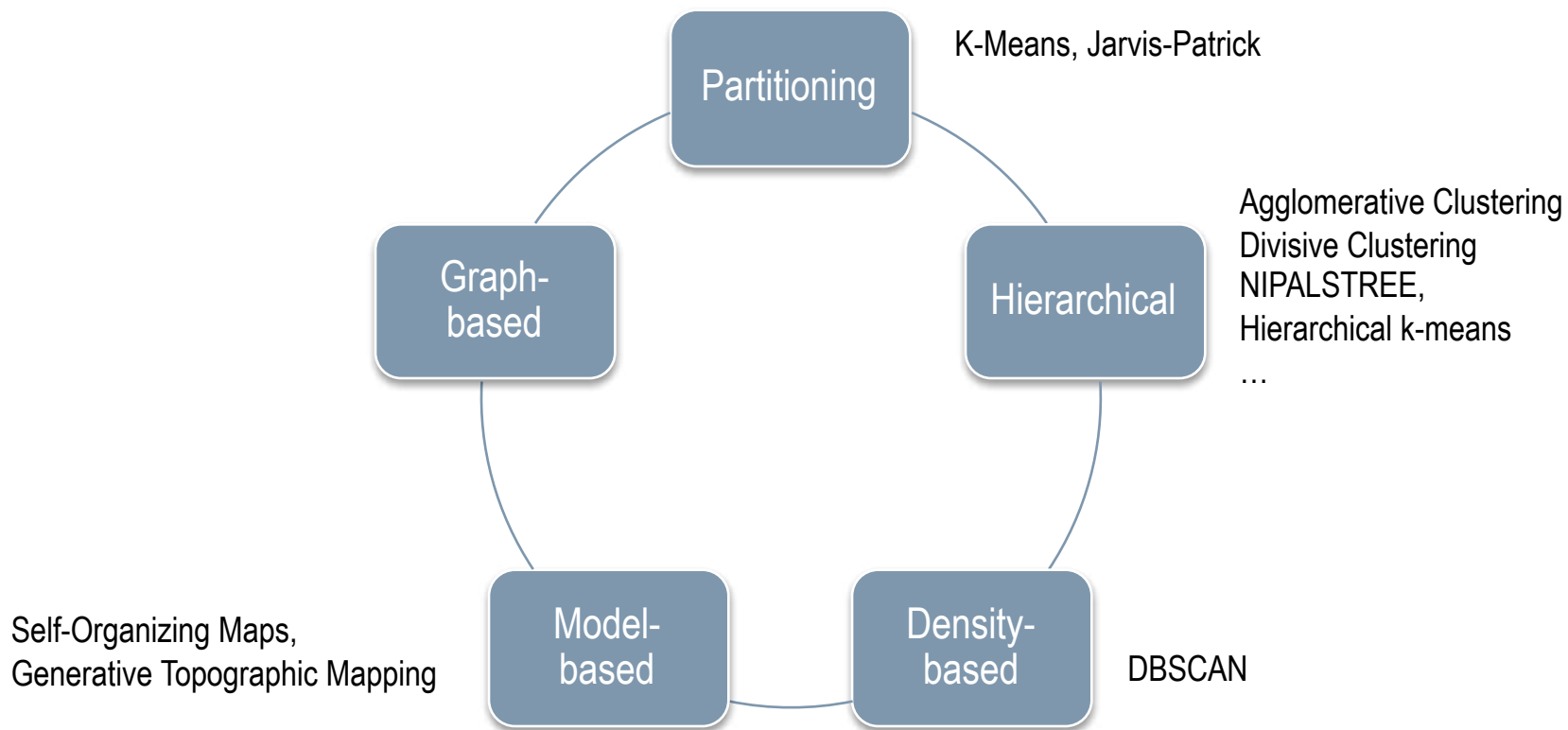
Если заданное количество соединений не отобрано повторно двух предыдущих шагов



ACS Chem. Biol. 2011, 6, 208–217

1. Задать пороговое значение отличия t
2. Выбрать соединение
3. Исключить все соединения с отличием $< t$
4. Если соединения в библиотеке после предыдущего шага остались, вернуться к шагу 2

МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ



Общий принцип:

- Разбить соединения на категории (группы, кластеры): по принципу сходства молекул, относящихся к одному кластеру
- Выбрать одного или нескольких представителей каждой группы

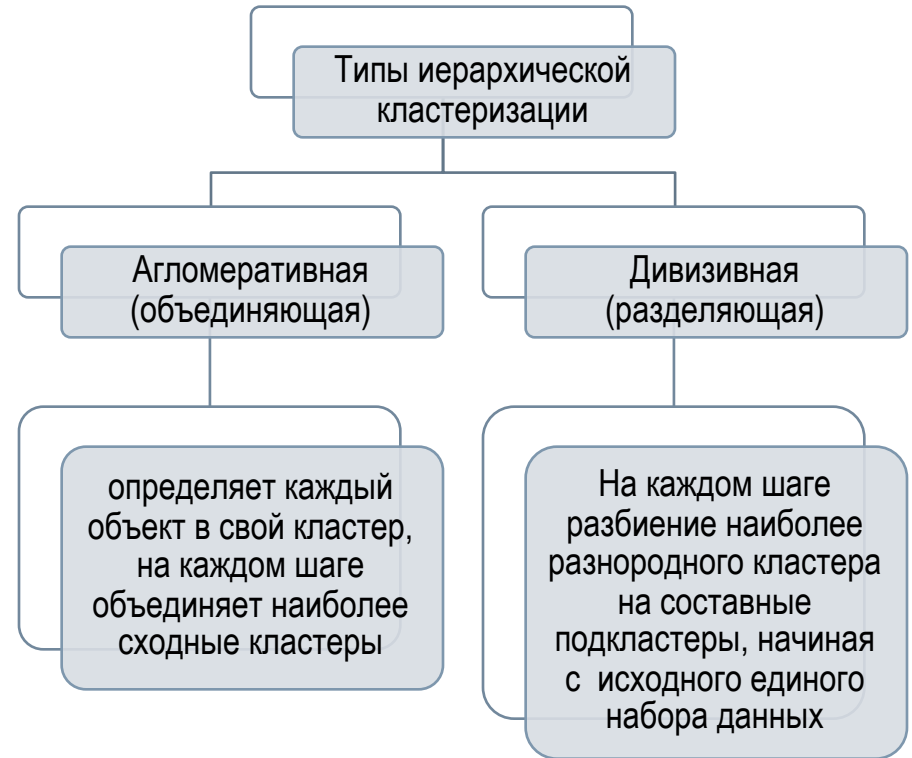
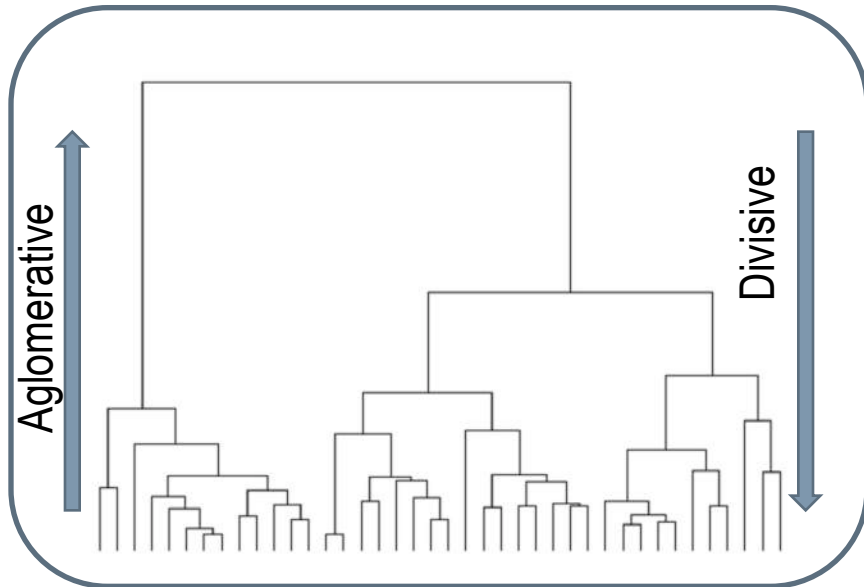
КЛАСТЕРИЗАЦИЯ: МЕТОД ДЖАРВИСА-ПАТРИКА

Расчет попарных расстояний, определение ближайших соседей для каждого соединения



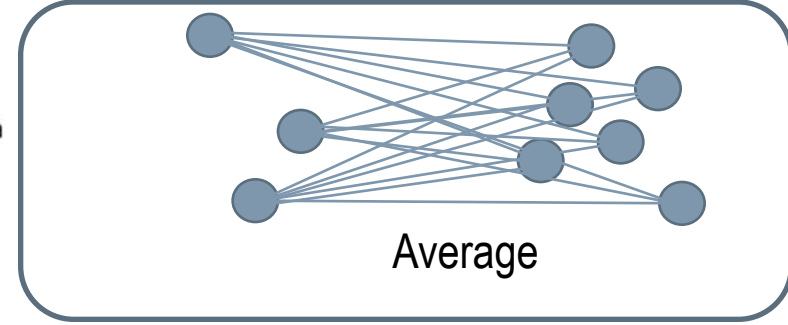
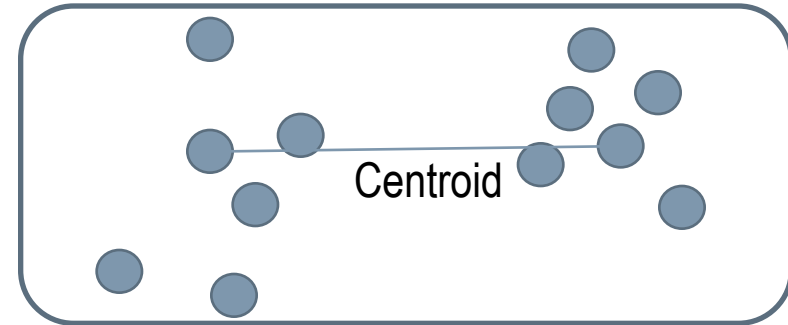
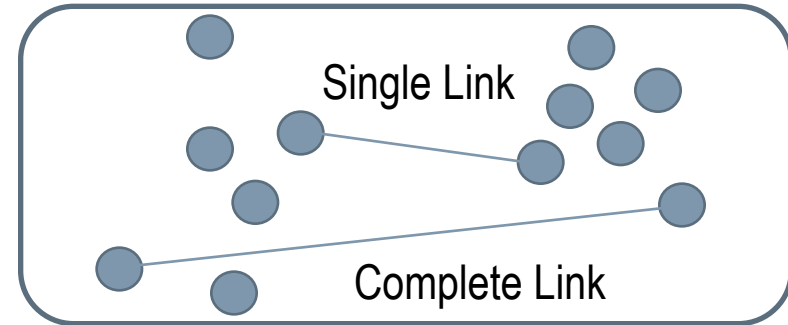
Два соединения попадают в один кластер, если они сами и соединения из их списка находятся в списке ближайших соседей друг друга

КЛАСТЕРИЗАЦИЯ: ИЕРАРХИЧЕСКАЯ

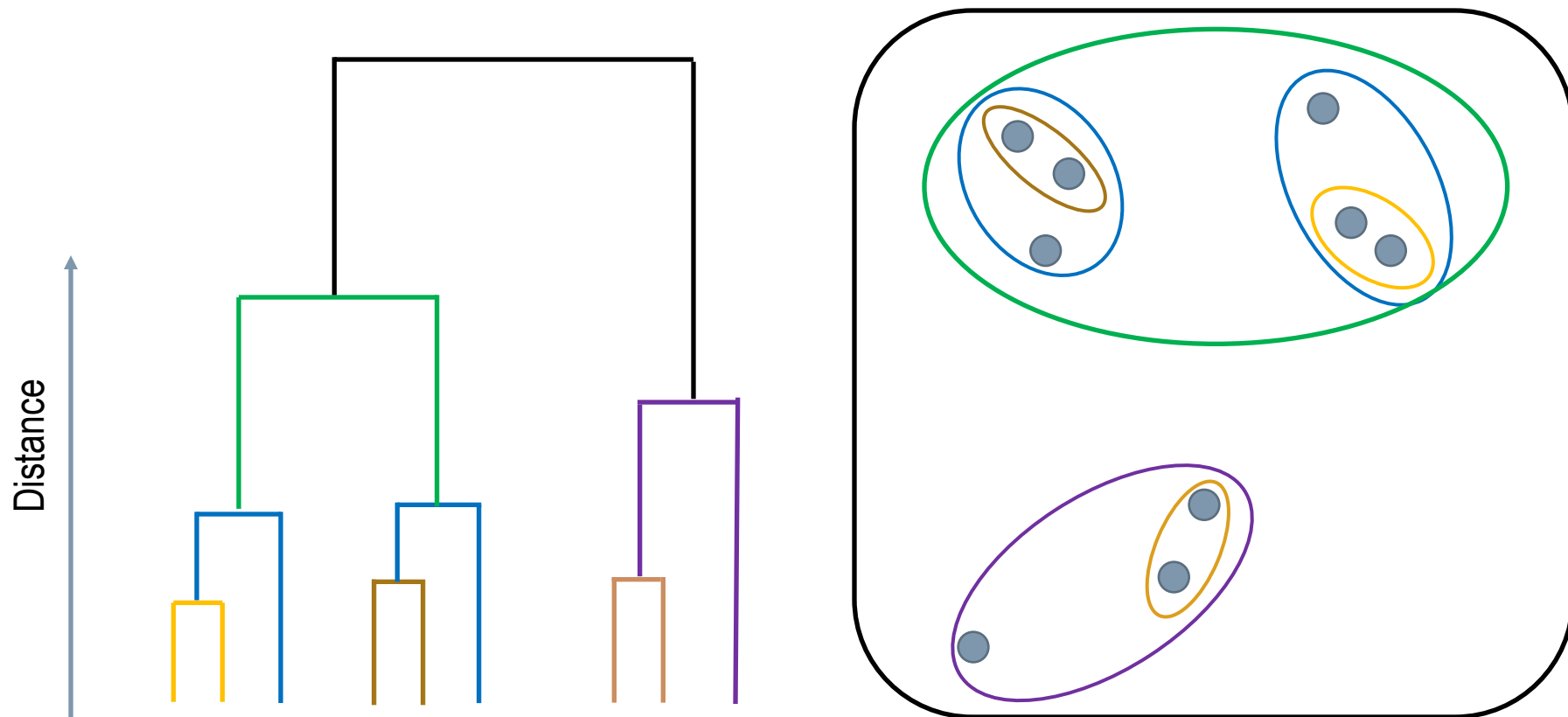


ИЕРАРХИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ: СПОСОБЫ ОПРЕДЕЛЕНИЯ РАССТОЯНИЙ

- **Single link:** smallest distance between an element in one cluster and an element in the other, i.e., $\text{dist}(K_i, K_j) = \min(t_{ip}, t_{jq})$
- **Complete link:** largest distance between an element in one cluster and an element in the other, i.e., $\text{dist}(K_i, K_j) = \max(t_{ip}, t_{jq})$
- **Centroid:** distance between the centroids of two clusters, i.e., $\text{dist}(K_i, K_j) = \text{dist}(C_i, C_j)$
- **Average:** avg distance between an element in one cluster and an element in the other, i.e., $\text{dist}(K_i, K_j) = \text{avg}(t_{ip}, t_{jq})$



КЛАСТЕРИЗАЦИЯ: ИЕРАРХИЧЕСКАЯ АГЛОМЕРАТИВНАЯ

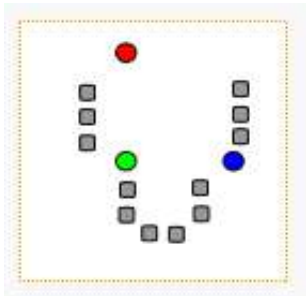


КЛАСТЕРИЗАЦИЯ: МЕТОД К- СРЕДНИХ (K-MEANS)

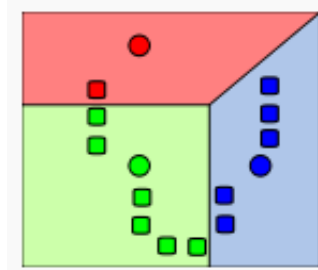
Разделение набора данных на k поднаборов проводится так, чтобы минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров

$$\min \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

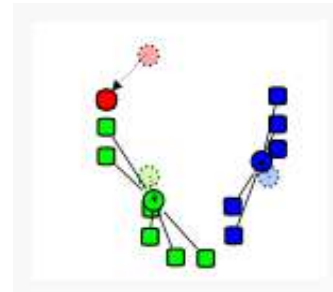
где $\boldsymbol{\mu}_i$ центр масс кластера S_i .



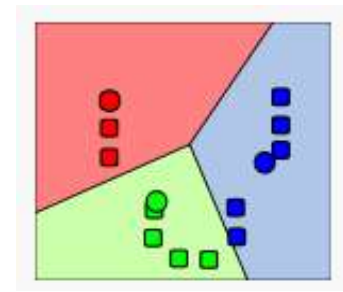
1



2



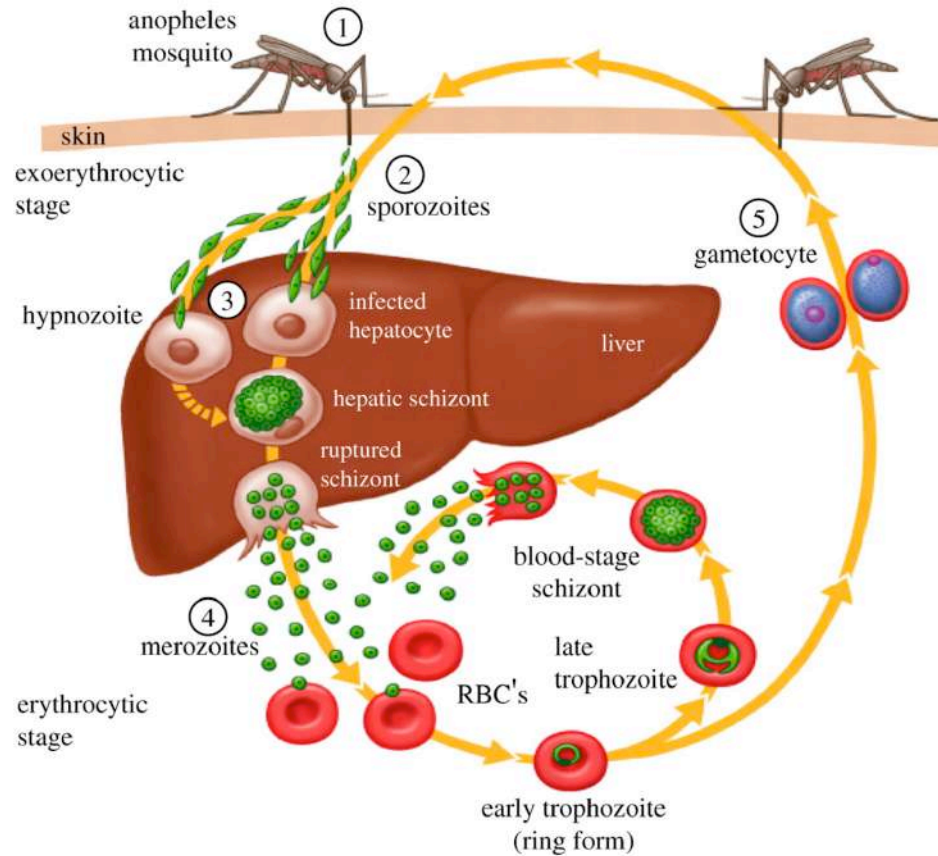
3



4

1. Случайный выбор c начальных точек (соединений).
2. Другие соединения относятся к ближайшим, из определенных на предыдущем шаге, образуя c исходных кластеров
3. Определяются центроиды кластеров.
4. Центры кластеров смещаются в их центроиды.
5. Повторение шагов 2-4 до схождения

МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ: КЛАСТЕРИЗАЦИЯ (CASE STUDY OF APPROACH APPLICATION)



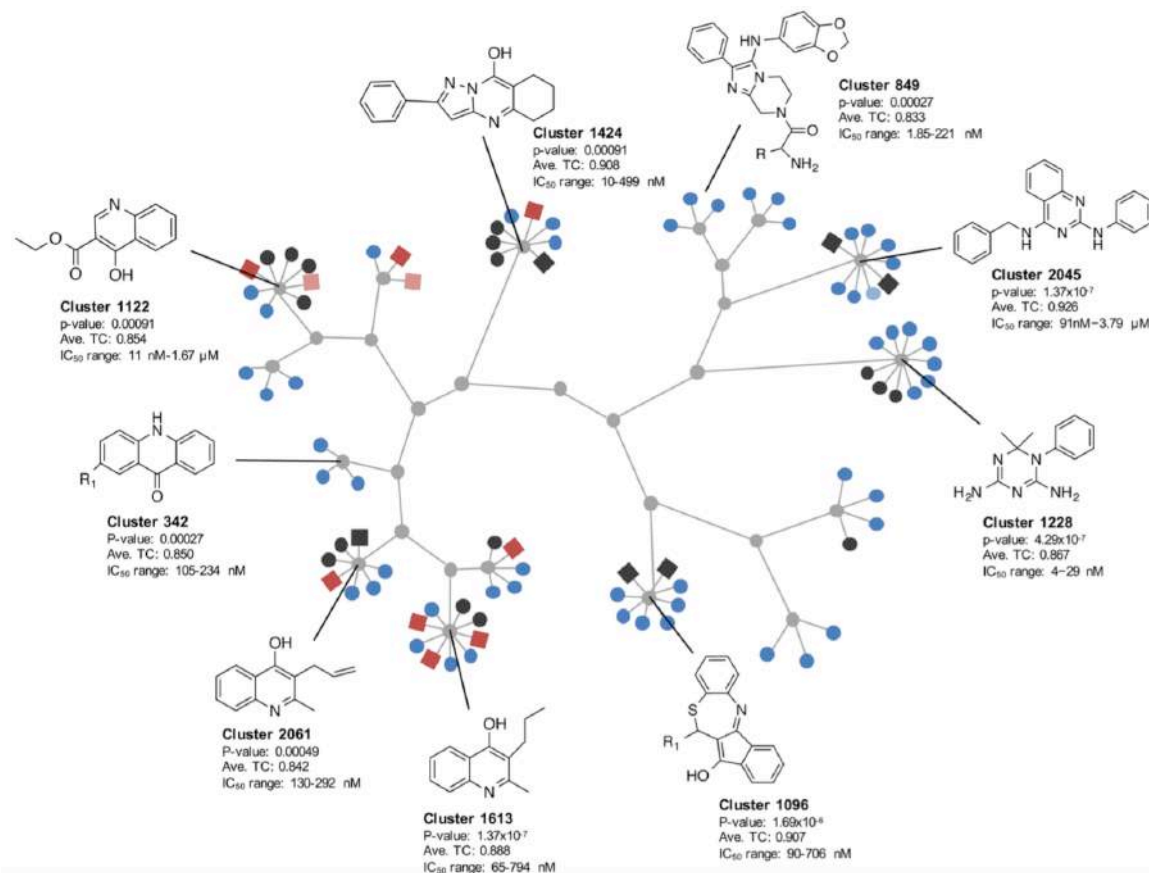
Phil. Trans. R. Soc. B 2011 366 2806-2814; DOI: 10.1098/rstb.2011.0091

High-Throughput Luciferase-Based Assay for the Discovery of Therapeutics That Prevent Malaria
J Swann *ACS Infectious Diseases* 2016 2 (4), 281-293 DOI: 10.1021/acsinfecdis.5b00143

МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ: КЛАСТЕРИЗАЦИЯ

- 4822 compounds were clustered using a hierarchical clustering method based on substructure similarity (Tanimoto coefficient 0.65 => 2335 total clusters from 1 to 45 compounds

- 15 scaffold series showed specific enrichment in exoerythrocytic-stage activity



МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ: МЕТОДЫ С ЗАДАННОЙ ТОПОЛОГИЕЙ

Определить
пространство малой
размерности

Определить
принадлежность
каждого из соединений к
одной из ячеек карты
(многообразия)

Выбрать одного или
нескольких
представителей от
каждой ячейки или
вершины

МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ: ИСПОЛЬЗОВАНИЕ ОПТИМИЗАЦИОННЫХ АЛГОРИТМОВ

Отбор соединений в поднабор случайным образом, оценка разнородности данных
(расчет значения соответствующего параметра)

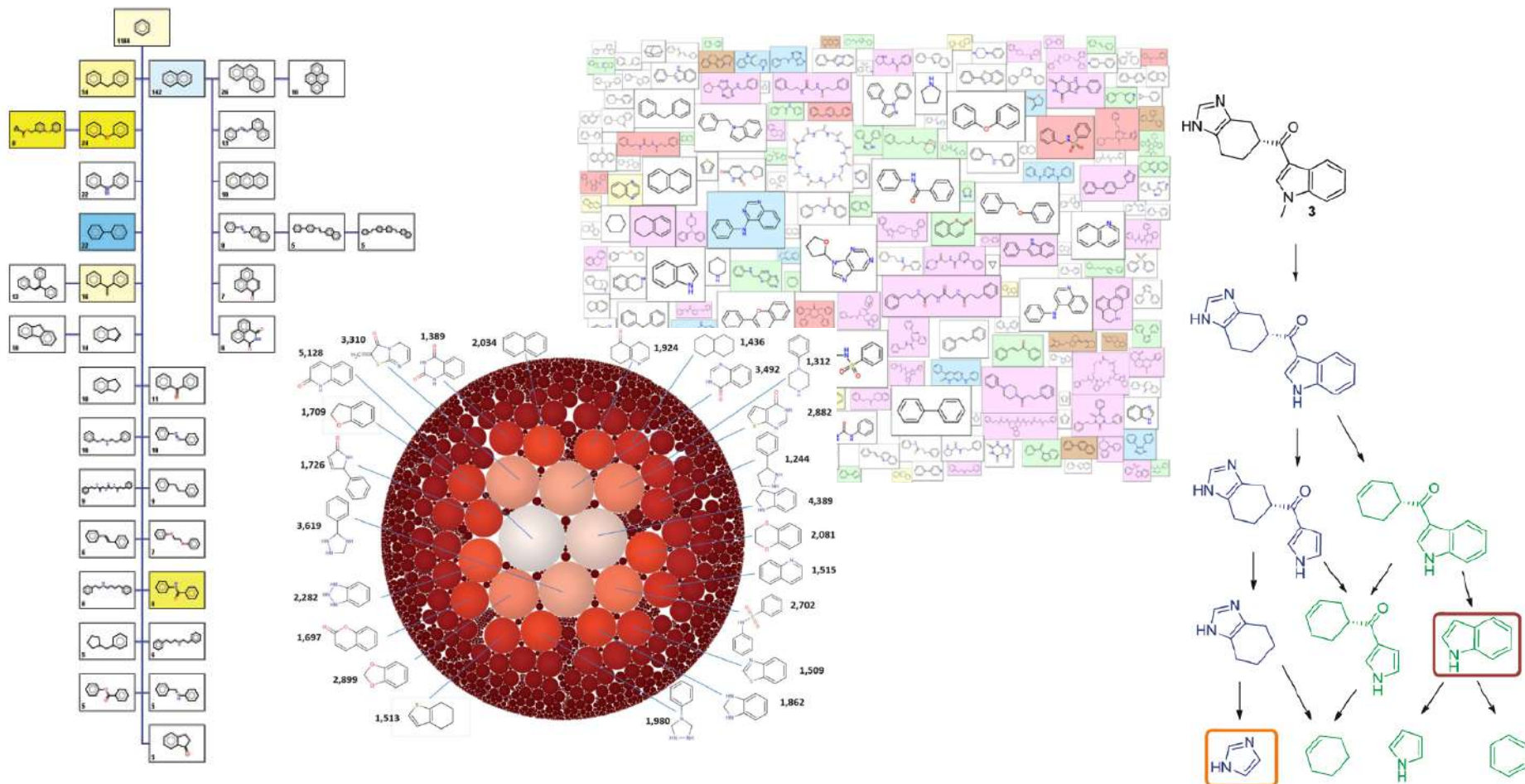


Новый набор данных получается из предыдущего случайной заменой отдельных соединений



Оценивается разнородность нового набора данных, если значение превосходит предыдущее, то полученный вариант принимается, если нет, то проводится дополнительная оценка

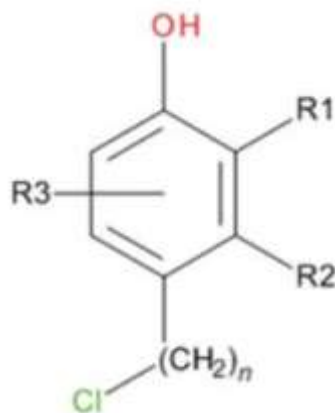
МЕТОДЫ СОЗДАНИЯ ДИВЕРСИФИЦИРОВАННЫХ БИБЛИОТЕК СОЕДИНЕНИЙ, ОСНОВАННЫЕ НА ОПИСАНИИ ДАННЫХ ПРИ ПОМОЩИ МОЛЕКУЛЯРНЫХ ГРАФОВ



См. материалы предыдущего семестра

МЕТОДЫ ДИЗАЙНА СФОКУСИРОВАННЫХ БИБЛИОТЕК СТРУКТУРЫ МАРКУША (ИЛИ GENERIC STRUCTURES)

- ❖ Структура Маркуша – способ отображения серии соединений с помощью общего для них ядра и одного или нескольких вариантов заместителей, объединяемых в группу альтернативных структур.
- ❖ Одна структура Маркуша может соответствовать нескольким миллионам конкретных химических соединений, относящихся к различным классам
- ❖ Используется как в патентах, так и при описании комбинаторных библиотек



Substituent variation

R1 = methyl or ethyl

Homology variation

R2 = alkyl

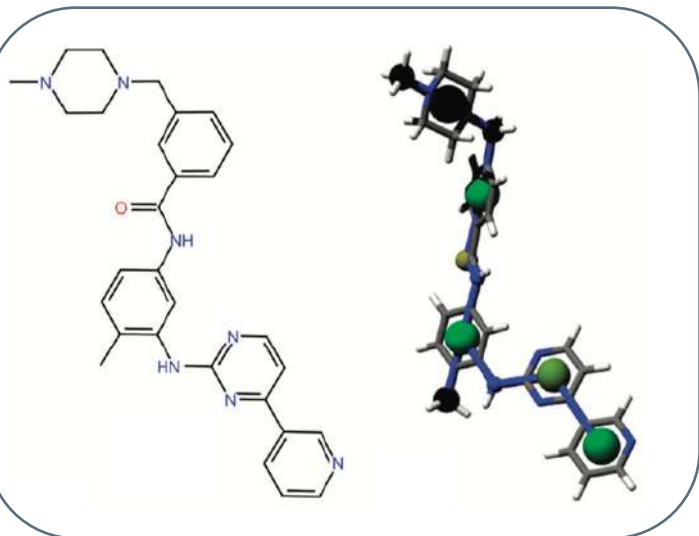
Position variation

R3 = amino

Frequency variation

$n = 1-3$

LOFT: SIMILARITY-DRIVEN MULTIOBJECTIVE FOCUSED LIBRARY DESIGN

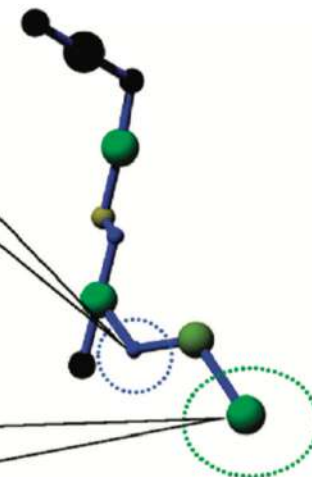
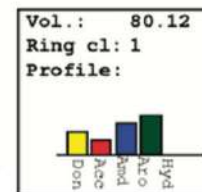
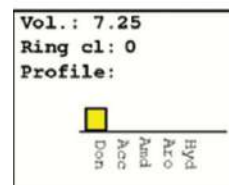


number of links
molecular weight
number of non-hydrogen atoms
number of non-hydrogen bonds
smallest set of smallest rings
number of ring systems
number of H-bond acceptors
number of H-bond donors
number of rotatable bonds
maximal path of contiguous rotatable bonds¹
topological polar surface area*⁴²
number of EZ stereo centers
number of RS stereo centers
calculated logP value^{65,66}
molar refractivity
polar surface area
link type

inclusion SMARTS

exclusion SMARTS

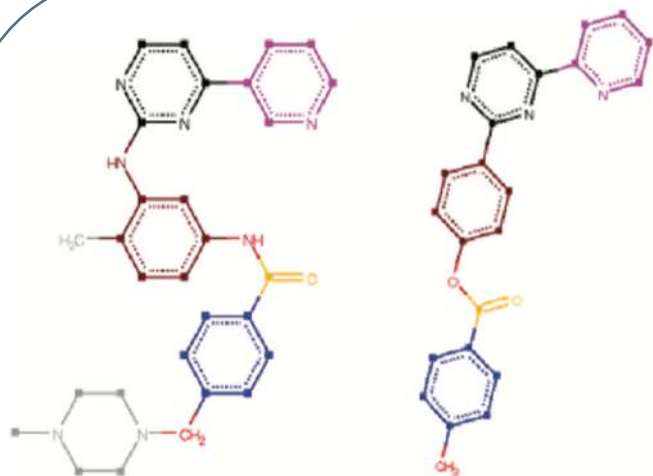
user defined properties* (strictly additive)



LoFT: Similarity-Driven Multiobjective Focused Library Design

J. R. Fischer et al *Journal of Chemical Information and Modeling* 2010 50 (1), 1-21 DOI: 10.1021/ci900287p

LOFT: SIMILARITY-DRIVEN MULTIOBJECTIVE FOCUSED LIBRARY DESIGN



Imatinib **4**

11P-230 **5**

Overall similarity: 0.9099

Matched fragments:

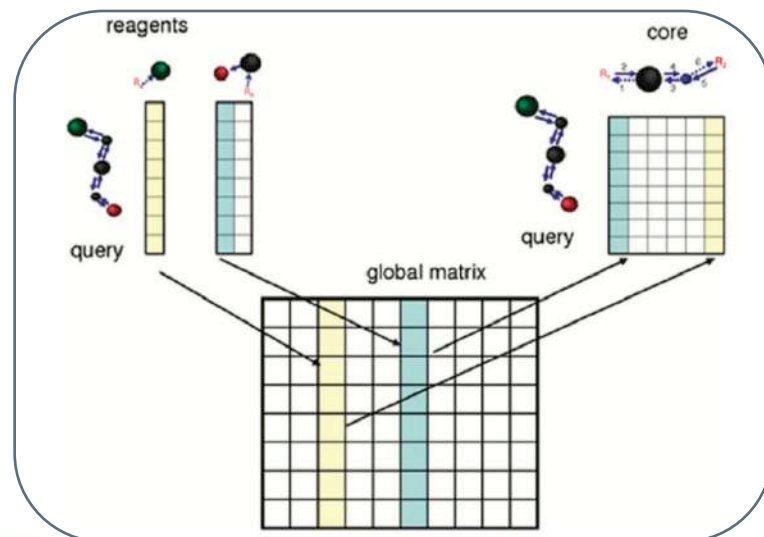
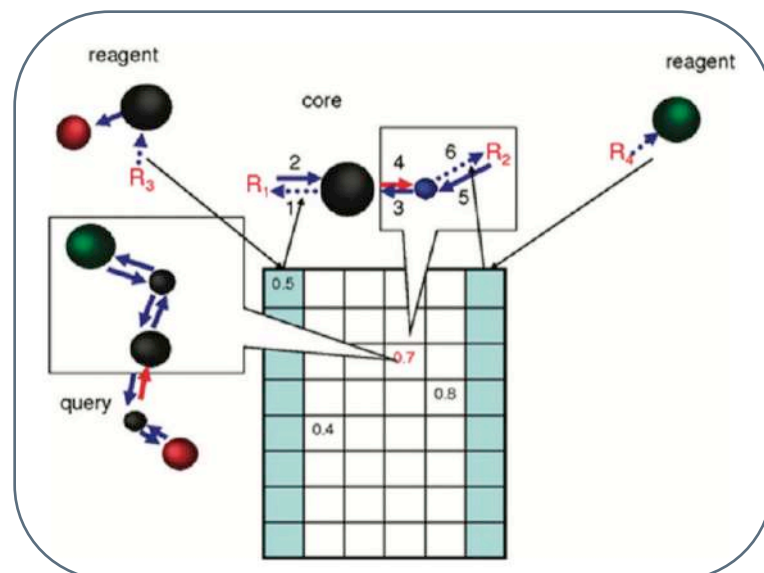
C
c1ccccc1
O=C
c1ncccc1
c1ncccn1
c1ccccc1,N
N
 Not matched

Matched fragments:

C
c1ccccc1
O=C
c1ncccc1
c1ncccn1
c1ccccc1
O
 Not matched

Local similarity:

0.9554
1.0000
0.9982
1.0000
0.9981
0.8921
0.2854
 Not matched



LOFT: Similarity-Driven Multiobjective Focused Library Design

J. R. Fischer et al *Journal of Chemical Information and Modeling* 2010 50 (1), 1-21 DOI: 10.1021/ci900287p

LOFT: SIMILARITY-DRIVEN MULTIOBJECTIVE FOCUSED LIBRARY DESIGN

$$qsim(p) = \max(\text{sim}(i,p) | i \in q)$$

Maximum similarity of product p to any query molecule i from query set q .

$$qsim(p) = \frac{1}{|q|} \sum_{i=1}^{|q|} \text{sim}(i,p)$$

Sum of the similarity values of product p compared to all queries q divided by the number of queries.

$$asim(p) = 1 - \max(\text{sim}(i,p) | i \in a)$$

Maximum dissimilarity (1-similarity) of product p to any antiquery molecule i from the antiquery set a .

$$asim = \frac{1}{|a|} \sum_{i=1}^{|a|} (1 - \text{sim}_i)$$

Sum of the dissimilarity values of product p compared to all antiqueries a divided by the number of antiqueries.

$$pscore(p) = \sum_{i=1}^n w_i s_i$$

Weighted sum over all property scores to score a product p where n is the number of properties, w_i the weight and s_i the score of a single property.

$$P(r) = \{\text{products } p | p \text{ contains } r\}$$

Set of all products which contain reagent r .

$$rscore(r) = \frac{1}{|P(r)|} \sum_{p \in P} pscore(p)$$

The score of a reagent r is the sum of the product scores divided by the number of products. Only products which contain r are taken into account.

$$R(l) = \{\text{reagents } r | l \text{ contains } r\}$$

All reagents r which are part of sublibrary l .

$$libscore(l) = \frac{1}{|R(l)|} \sum_{r \in R} rscore(r)$$

Arithmetic mean score over the reagent scores of sublibrary l .

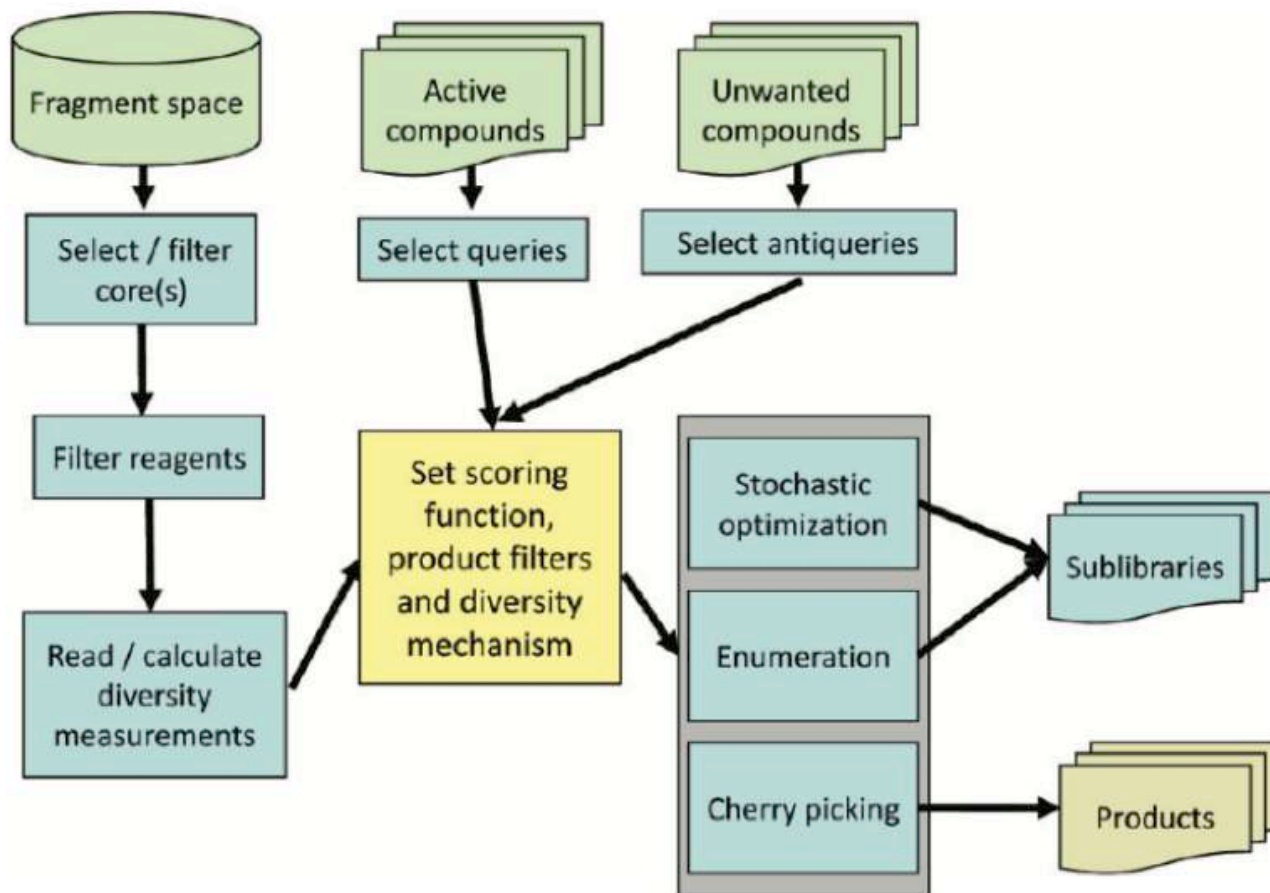
$$libscore(l) = \left(\prod_{r \in R} rscore(r) \right)^{1/|R|}$$

Geometric mean score over the reagent scores of sublibrary l .

$$libscore(l) = \sqrt{\frac{1}{|R(l)|} \sum_{r \in R} rscore(r)^2}$$

Quadratic mean score over the reagent scores of sublibrary l .

LOFT: SIMILARITY-DRIVEN MULTIOBJECTIVE FOCUSED LIBRARY DESIGN



LoFT: Similarity-Driven Multiobjective Focused Library Design

J. R. Fischer et al *Journal of Chemical Information and Modeling* 2010 50 (1), 1-21 DOI: 10.1021/ci900287p

FRAGMENT LIBRARIES: FRAGMENTS' CONVENTIONAL CHARACTERISTICS

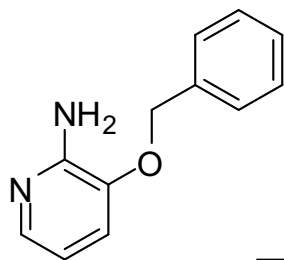
Molecular Weight M_r ~300 Da

H-bond donors (HBD) <3

H-bond acceptors (HBA) <3

Clog P <3 (a measure of lipophilicity of a compound)

Polar Surface Area (PSA) <60 (a measure of permeability through the cell membrane)



Fragment

$IC_{50} = 1.3\text{mM}$

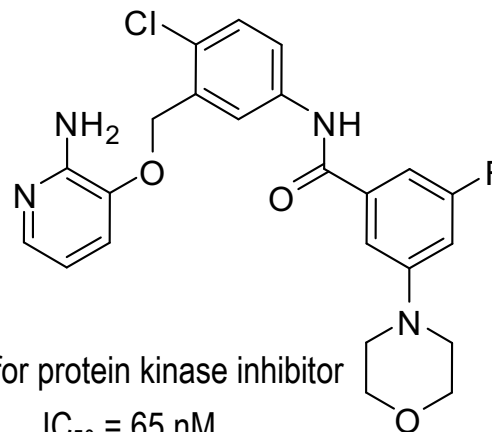
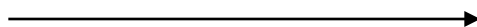
$M_r = 200$

HBD= 2

HBA=3

Clog P=1.92

PSA=48.14



Lead for protein kinase inhibitor

$IC_{50} = 65\text{ nM}$

$M_r = 456$

FRAGMENT LIBRARIES: ADDITIONAL REQUIREMENTS

Sample relevant chemical space by including pharmacophores that can be responsible for fragment binding

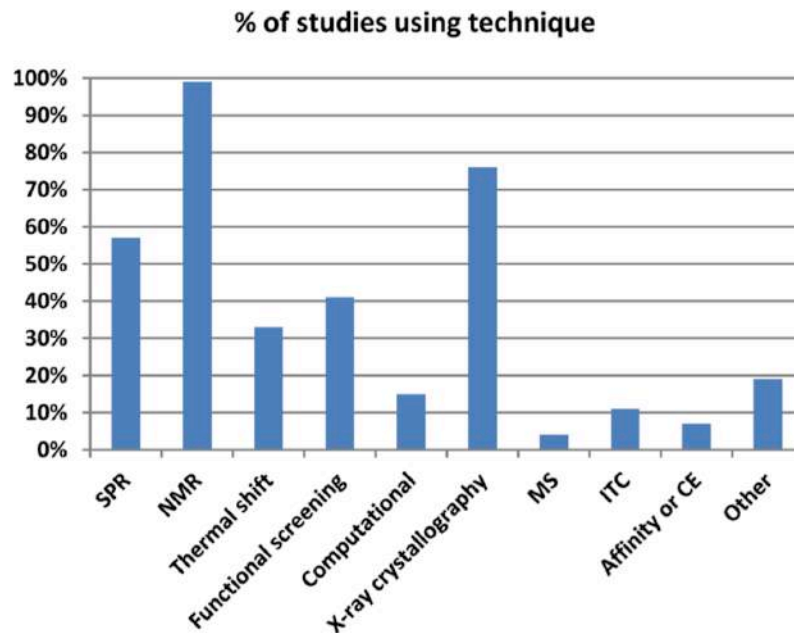
Contain an appropriate size distribution and a balance of differently shaped fragments of appropriate complexity

Contain a diversity of synthetically accessible growth vectors so that fragment hits can be effectively optimized into lead compounds

Avoid groups known to be associated with high reactivity, aggregation in solution, or persistent false positive data

FRAGMENT LIBRARIES

Fragment screening technologies used in published FBDD studies

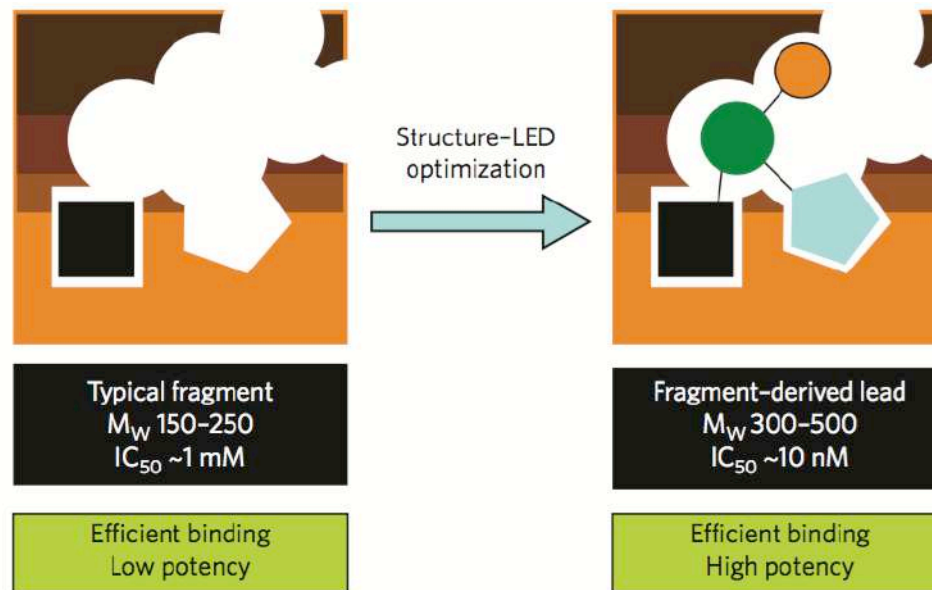
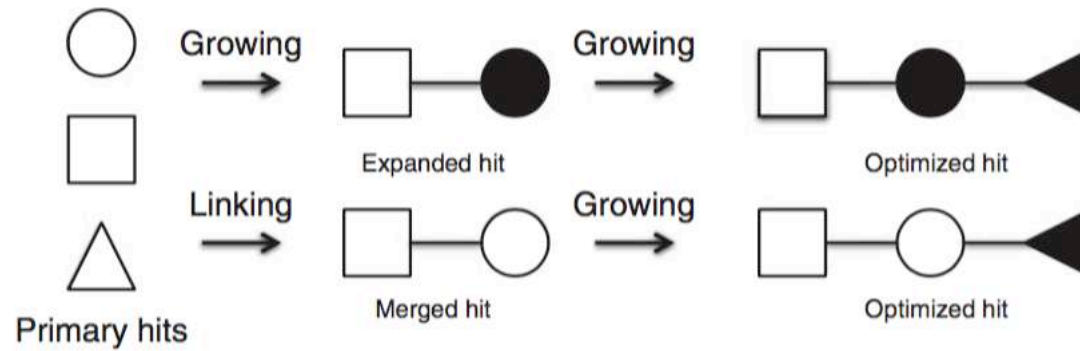


method	sensitivity limit	specificity assessment	throughput	structural information	propensity for false positive/false negative
biochemical	high μM	+	high	none	high FP/FN
ligand-NMR	low mM	-	medium	some	medium FP
protein-NMR	low mM	+	low	high	low FP/FN
SPR	high μM	+	medium	none	medium FP
thermal shift	high μM , low mM	-	high	none	high FP/FN
X-ray	mid mM	+	low	high	low FP/high FN

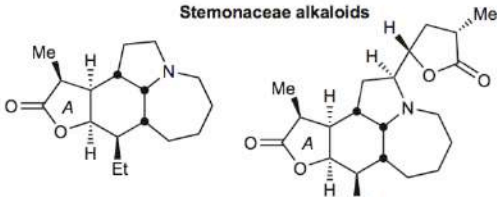
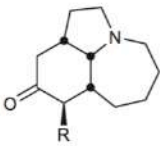
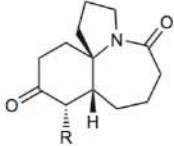
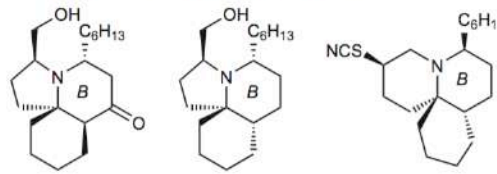
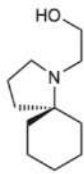
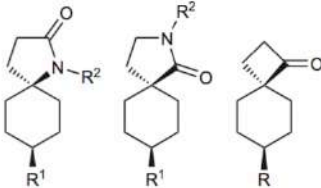
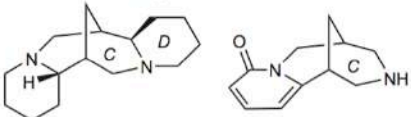
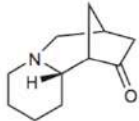
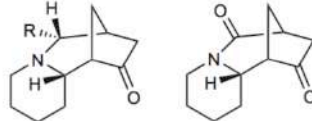
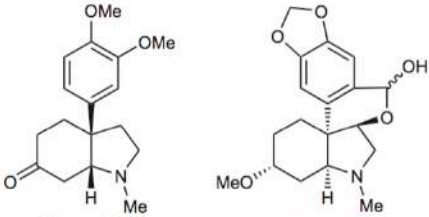
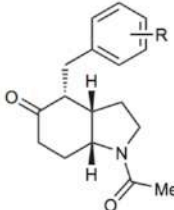
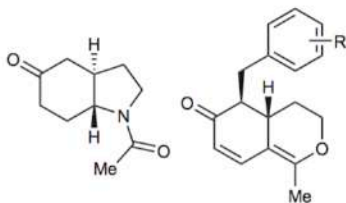
FRAGMENT LIBRARIES

ACB Blocks	www.acbblocks.com	¹⁹ F NMR-oriented, RO3 compliant, predicted to be soluble, purity >96%	1280
Asinex	www.asinex.com	fragment library	22524
Beactica	www.beactica.com	SPRINT: validated for SPR. 2000 purchased fragments	1946
ChemBridge	www.chembridge.com	ChemBridge Fragment Library, RO3 compliant with predicted solubility; minimum purity 90% by ¹ H NMR	>7000
ChemDiv	www.chemdiv.com	3D designed fragment library	4283
Enamine	www.enamine.com	RO3 compliant	18108
		golden	1794
		fragment library (diverse subset of full library), "simple" fragment library: RO3 compliant ≤20 heavy atoms from screening collection	126597
AnCoreX	www.ancorex.com	MetaKel (metal chelating. MW < 300)	>500
		TCI-Frag (targeted covalent inhibitor fragment screening; mildly reactive functionalities, RO3 compliant)	>100
Key Organics	www.keyorganics.net	fragment library	26000
		2nd generation with assured aqueous solubility, RO3 compliant	1166
		fragments from nature: RO3 compliant, assured solubility and high Fsp ³ content	183
		CNS fragment library: more stringent filters (e.g., mw <240)	700
Life Chemicals	www.lifechemicals.com	general	31000
		RO3 compliant (and subsets of predicted soluble, fluorinated, brominated, and Fsp ³ enriched, covalent and PPI focused)	14000
Maybridge	www.maybridge.com	RO3 compliant diversity fragment library with assured solubility in DMSO and PBS buffer; 1000 fragment subset available	2500
		fragment collection, filtered by purity, mw <350 and substructures	>30000
Otava	www.otavachemicals.com	general RO3 compliant, predicted to be soluble	12486
		assured solubility in DMSO and PBS	1000
		fluorine	1217
		metal chelator	1023
		halogen-enriched with bromine for X-ray studies	618
Prestwick Chemical	www.prestwickchemical.com	Prestwick Fragment Library mainly derived from drug fragments, RO3 compliant	910

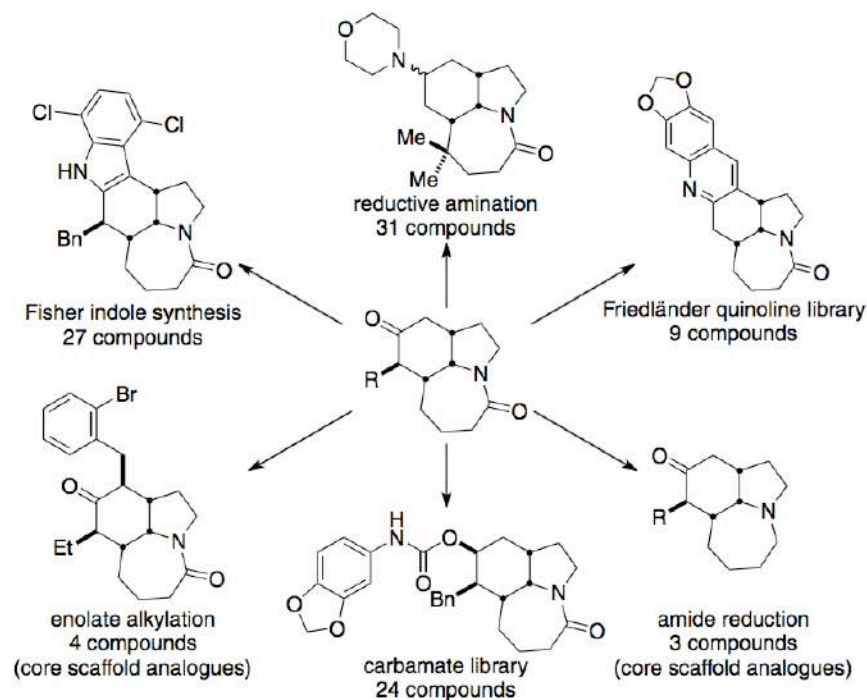
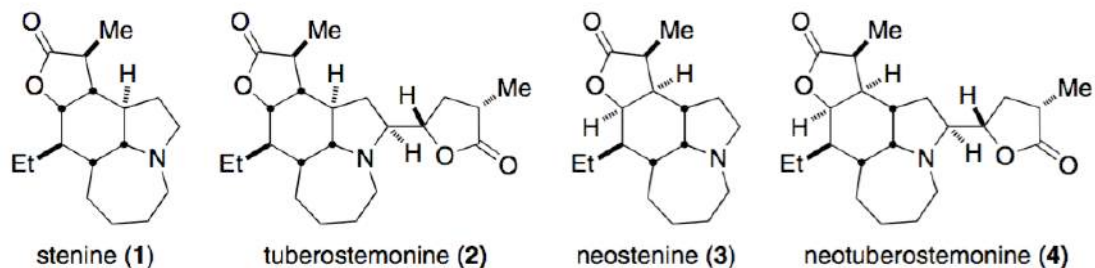
МОЛЕКУЛЯРНЫЙ ДИЗАЙН DE NOVO, ОСНОВАННЫЙ НА ФРАГМЕНТАХ



NATURAL PRODUCT LIBRARIES: ALKALOID-INSPIRED LIBRARY CASE

Natural product family (representative members)	Representative biological activities	Primary scaffold	Secondary scaffold
<p data-bbox="407 205 600 224">Stemonaceae alkaloids</p>  <p data-bbox="291 418 382 436">Neostenine</p> <p data-bbox="513 418 678 436">Neotuberostemonine</p>	<ul style="list-style-type: none"> • Antitussive • Insecticidal/antifeedant/larvicidal • Anthelmintic • Acetylcholinesterase inhibition • P-glycoprotein modulation 	 <p data-bbox="1151 389 1215 408">[ref. 29]</p> <p data-bbox="1122 412 1244 432">Ring A excised</p>	
<p data-bbox="305 468 683 486">Pyrrolo- and pyridoquinoline marine alkaloids</p>  <p data-bbox="253 675 363 694">Cylindricine C</p> <p data-bbox="407 675 517 694">Lepadiformine</p> <p data-bbox="600 675 691 694">Fascicularin</p>	<ul style="list-style-type: none"> • General cytotoxicity • Cardiovascular actions 	 <p data-bbox="1070 682 1296 722">1. Ring B excised 2. Alcohol attachment moved</p>	
<p data-bbox="413 729 542 748">Lupin alkaloids</p>  <p data-bbox="343 875 421 893">Sparteine</p> <p data-bbox="600 875 672 893">Cytisine</p>	<ul style="list-style-type: none"> • Anti-arrhythmic • Acetylcholine agonism • Nicotinic receptor agonism/antagonism 	 <p data-bbox="1103 889 1267 925">1. Ring D excised 2. Ring C contracted</p>	
<p data-bbox="359 925 568 943">Amaryllidaceae alkaloids</p>  <p data-bbox="285 1175 382 1193">Mesembrine</p> <p data-bbox="523 1175 620 1193">Pretazettine</p>	<ul style="list-style-type: none"> • Serotonin reuptake inhibition • 5-Hydroxytryptamine reuptake inhibition • PDE4 inhibition • Acetylcholinesterase inhibition • Antileukemic • Cytotoxicity 	 <p data-bbox="1083 1179 1286 1218">1. Fused bicycle retained 2. Aryl attachment moved</p>	

NATURAL PRODUCT LIBRARIES: ALKALOID-INSPIRED LIBRARY CASE

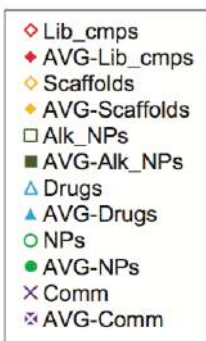
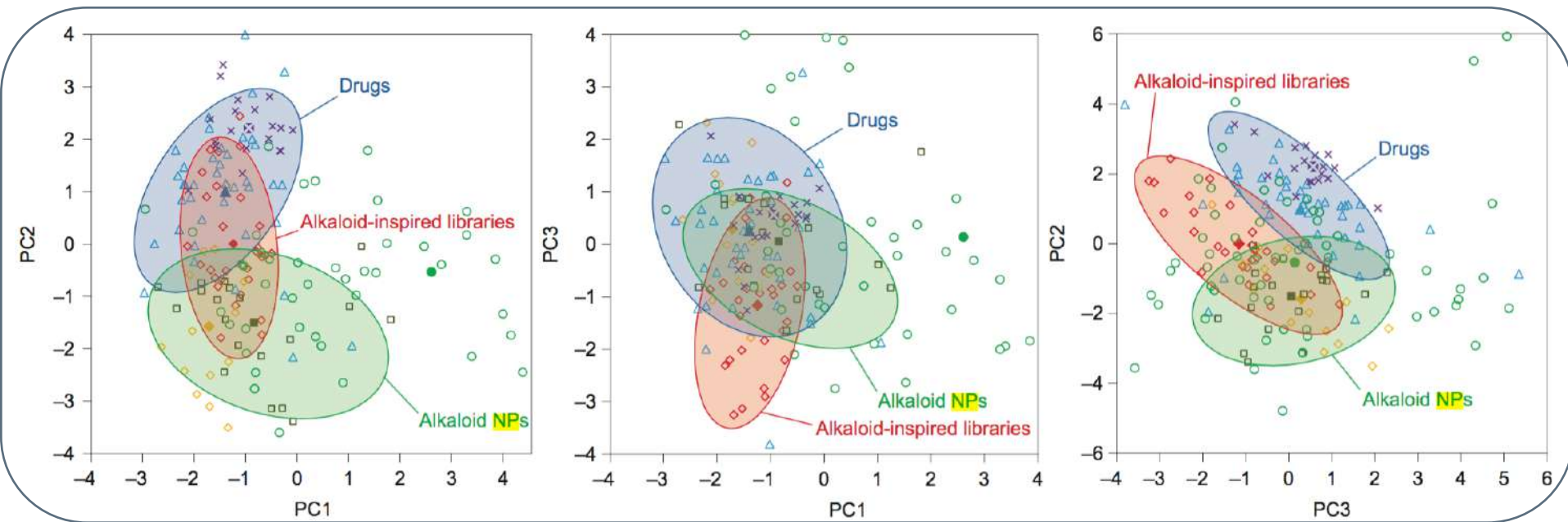


McLeod et al Probing chemical space with alkaloid-inspired libraries (2014) 6 133

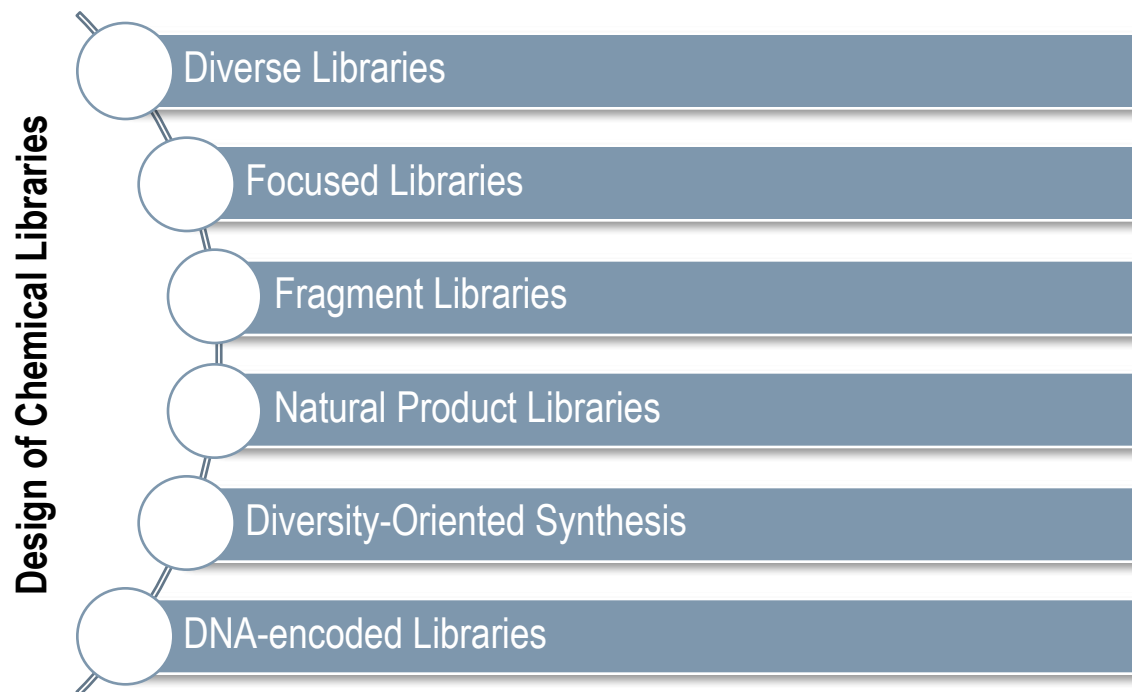
Frankowski et al Synthesis and receptor profiling of *stemona* alkaloid analogues reveal a potent class of sigma ligands
PNAS (2011) 108 (17) 6727-6732

NATURAL PRODUCT LIBRARIES: ALKALOID-INSPIRED LIBRARY CASE

Principal Component Analysis

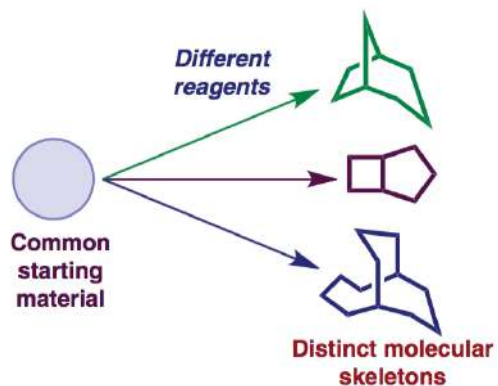


ДИЗАЙН БИБЛИОТЕК СОЕДИНЕНИЙ

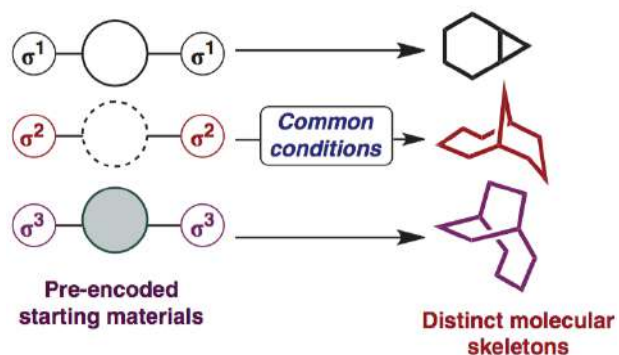


DIVERSITY-ORIENTED SYNTHESIS

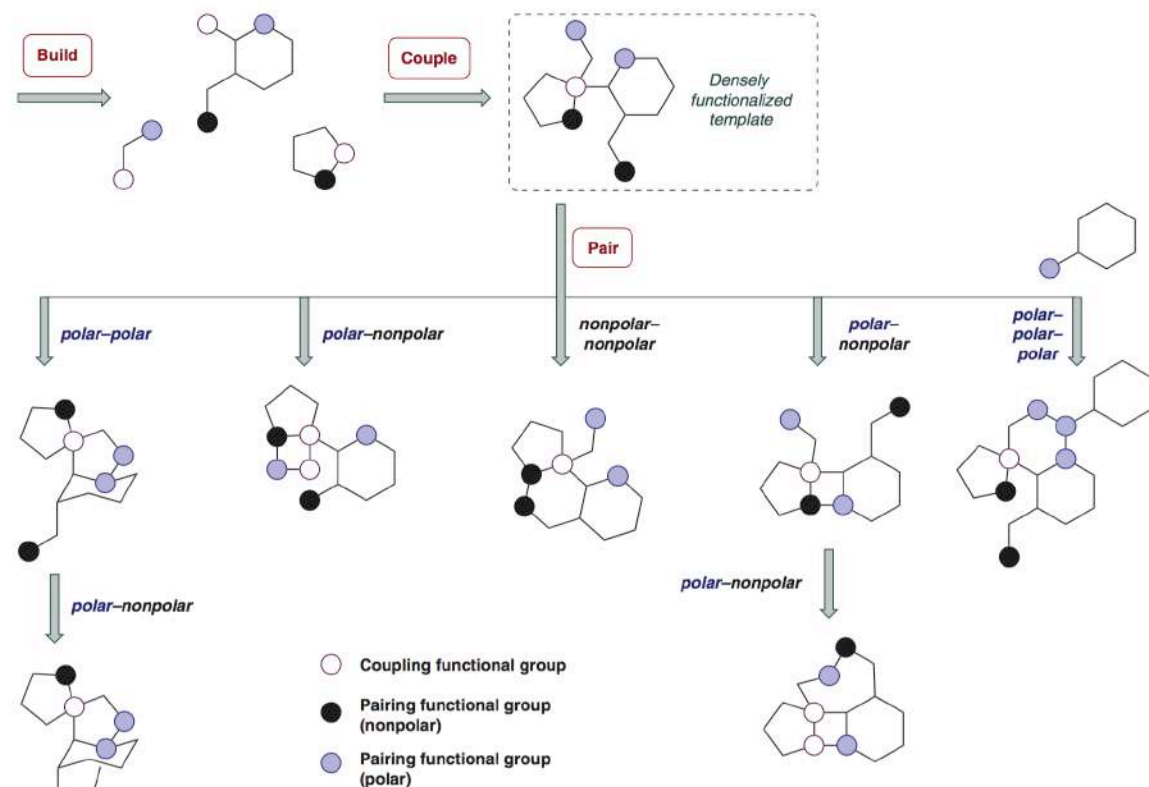
Reagent-based approach



Substrate-based approach



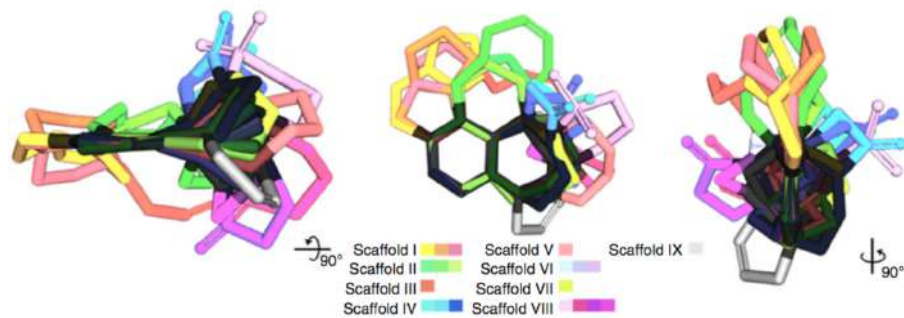
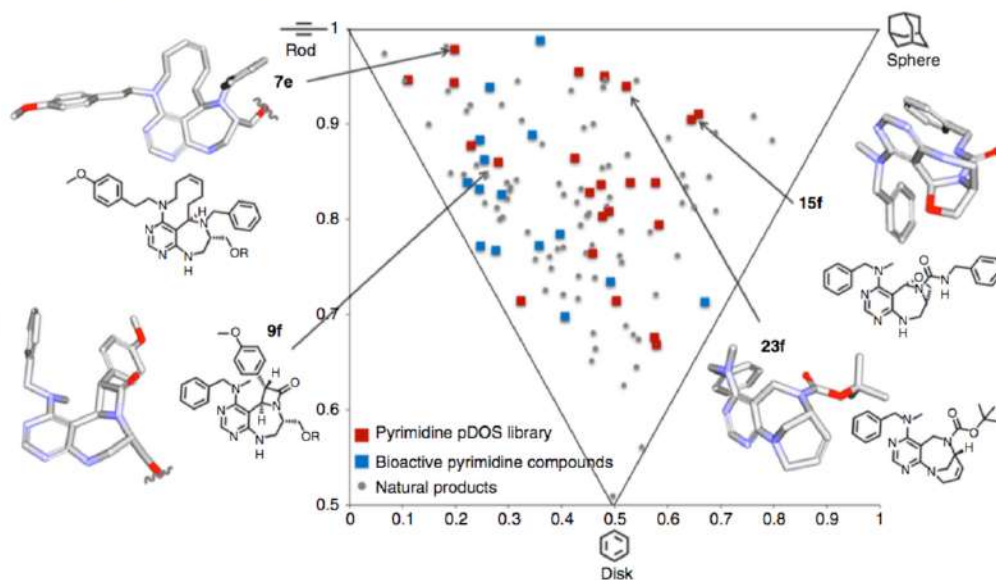
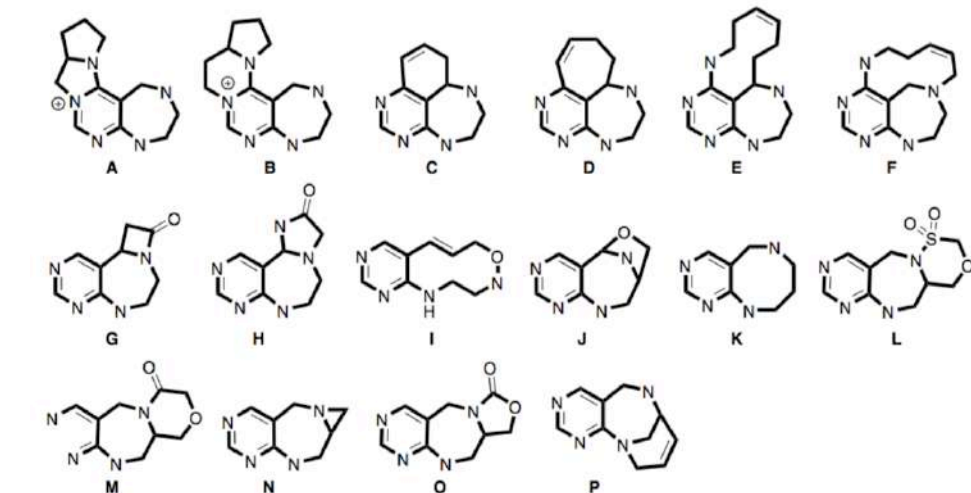
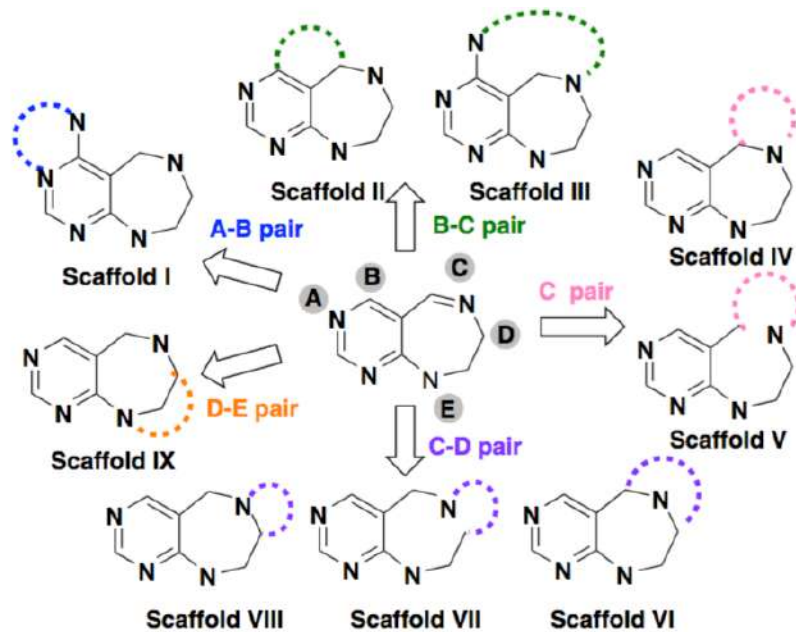
Build/couple/pair strategy



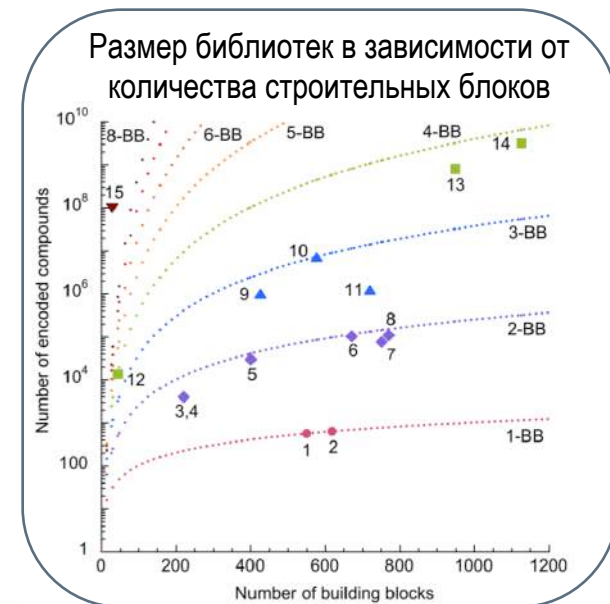
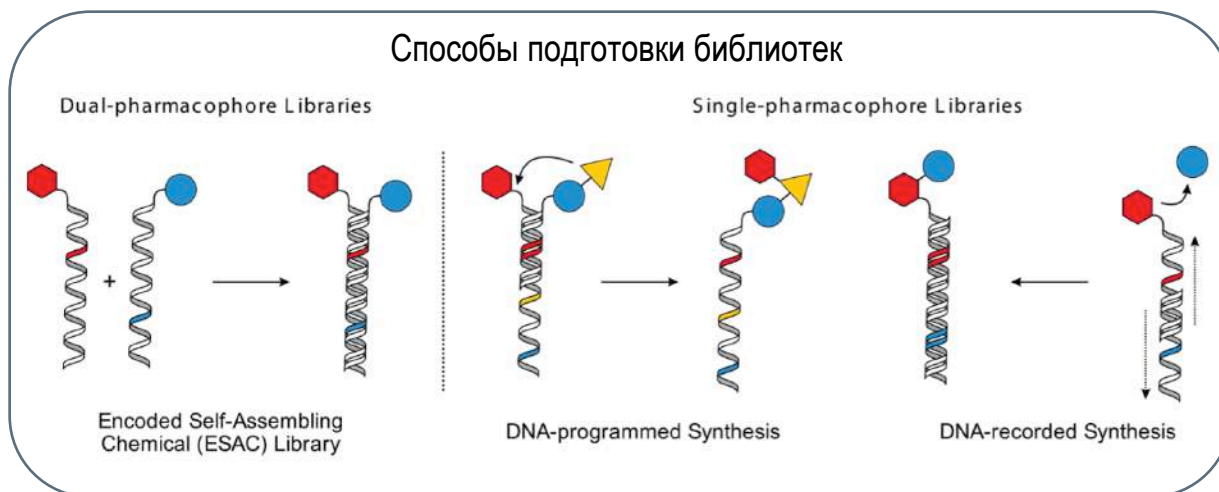
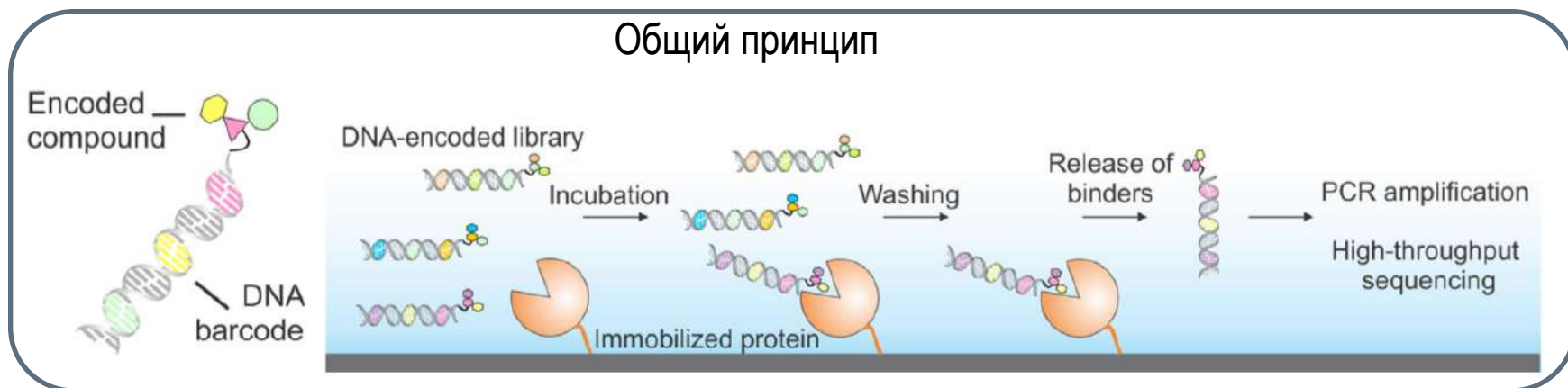
Galloway et al Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules
Nature Commun (2010) 1 80

Kim et al Diversity-oriented synthetic strategy for developing a chemical modulator of protein-protein interaction
Nature Commun (2016) 7 13196

DIVERSITY-ORIENTED SYNTHESIS



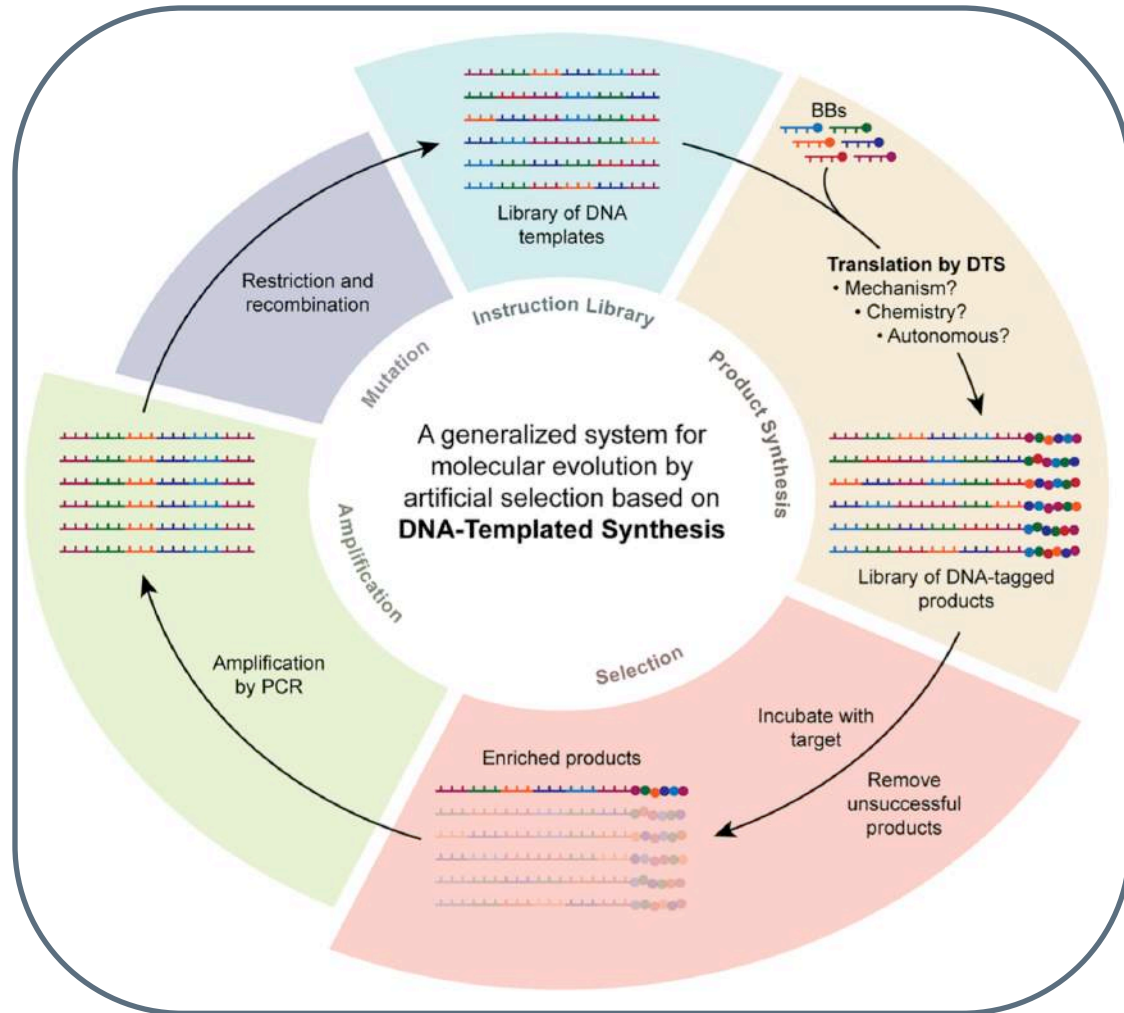
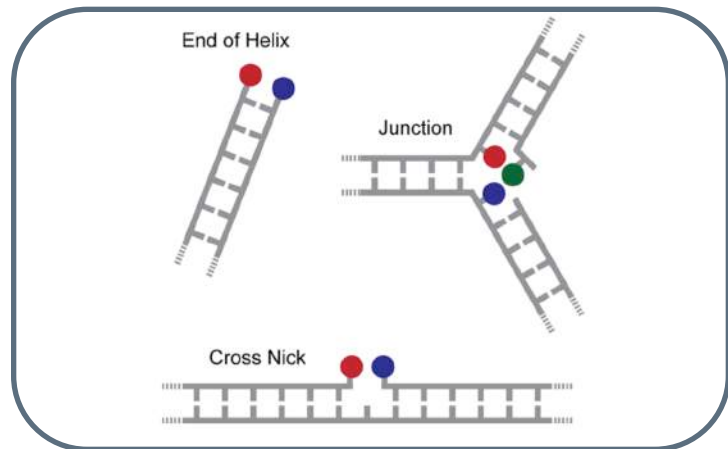
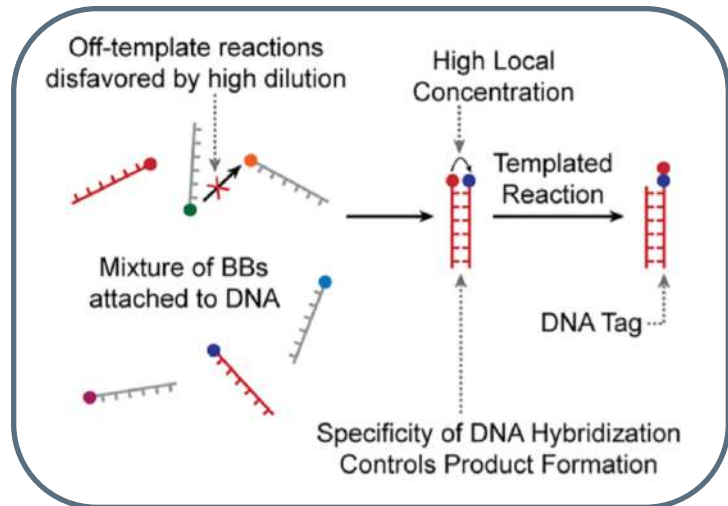
DNA-ENCODED LIBRARIES: DRUG DESIGN



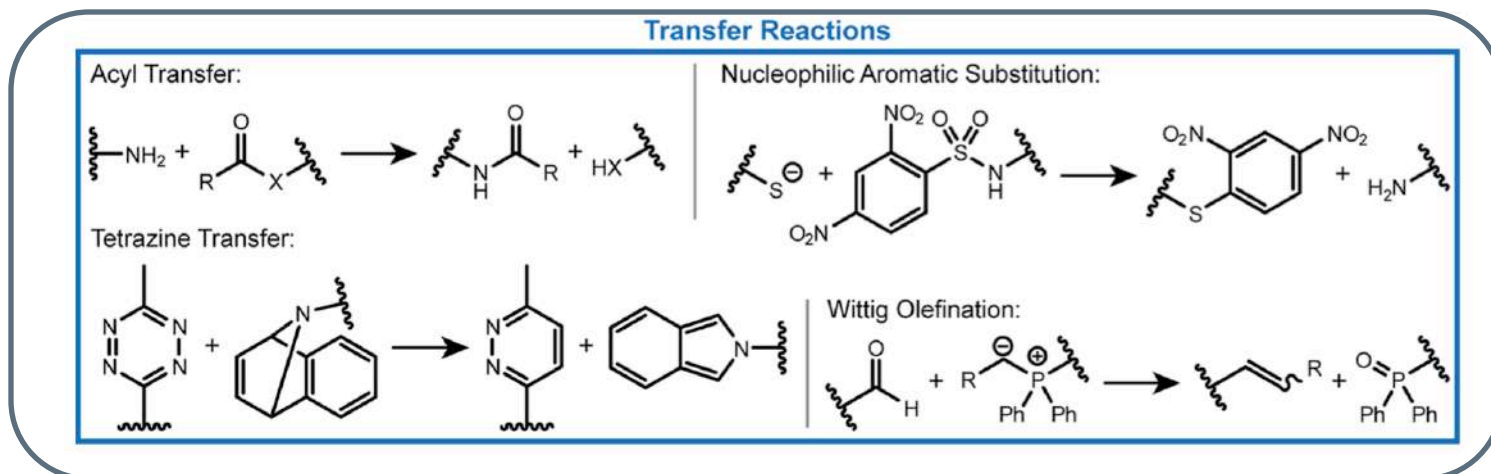
Franzini et al **DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries** *Acc Chem Res* **2014** 47 (4), 1247-1255

Yuen et al **Achievements, Challenges, and Opportunities in DNA-Encoded Library Research: An Academic Point of View** *ChemBioChem* **2017**, 18, 1 – 9

DNA-ENCODED LIBRARIES



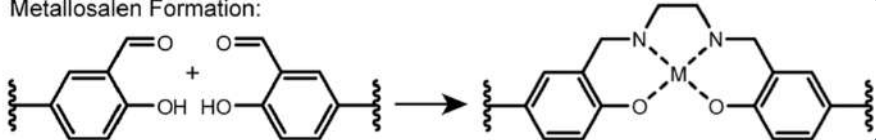
DNA-ENCODED LIBRARIES: MOST COMMON REACTION TYPES



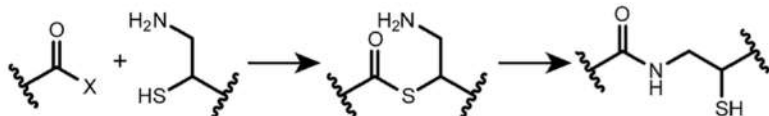
- DNA adapter cleavage
- Best yields per step -> 80%

DNA-ENCODED LIBRARIES: COMMON REACTION TYPES

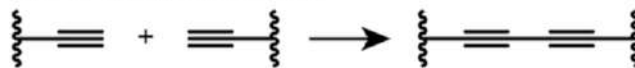
Metallosalen Formation:



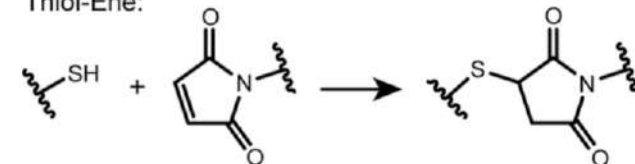
Native Chemical Ligation:



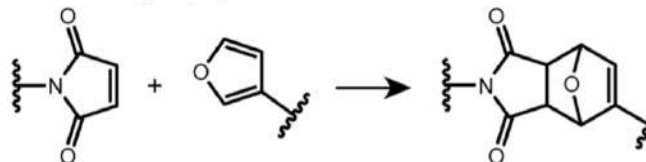
Glaser-Eglinton Coupling:



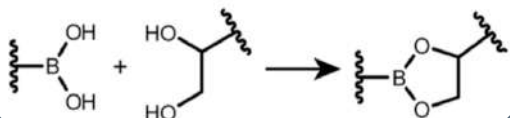
Thiol-Ene:



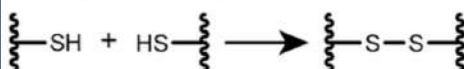
Diels-Alder [4+2] Cycloaddition:



Boronic ester formation:



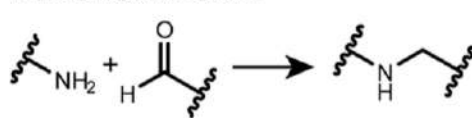
Disulfide Formation:



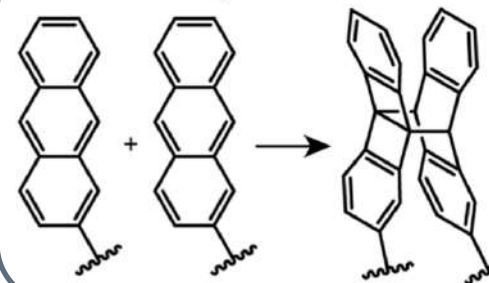
Copper-Catalyzed Azide-Alkyne Cycloaddition:



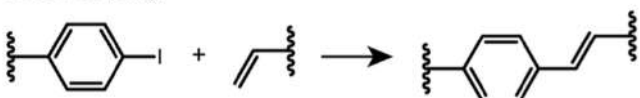
Reductive Amination:



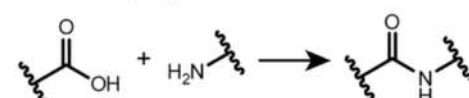
Photocrosslinking:



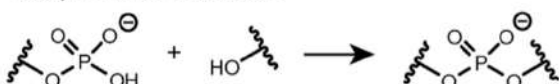
Heck Coupling:



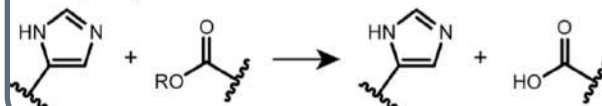
Amide Coupling:



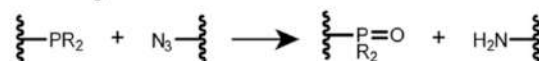
Phosphodiester Formation:



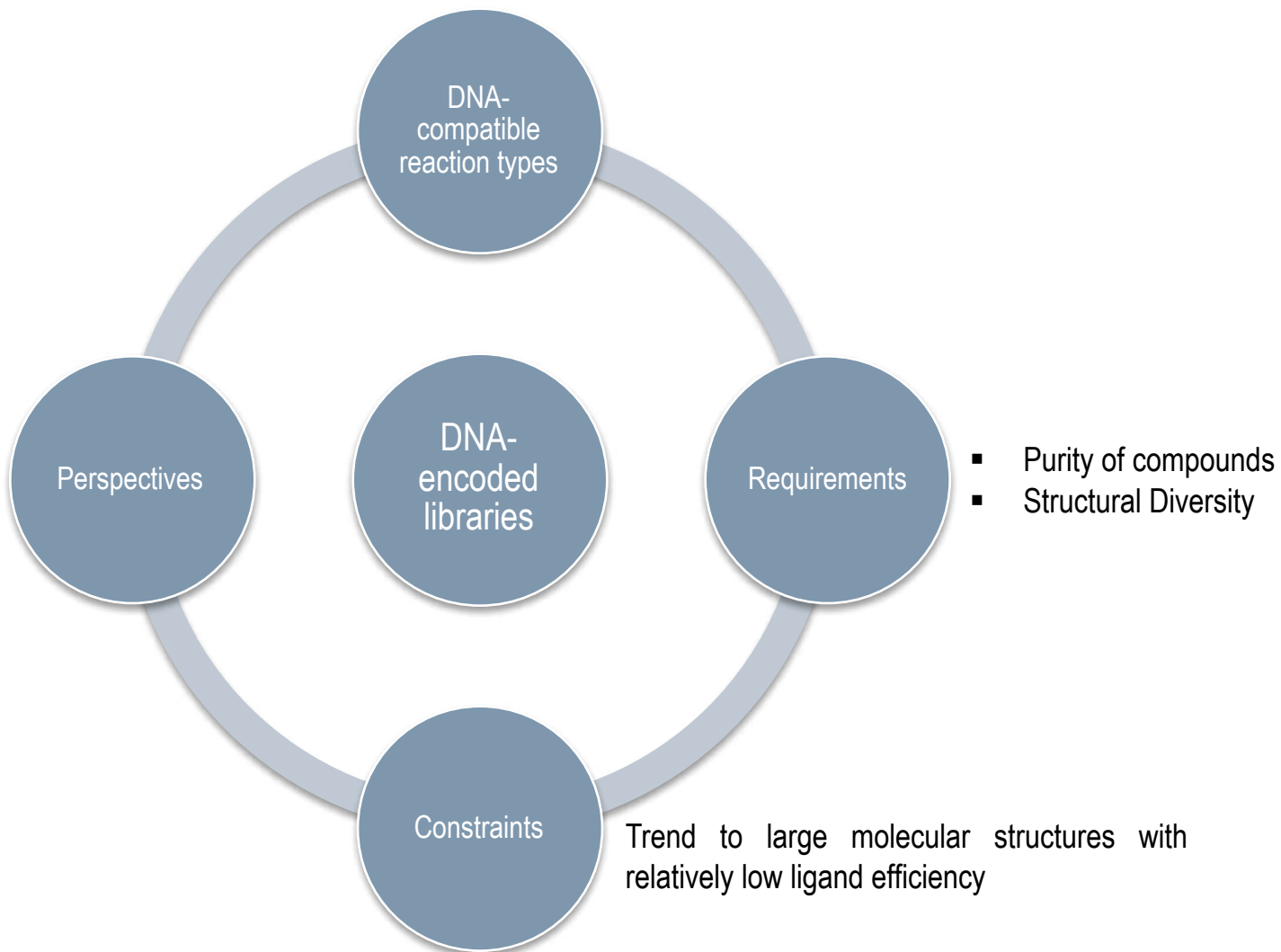
Ester Hydrolysis:



Staudinger Reduction:



DNA-ENCODED LIBRARIES

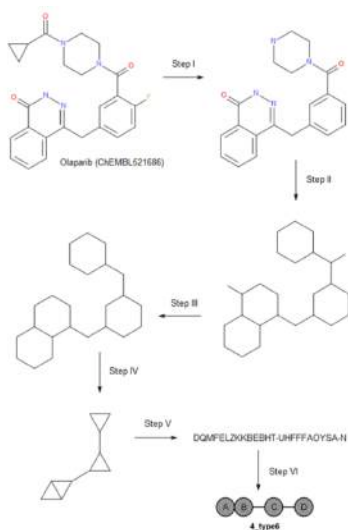


Cheminformatics insight:

- Assisting in developing novel and optimizing actual synthetic schemes
- Diversity analysis
- "Chemogenomics"-related tasks: association of target families with given type of DNA-encoded library
- Ligand efficiency and structural complexity control
- Filtering undesired structures invoking DNA reactivity (oxidative damage, alkylation, deamination etc)

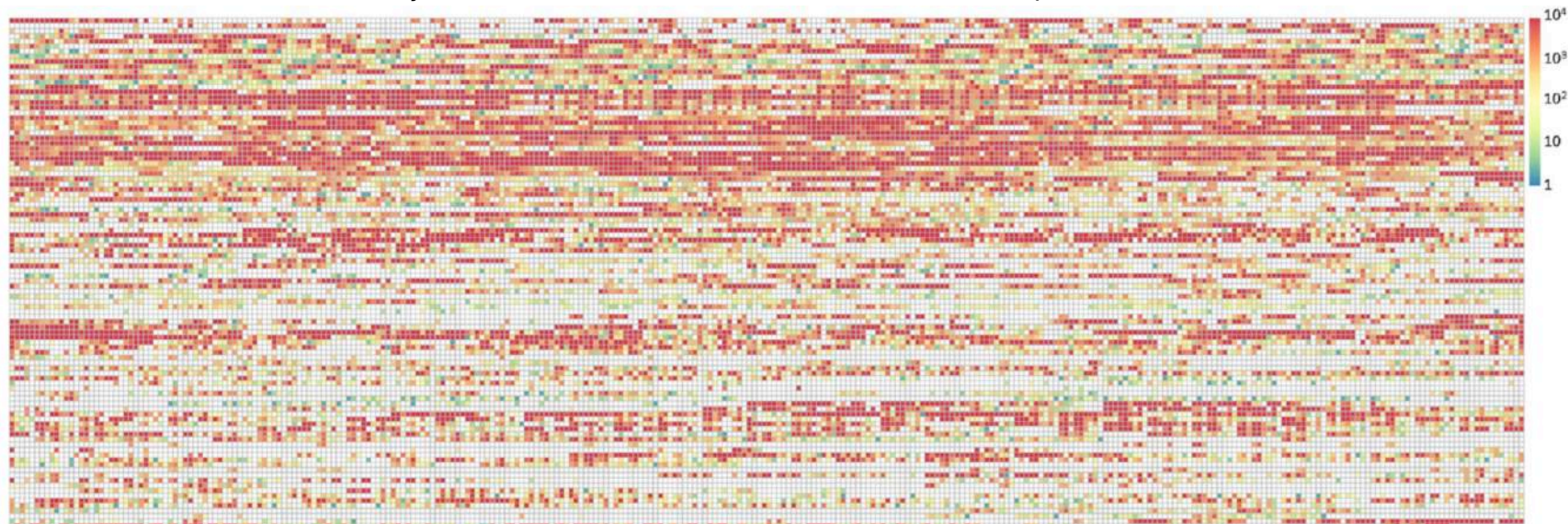
MAPPING OF DRUG-LIKE CHEMICAL UNIVERSE WITH REDUCED COMPLEXITY MOLECULAR FRAMEWORKS

Reduced Complexity Molecular Framework (RCMF) + three descriptors types (ring, linker, bond angle information)

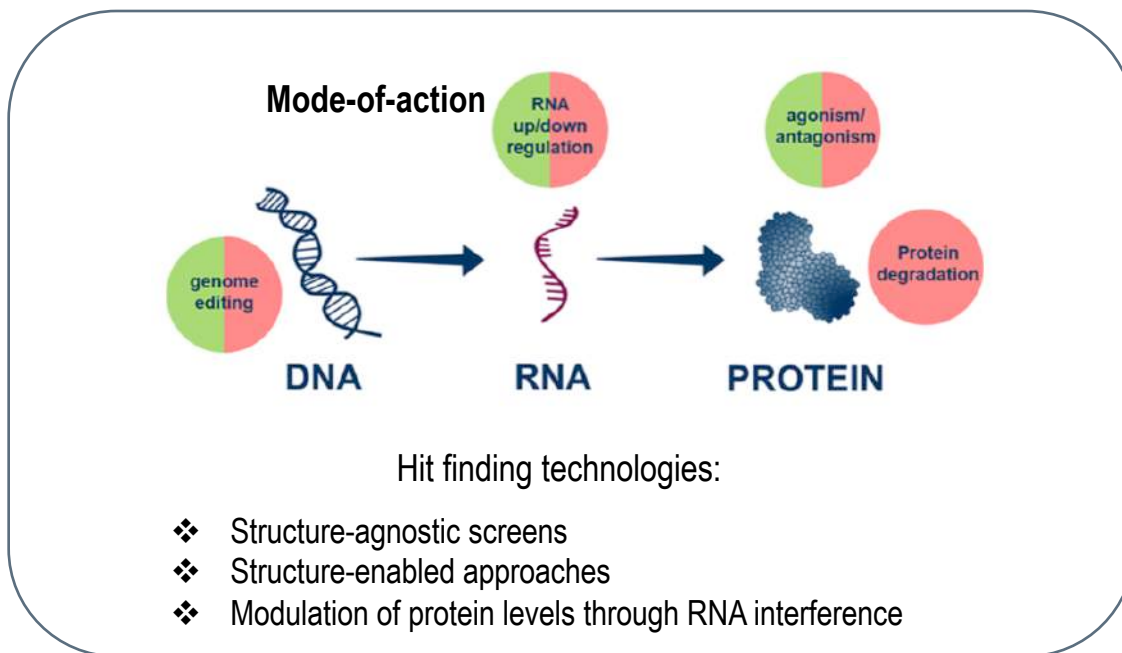


- I) Generation of a Murcko scaffold (stereochemistry and 3-member terminal rings are removed)
- II) Conversion of all atoms to carbons, conversion of all bonds to single
- III) Extraction of Murcko scaffold
- IV) All linkers are reduced to become a single bond, all rings are reduced to a minimal possible size to keep original connectivity patterns between the rings
- V) InChIKey is generated for the structure formed at previous step
- VI) Comparison of InChIKey to a database of predefined set of InChIKeys and retrieving their corresponding RCM framework type, which is "4_type6" for Olaparib

Projection of four Nuevolution DELs on 2D RCMF map



МЕХАНИЗМ ДЕЙСТВИЯ (MODE-OF-ACTION) КАК ОСНОВА ОПРЕДЕЛЕНИЯ СТРАТЕГИИ СКРИНИНГА НОВЫХ СОЕДИНЕНИЙ

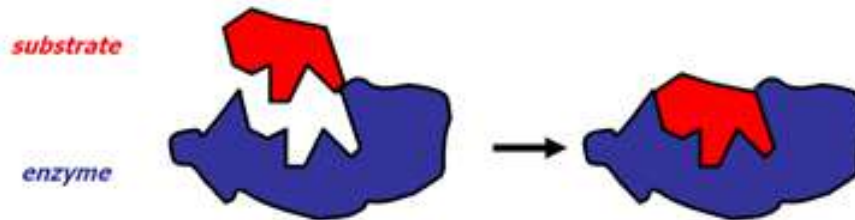


Examples of possible cheminformatics support:

- ❖ Aid new technologies (e.g. DNA-encoded libraries DEL)
- ❖ Data sharing and analysis
- ❖ Integrating new machine learning technologies (Deep learning, one-shot-learning, machine learning for dynamic systems, etc), rational use of already known approaches

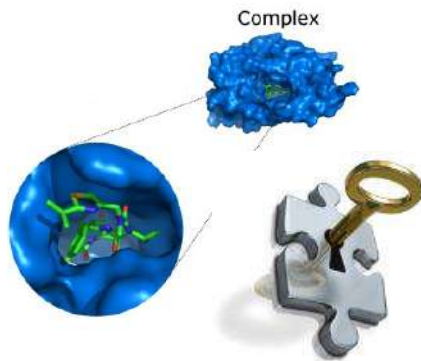
PROTEIN AGONISM / ANTAGONISM: DOCKING AND 3D QSAR METHODS

Принцип «ключ-замок»



Structure- and ligand- based methods

Docking



3D QSAR

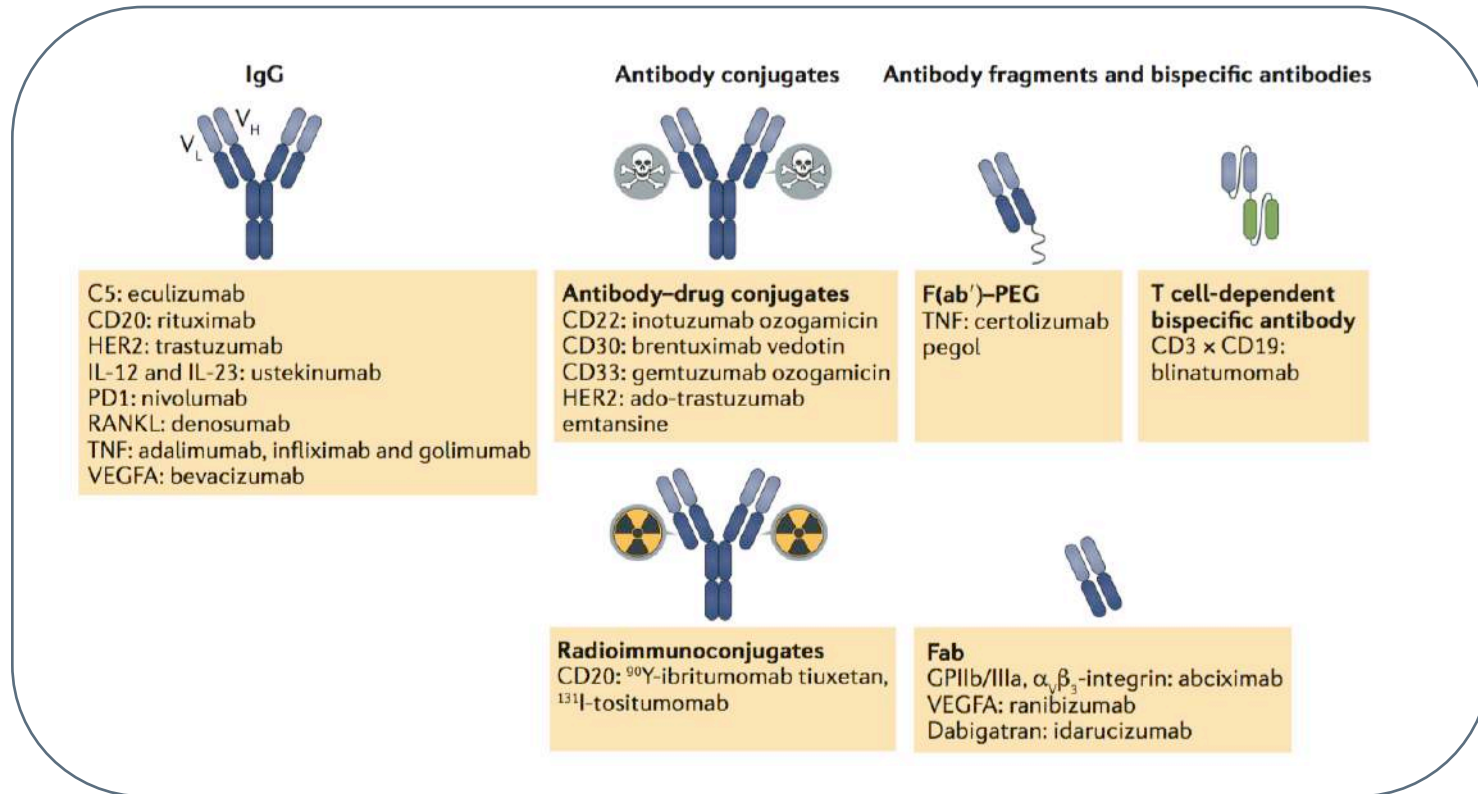
Большая часть алгоритмов
3D QSAR



3D QSAR, основанный
на знании строения
рецептора



АГОНИЗМ / АНТАГОНИЗМ: АНТИТЕЛА

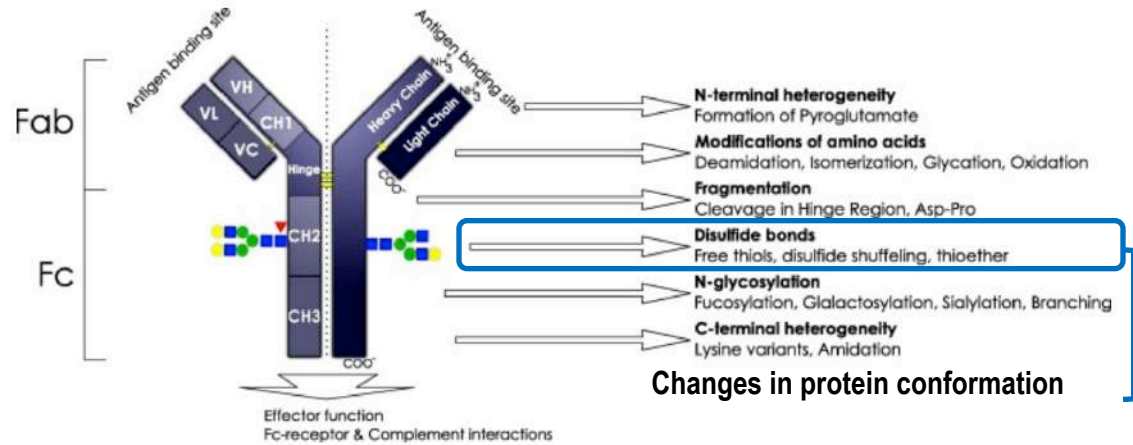
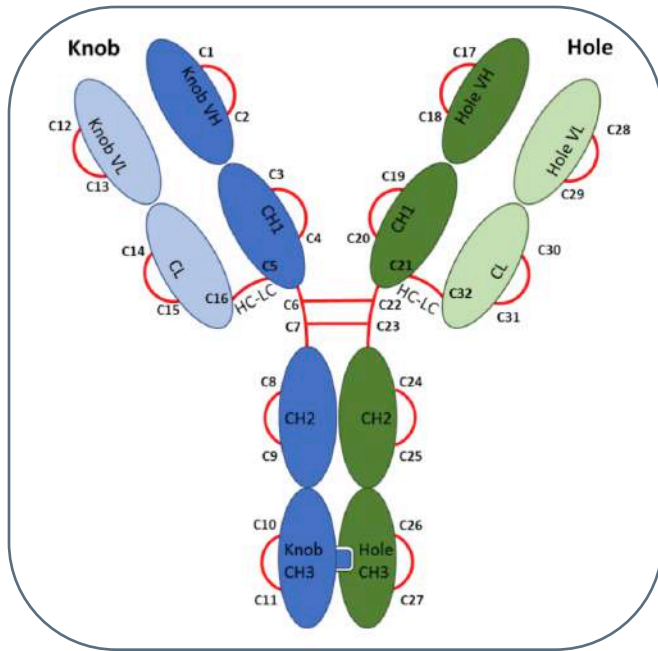


Antibody agonism occurs when receptor engagement by the antibody results in receptor activation and signalling. Antibodies have several advantages over natural ligands such as greater thermal or in-vivo stability and longer plasma half-lives

Types of receptor activation:

- Receptor activation through dimerization
- Receptor activation through clustering

АГОНИЗМ / АНТАГОНИЗМ: АНТИТЕЛА



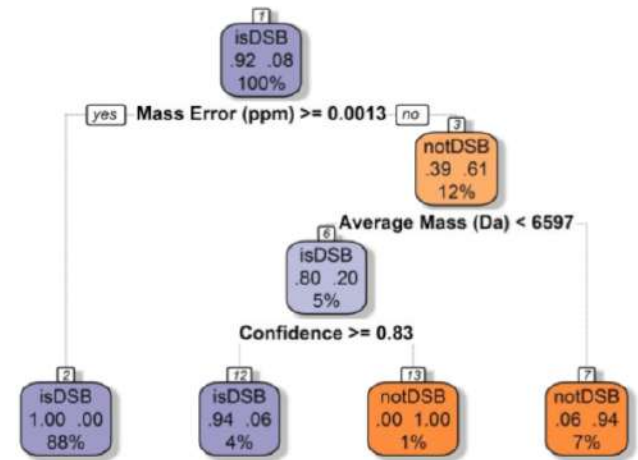
Analysis of bispecific antibody to ensure that the protein is properly folded with minimal traces of cysteine-related by-products that may affect structural integrity is required

Liquid Chromatography + Mass Spectrometry (LC-MS): analysis of cysteine-related characteristics (verification of expected disulfide bonds and identification of mispaired disulfide bonds)

Machine learning: analysis of LC-MS data using Decision trees with stochastic gradient descent

Quantitative attributes such as MS areas, monoisotopic mass, and confidence score were used for identifying observations with greater than zero monoisotopic mass, whose origins and significance were unknown followed by using for model obtaining

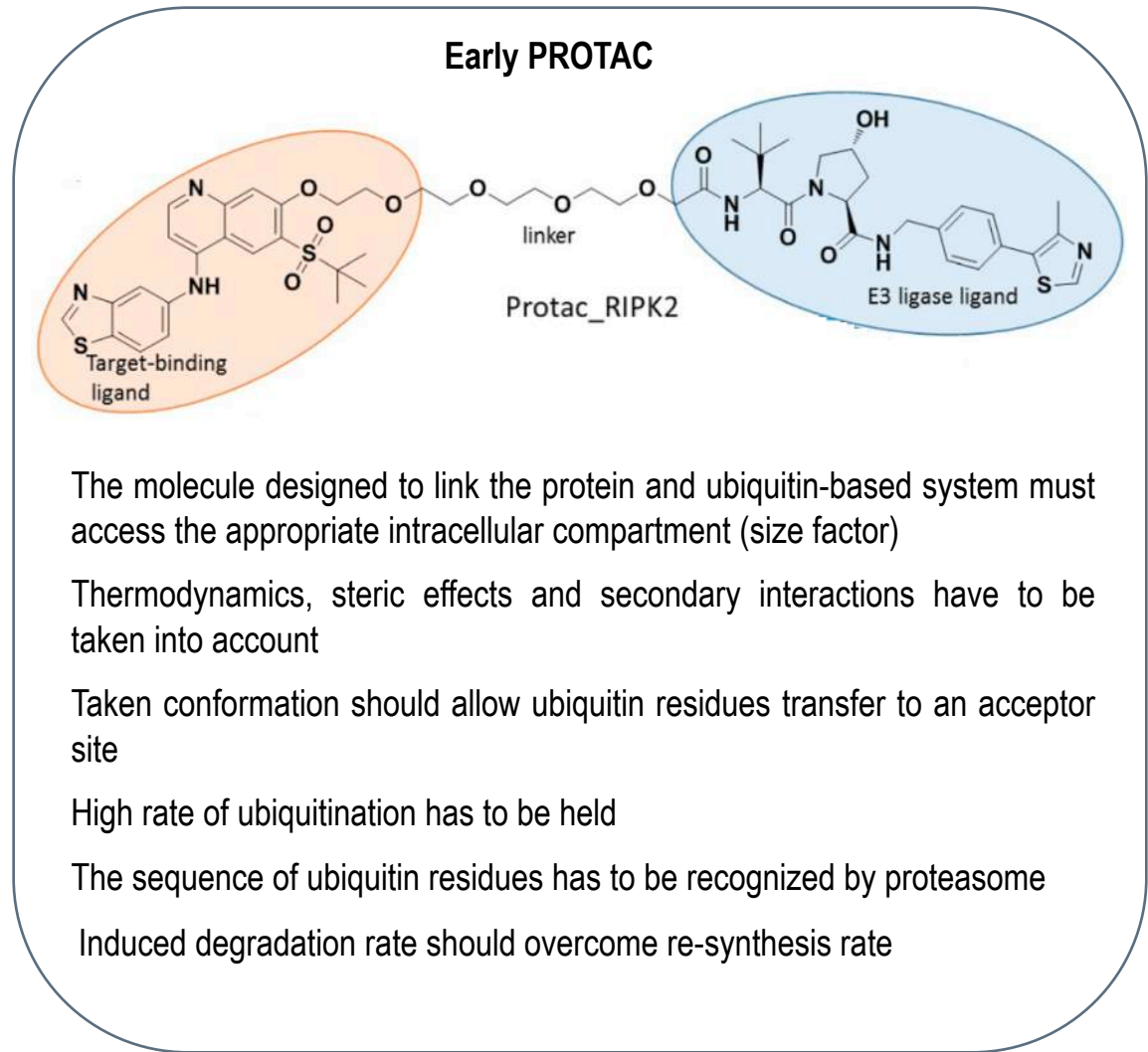
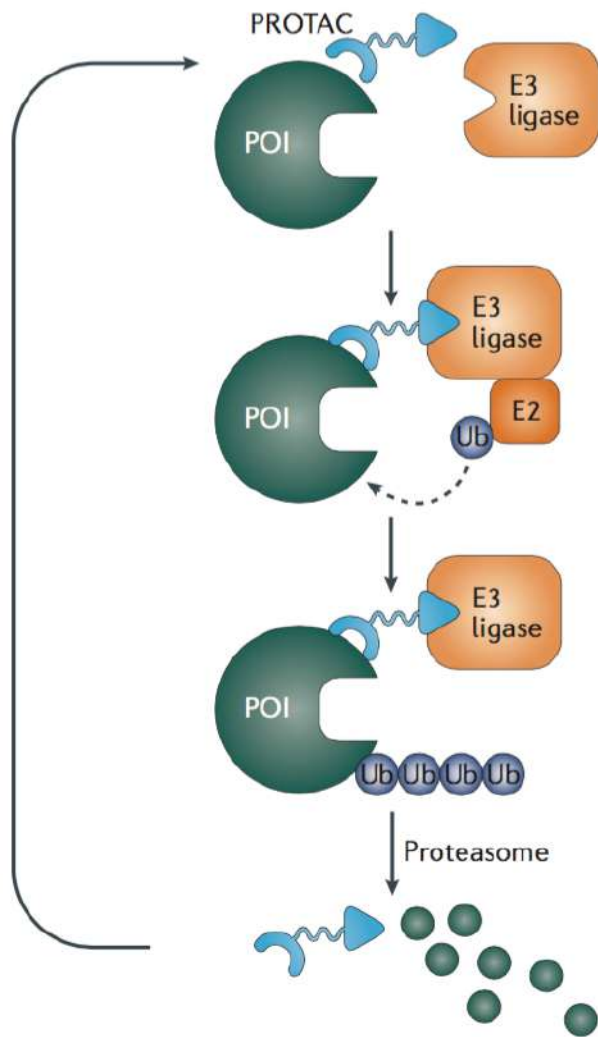
Decision trees + stochastic gradient boosting



Baker et al **Rapid Identification of Disulfide Bonds and Cysteine-Related Variants in an IgG1 Knob-into-Hole Bispecific Antibody Enhanced by Machine Learning**
Analytical Chemistry 2019 91 (1), 965-976

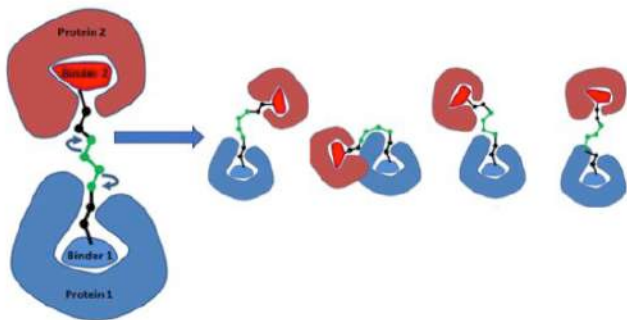
Higel et al **N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins**
European Journal of Pharmaceutics and Biopharmaceutics (2016) 100 94

УБИКВИТИН-ЗАВИСИМАЯ ДЕГРАДАЦИЯ БЕЛКА



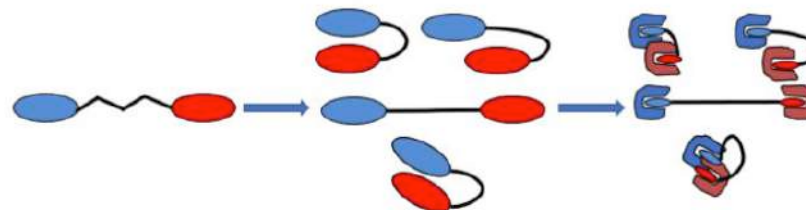
IN SILICO MODELING OF PROTAC-MEDIATED TERNARY COMPLEXES

Method 1



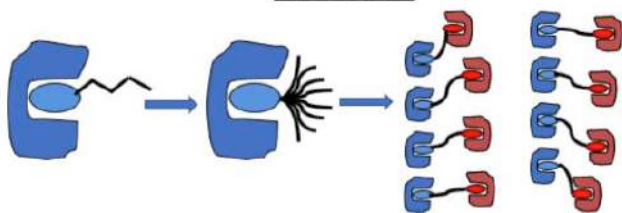
The entire ternary complex is sampled

Method 2



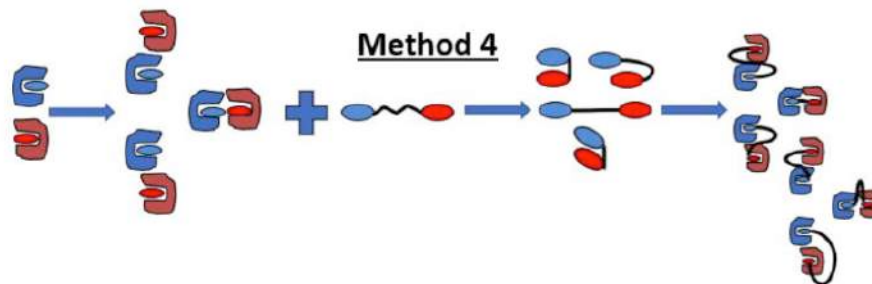
PROTAC conformations are sampled independently

Method 3



PROTAC is sampled based on one of the proteins structure

Method 4



PROTAC conformations are sampled independently, ligase-protein arrangements were evaluated