- Chinzo Dandar
- Khanh Tang
- Julia Khurelbataar
- Ben Tingle
- John Guttierez

- Jiankun Lyu
- Stefan Gahbauer
- Brian Bender
- Trent Balius

- Isha Singh
- Anat Levit
- Brian Shoichet
- Ying Yang
- Chase Webb
- Shuo Gu
- Shiming Peng

Thanks!

Roth Lab @ UNC
Bryan Roth
Sheng Wang
Tao Che

NextMove Software
Roger Sayle
John Mayfield

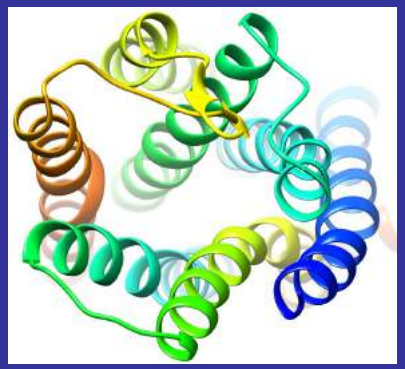
OpenEye Scientific
Molinspiration
Molecular Networks
ChemAxon
Schrodinger

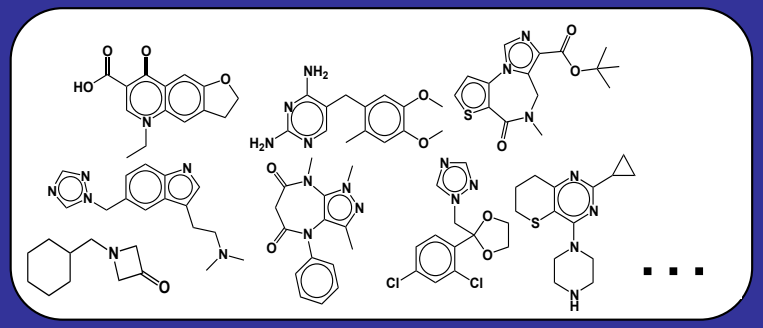
Enamine:
Yurii Moroz
A. Tolmachev

NIGMS
GM71896
GM133836

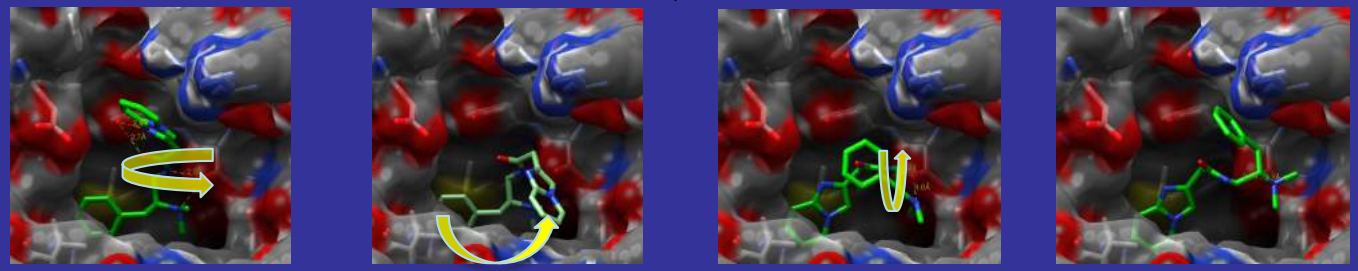
DARPA

docking large libraries for new chemotypes





library of ***available*** molecules



...

each in ~10^6 orientations & conformations

Score for fit

Test high-ranking molecules

Outline

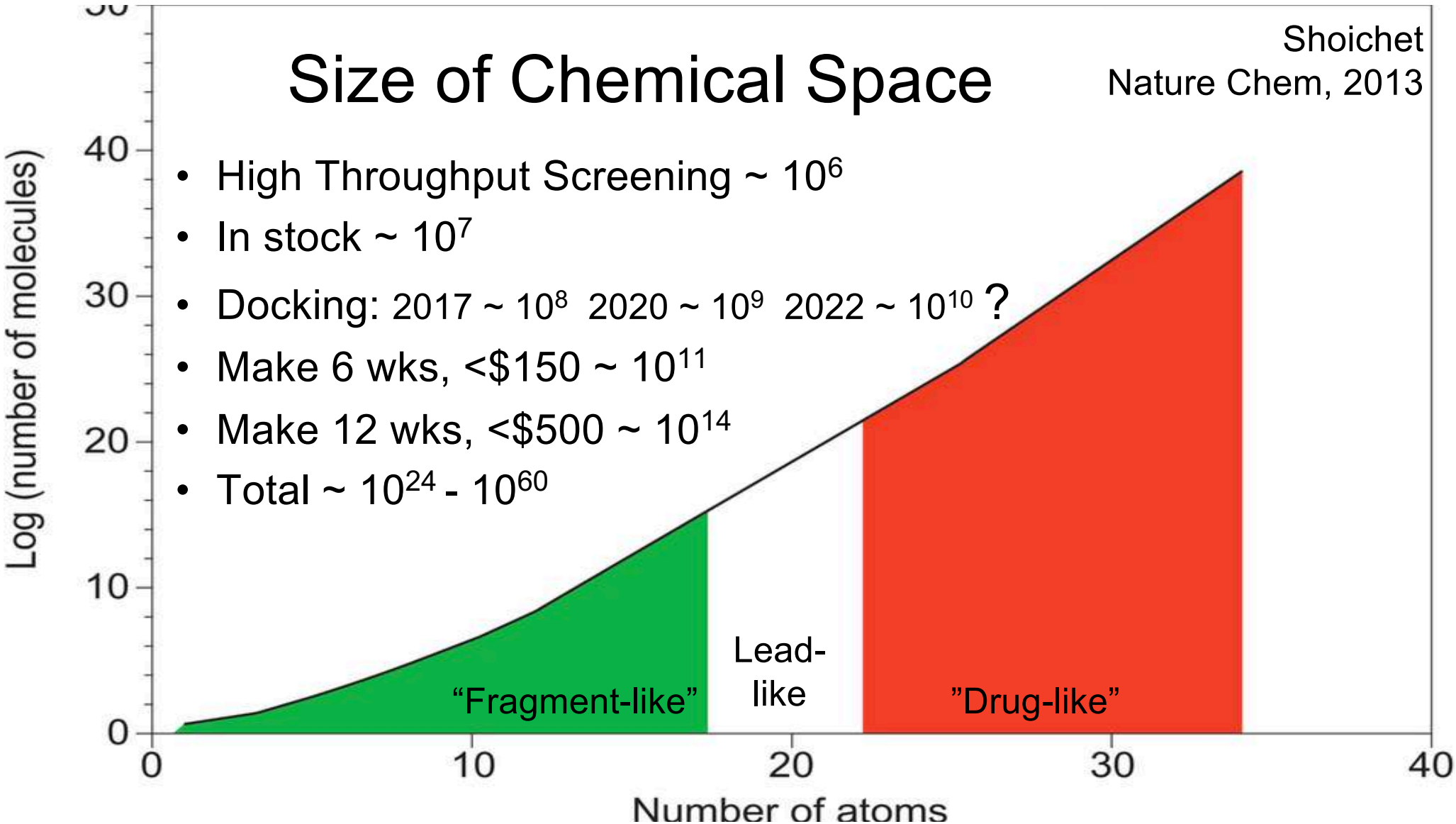
- Chemical space is big
  - …but it is where we find new drugs
  - and can now be accessed quickly at low cost
- Docking screens in lead-like space with docking
  - How does it work
  - What can I expect?

- Free public resources <u>you can use</u>
  - Smallworld, Arthor, DataWarrior
  - ZINC, SEA, and much more

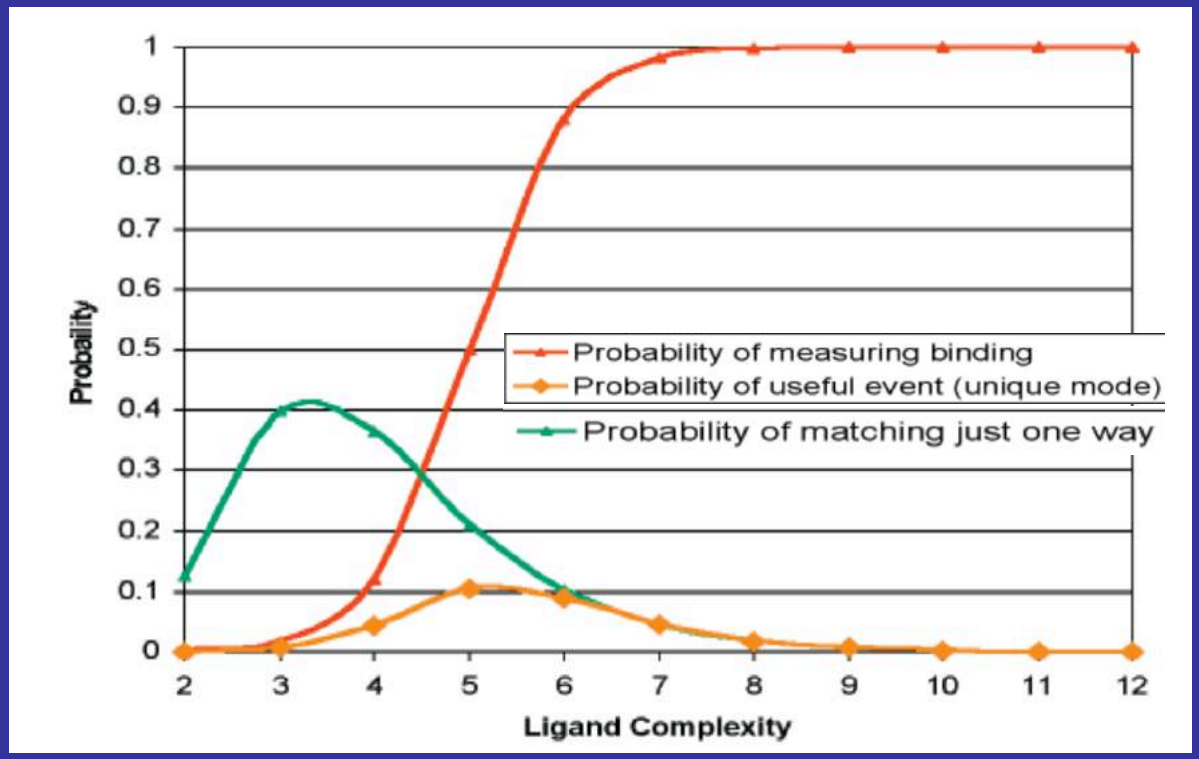
Size of Chemical Space

Shoichet
Nature Chem, 2013

- High Throughput Screening ~ 10^6
- In stock ~ 10^7
- Docking: 2017 ~ 10^8 2020 ~ 10^9 2022 ~ 10^{10} ?
- Make 6 wks, <$150 ~ 10^{11}
- Make 12 wks, <$500 ~ 10^{14}
- Total ~ 10^{24} - 10^{60}



Why fragments?
The Hann complexity argument



Feature Position	1	2	3	4	5	6	7	8	9
Receptor features	-	-	+	-	+	-	-	+	-
Ligand mode 1	+	+	-						
Ligand mode 2						+	+	-	
(reverse mode 3)					-	+	+		
(end wrap mode 4)	+							-	+

Hann, Leach, Harper JCICS, 2001

Molecules you can buy

24M
(55M)
((99M))

625M
(676M)
((5B))

628M
(645M)
((100B))

LogP (up to)	200	250	300	325	350	375	400	425	450	500	>500	LogP
-1	35K	280K	1.3M	1.9M	3.7M	1M	304K	68K	44K	25K	5.4K	8.7M
0	174K	1.5M	6.3M	8.8M	17M	4.9M	2M	599K	403K	225K	4K	42M
1	468K	4.6M	20M	27M	53M	17M	9.2M	3.4M	2.4M	1.4M	8.5K	138M
2	624K	7.7M	37M	51M	102M	38M	26M	12M	8.8M	5.3M	23K	288M
2.5	240K	3.8M	21M	30M	61M	27M	21M	12M	8.8M	5.5M	24K	191M
3	140K	2.9M	19M	28M	56M	30M	25M	16M	12M	7.9M	39K	196M
3.5	63K	1.8M	14M	22M	45M	29M	26M	18M	15M	10M	64K	182M
4	19K	830K	8.3M	12M	18M	21M	24M	19M	17M	12M	95K	132M
4.5	2.5K	234K	4M	6.8M	12M	16M	19M	17M	16M	12M	132K	102M
5	93	35K	1.3M	2.9M	6.1M	9.4M	13M	13M	13M	10M	160K	69M
>5	28	891	46K	179K	557K	1.2M	2.1M	2.7M	3M	2.5M	818K	13M
Totals, by Weight	1.8M	24M	132M	190M	375M	195M	166M	114M	97M	67M	1.4M	1363M Substances 1.7K Tranches

Cheap
(all)
((ucsf))

3D Molecules you can buy

~~24M~~
20M

~~625M~~
281M

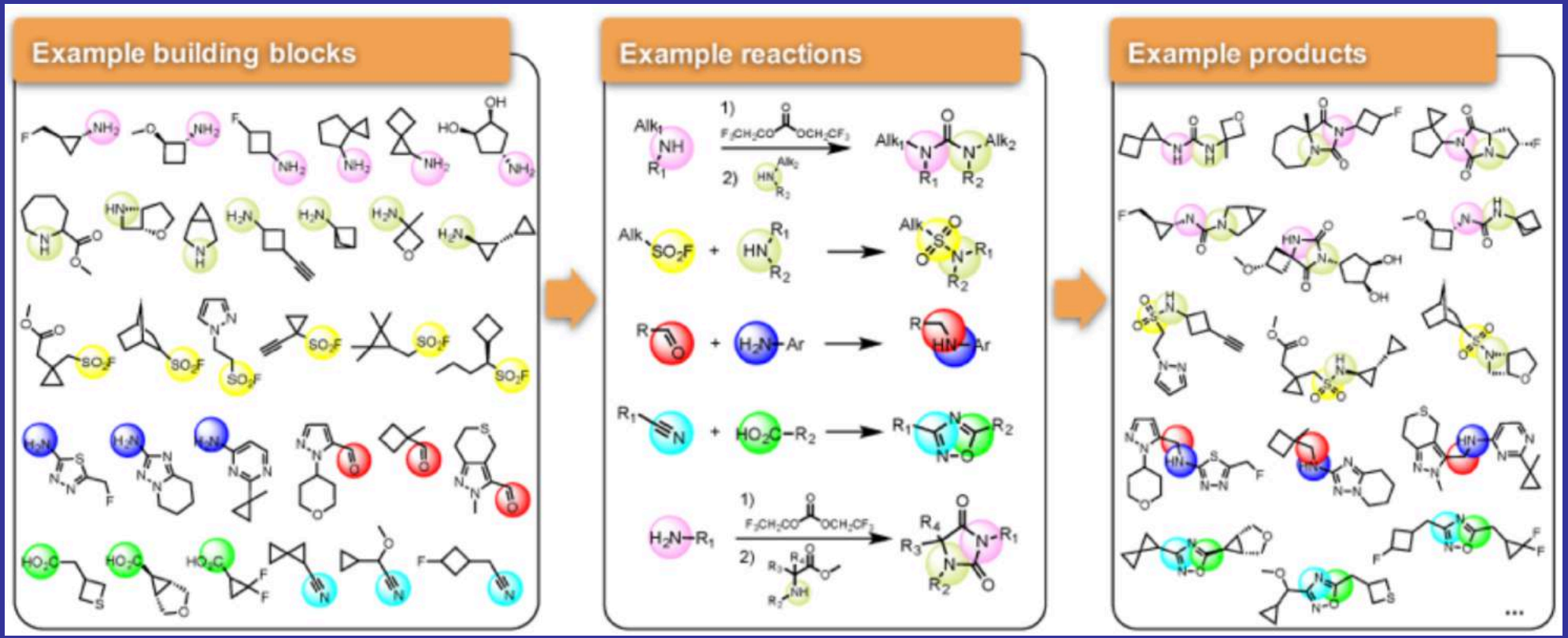
~~628M~~
142M

LogP (up to)	200	250	300	325	350	375	400	425	450	500	>500	LogP
-1	33K	77K	1.1M	1.4M	1.8M	372K	275K	65K	40K	19K	2.1K	5.2M
0	134K	745K	5.2M	7.7M	10M	2M	529K	296K	296K	156K	2.4K	27M
1	414K	3M	16M	10M	19M	5.6M	894K	298K	240K	301K	6.2K	57M
2	586K	7.1M	20M	37M	15M	24M	3.2M	341K	60K	311K	19K	107M
2.5	228K	3.5M	15M	16M	16M	11M	2.4M	393K	73K	65K	21K	65M
3	131K	2.7M	14M	9.6M	20M	28M	3M	439K	109K	317K	37K	79M
3.5	55K	1.2M	13M	20M	11M	13M	2.5M	183K	147K	458K	61K	62M
4	14K	390K	6.1M	7.2M	9.3M	3.2M	19M	5.3M	177K	466K	91K	51M
4.5	1.8K	50K	542K	6M	5.1M	4.7M	1.5M	4.4M	175K	519K	100K	23M
5	93	3.8K	93K	297K	67K	136K	111K	136K	145K	513K	156K	1.7M
>5	27	767	13K	49K	42K	107K	113K	167K	205K	440K	778K	1.9M
Totals, by Weight	1.6M	19M	92M	115M	108M	93M	34M	12M	1.7M	3.6M	1.3M	**481M** Protomers **6.5K** Tranches

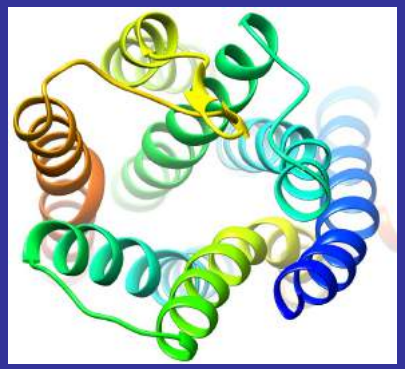
East Outskirts of Kiev – looking E

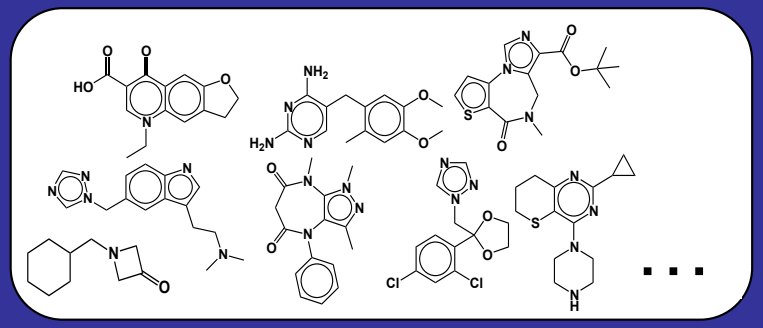
BB – 4 C right – RT left. #BB>200,000

make-on-demand library:
109 2-component reactions
200,000 building blocks

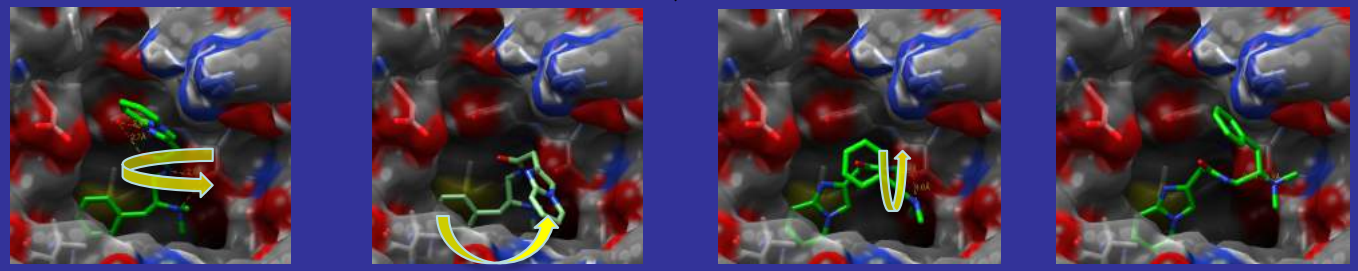


docking large libraries for new chemotypes





library of ***available*** molecules



...

each in ~10^6 orientations & conformations

Score for fit

Test high-ranking molecules

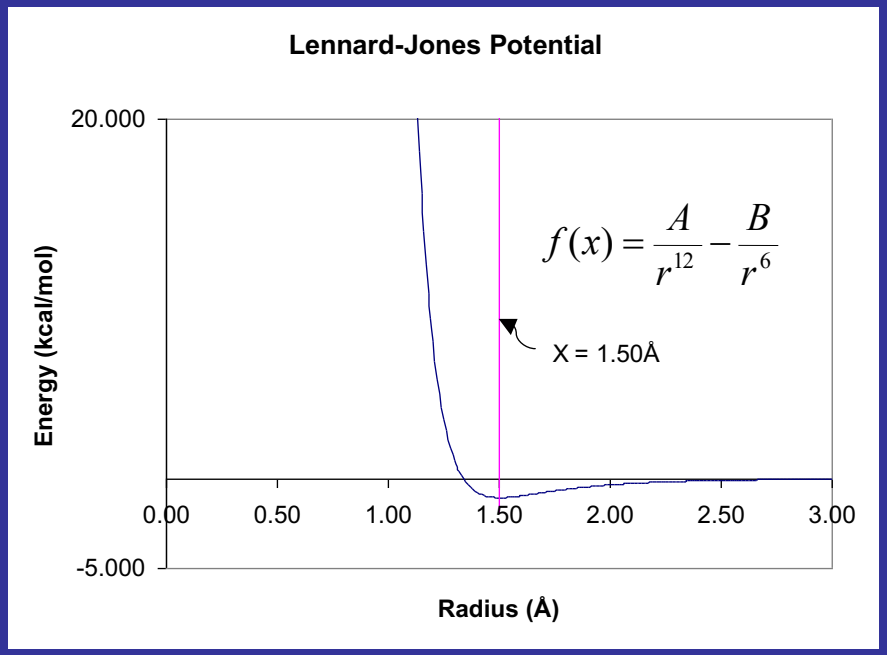
fast, grid-based DOCK3.7 scoring function
(a miracle it ever works)

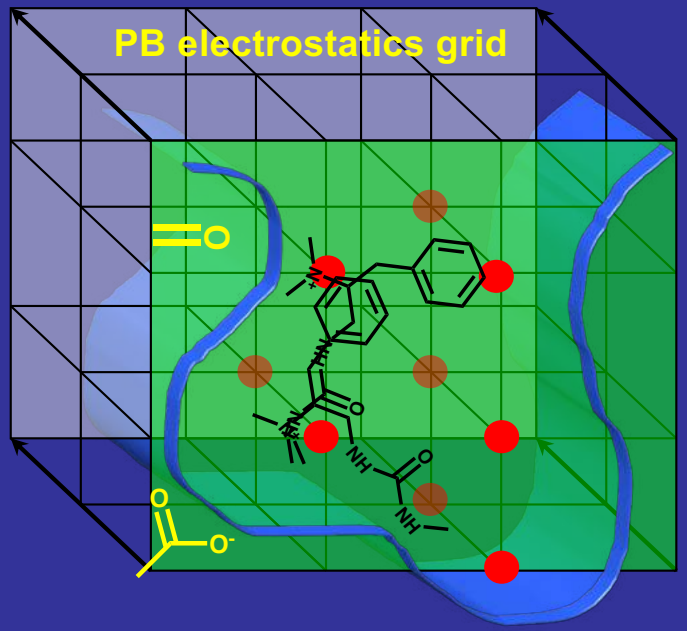
Mysinger, *JCIM* 2010
Fischer, *Nature Chem.* 2014
Balius, *PNAS*, 2017

$$\Delta G_{interact} - \Delta G_{solv,\,L} - \Delta G_{solv,\,R} = \Delta G_{bind}$$

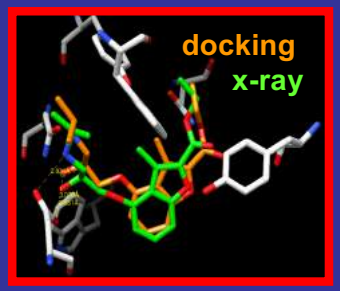
$$\Sigma(q_i{*}PB_i + v_i{*}P_{v,i}) = \Delta G_{inter}$$

$$GB_{ligand} + IST_{receptor} = \Delta G_{solv}$$

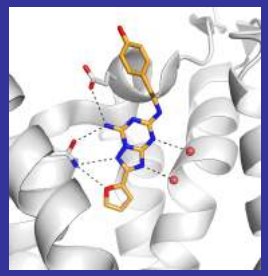




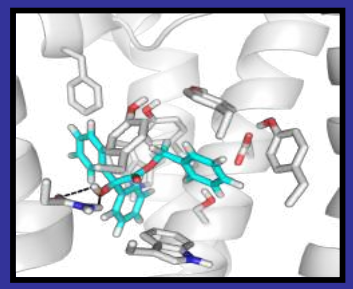
GPCR docking: 17 to 58% hit rates, nM activities, novel ligands, new biology, xray confirmation



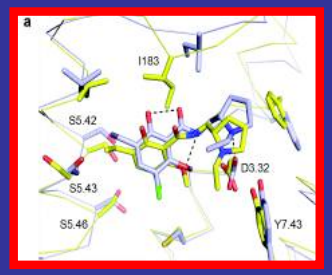
β2AR: 25%
0.01 to 3 uM
Kolb, *PNAS* 2009



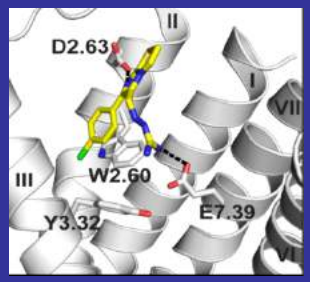
A2a: 35%
0.2 to 3 uM
Carlsson, *JMC* 2010



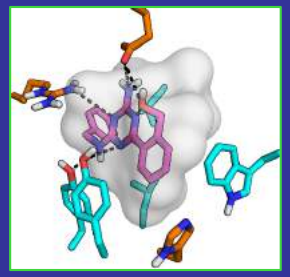
CHRM3: 58%
0.4 to 40uM
Kruze, *Mol Pharm* 2013



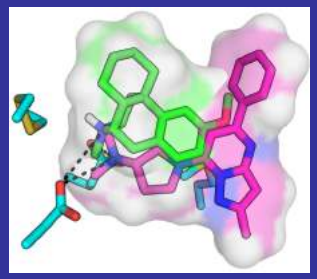
D3 23%
0.2 to 3uM
Carlsson, *NCB* 2011



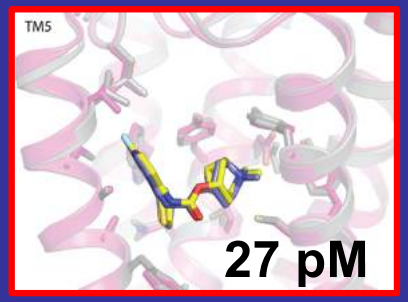
CXCR4:17%
0.3 to 30uM
Mysinger, *PNAS* 2012



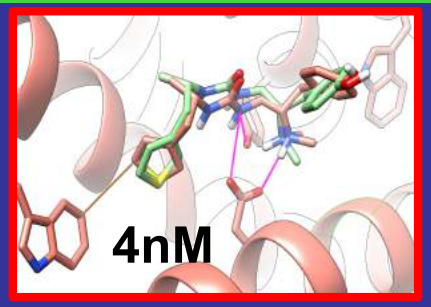
GPR68: 1uM PAM
Huang, *Nature* 2015



MRGPRX2. 6%
0.7 uM
Lansu *NCB* 2017



CHRM3
Liu, *PNAS* 2018

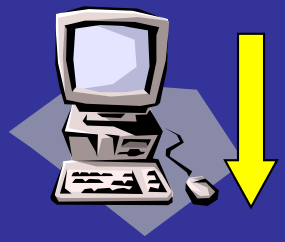


MOR: 30% 2 to 14 uM
Manglik *Nature* 2016
Kobilka, Skinniotis, unpub

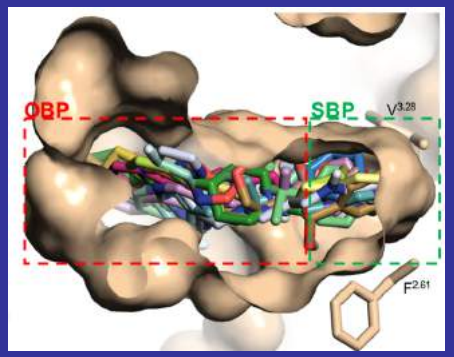
an explosion in readily-accessible molecules

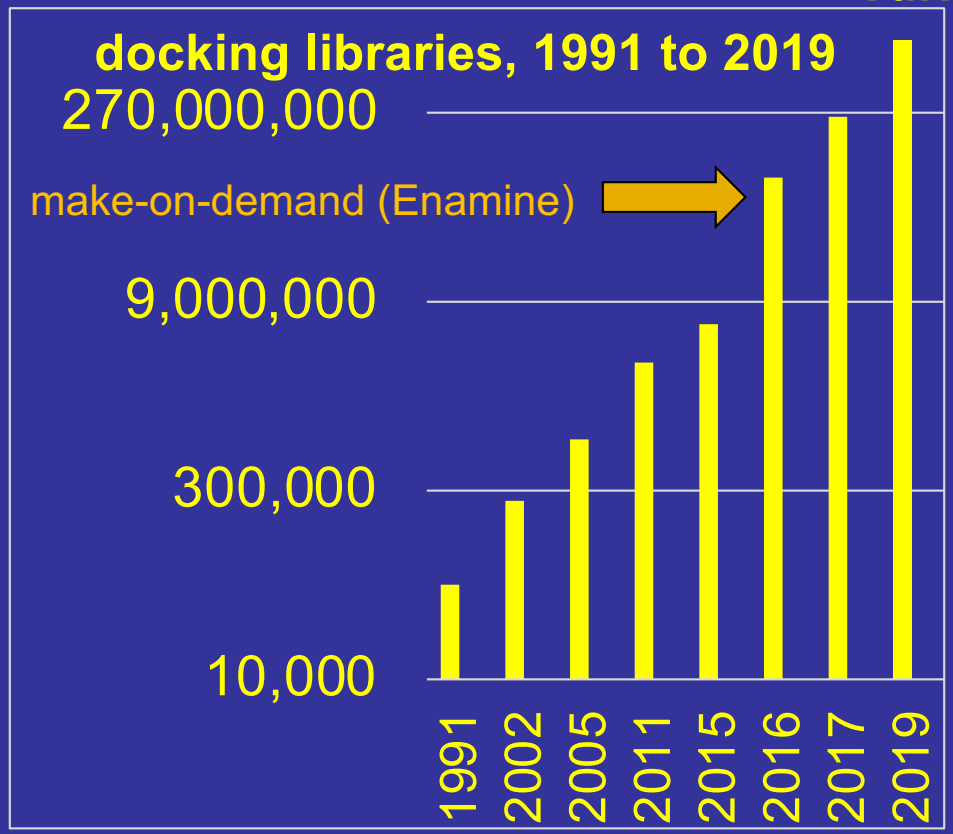
Yurii Moroz

3.5→350 million compounds

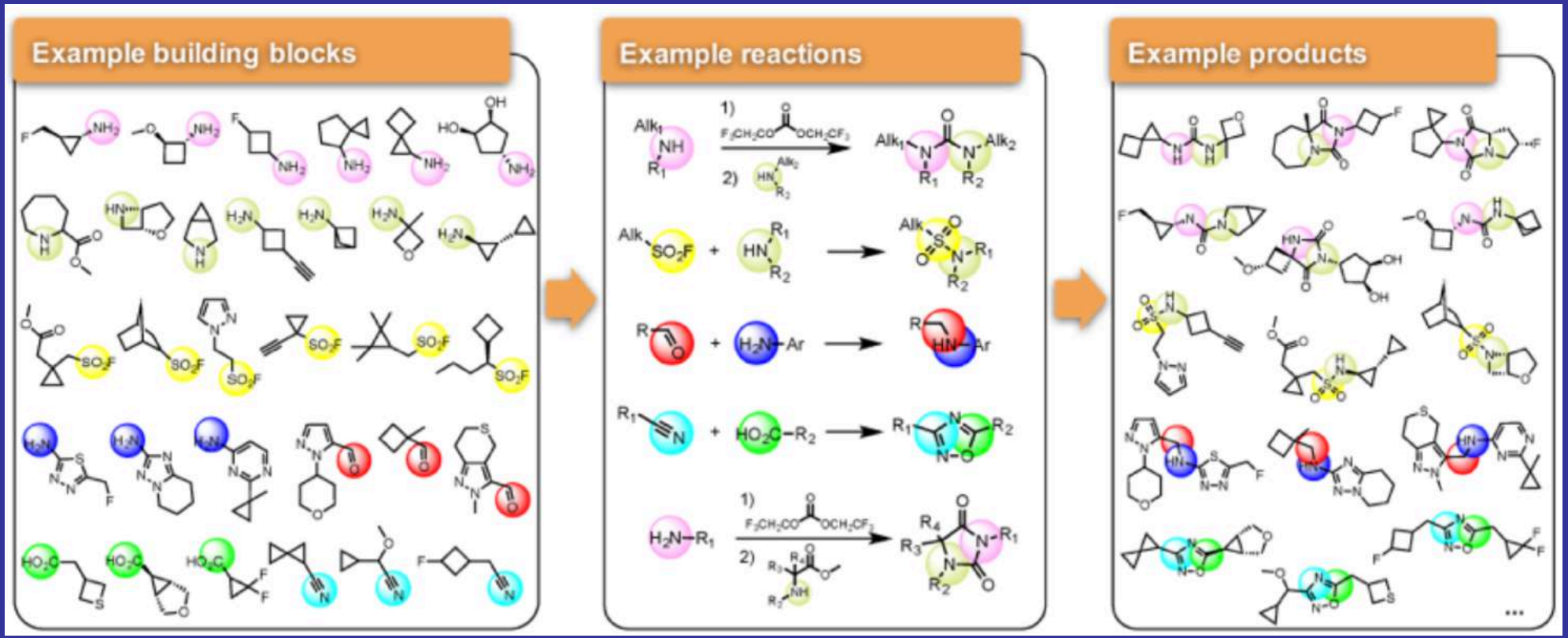


225 trillion complexes

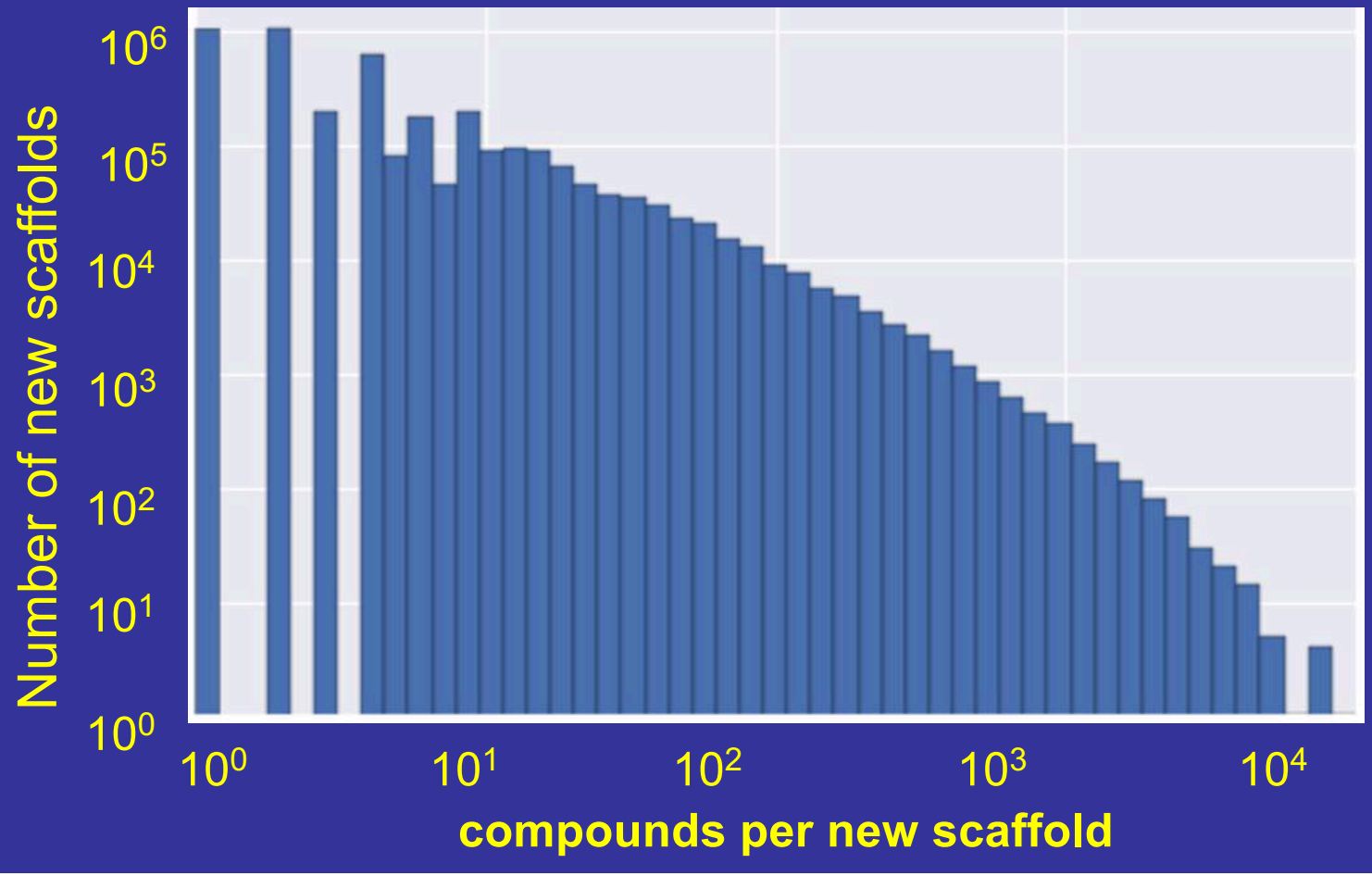


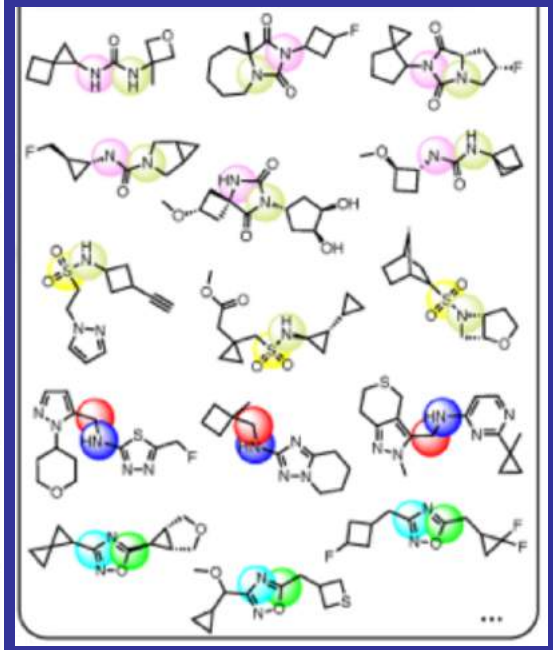


make-on-demand library:
109 2-component reactions
> 200,000 building blocks



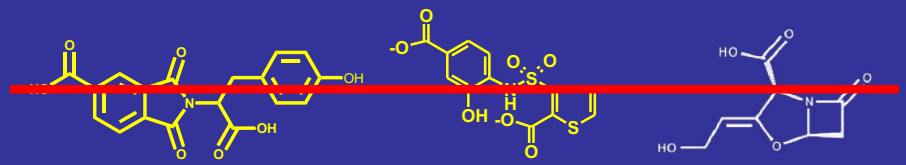
library is diverse: 7 million new scaffolds in 138 million compounds





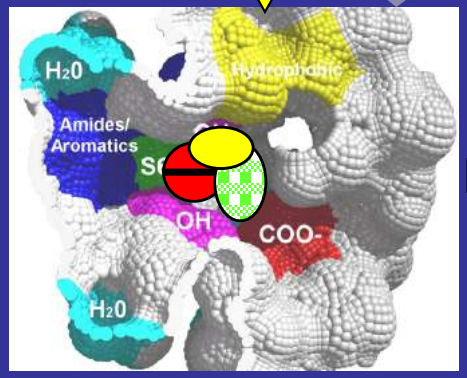
How most screening compounds get made

docking 99 million compounds vs β-lactamase predicts new chemotypes

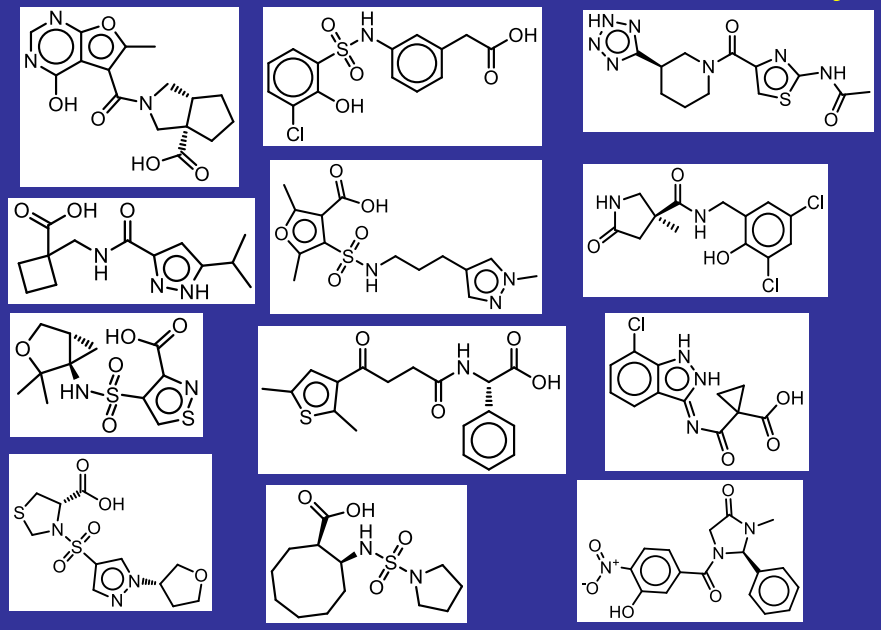


cannot resemble knowns nor 3.5 million library

99 million cmpds



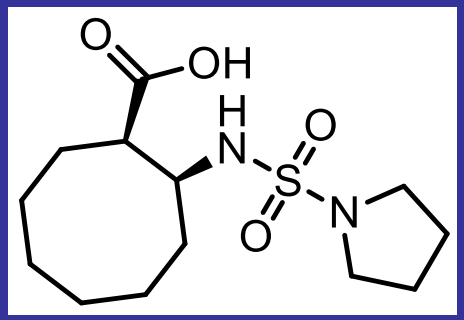
28,310 core hours

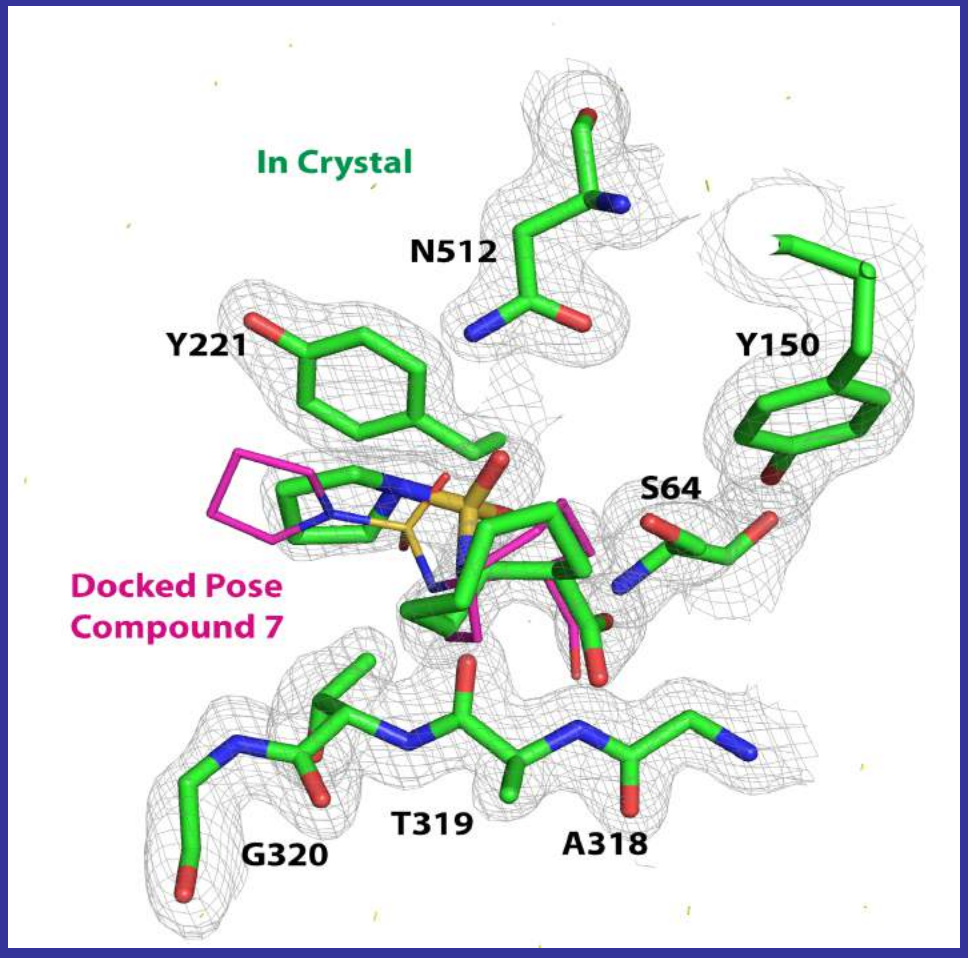


44 WACs picked from top 3000 clusters

JK Lyu, Trent Balius, Isha Singh, *Nature* 2019

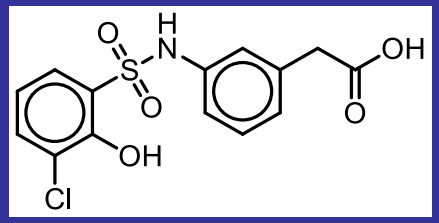
99 million vs β-lactamase: 11% hit rate,
x-ray structure confirms a docking prediction

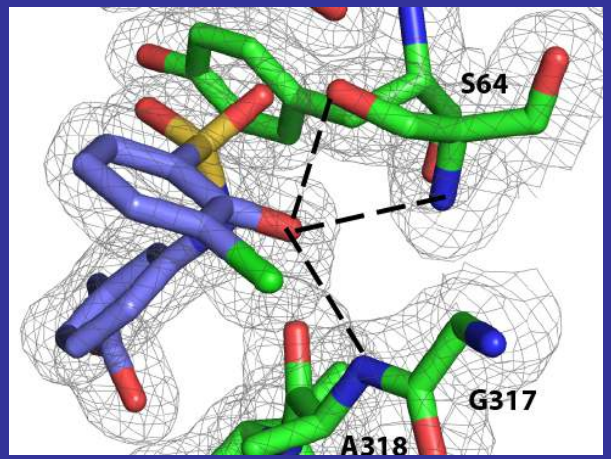


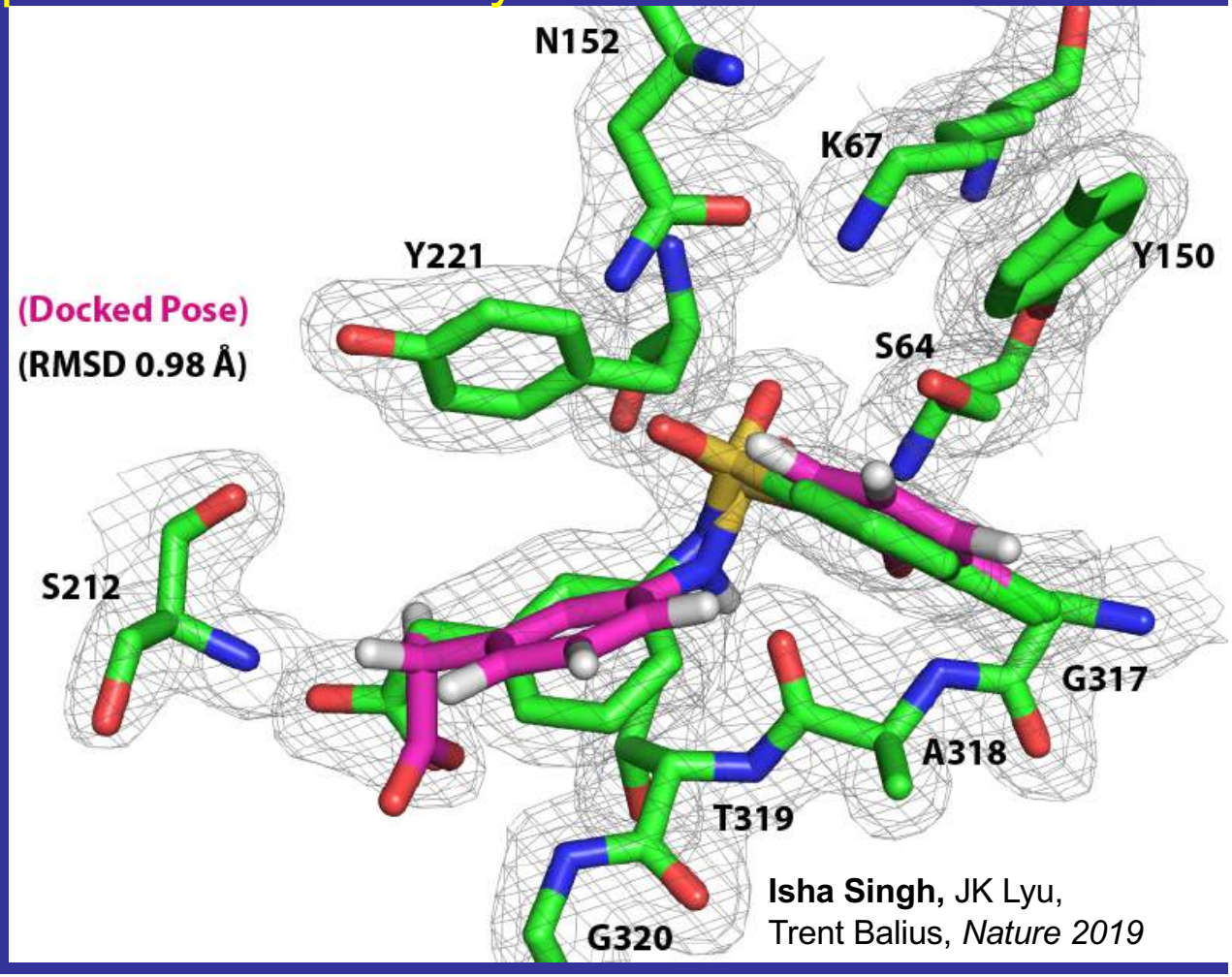


Isha Singh, JK Lyu,
Trent Balius, *Nature* 2019

99 million vs β-lactamase: 11% hit rate, unprecedented
1.3 uM phenolate in the oxyanion hole

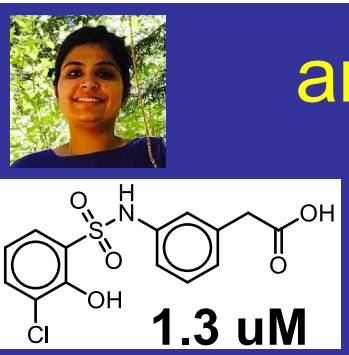






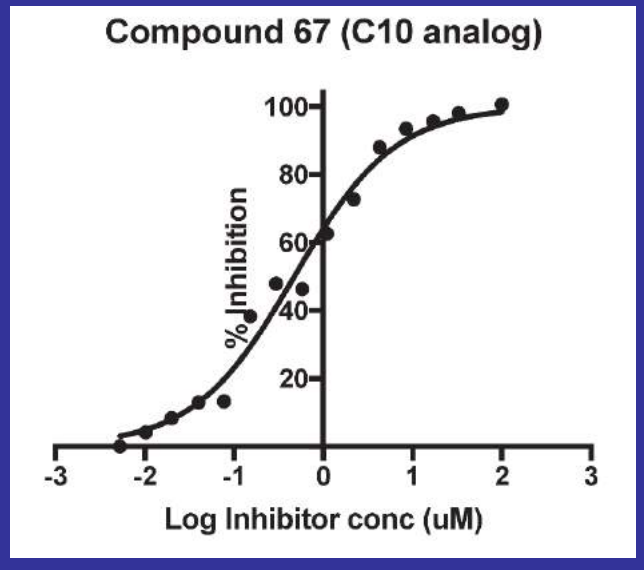
Isha Singh, JK Lyu, Trent Balius, *Nature 2019*

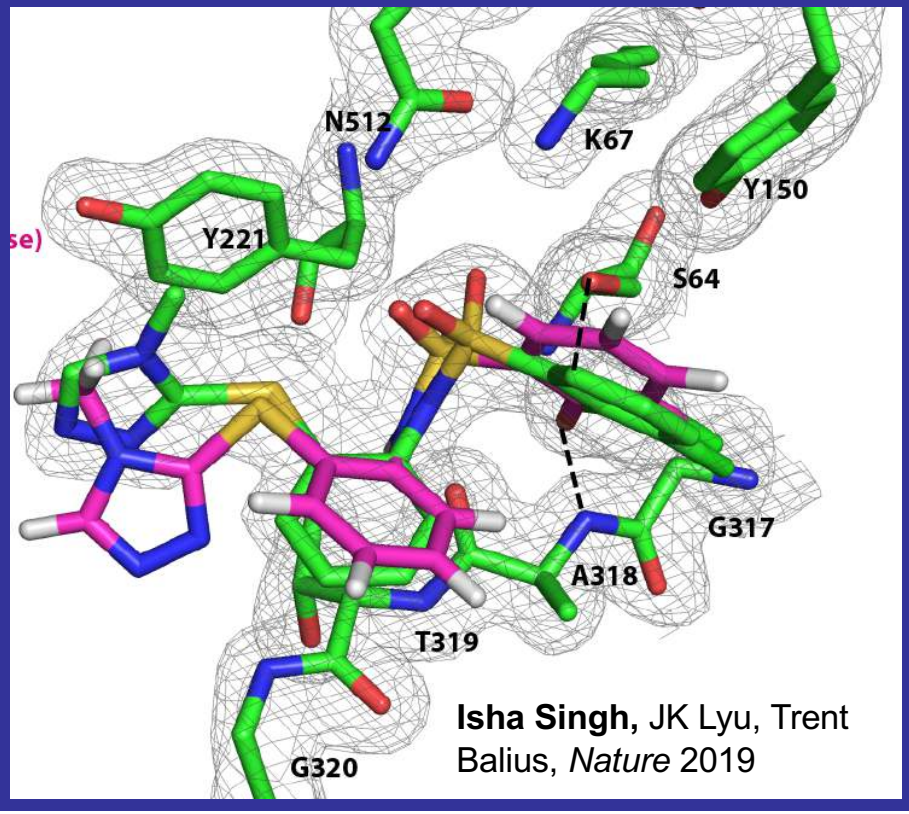
analog-by-catalog to 77 nM



77 nM

>400 analogs docked, test 11

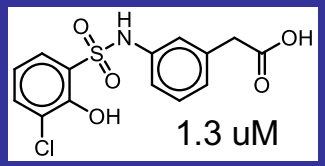




Isha Singh, JK Lyu, Trent Balius, *Nature* 2019

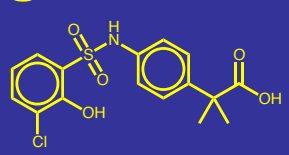
analog-by-catalog optimization:
>2000 possible analogs, which are best?

Isha Singh



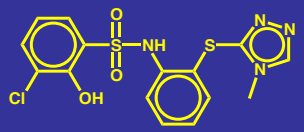
test 11

15 uM

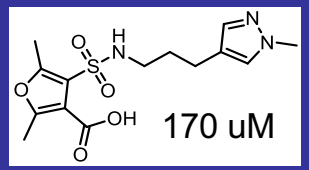


4 uM

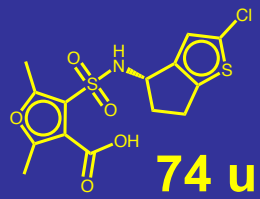
. . .



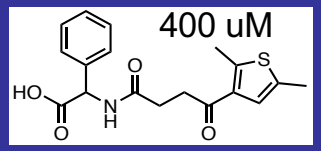
77 nM



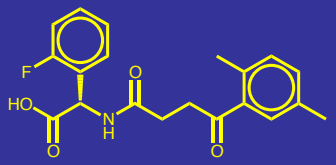
test 15



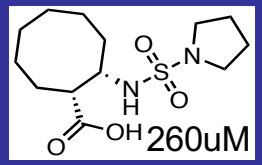
74 uM



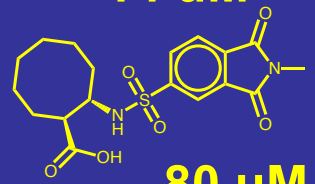
test 31



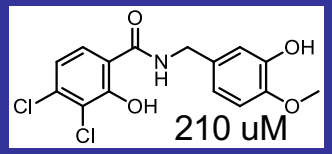
14 uM



test 19



80 uM

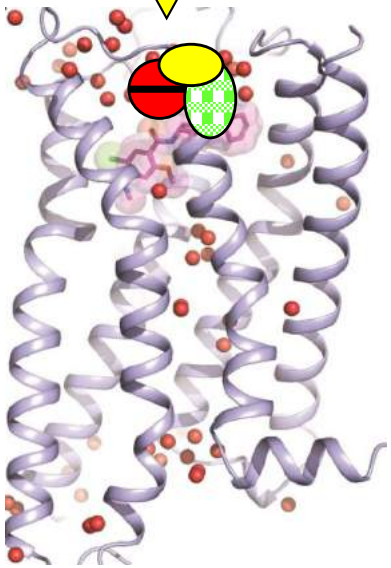


test 14

100 uM

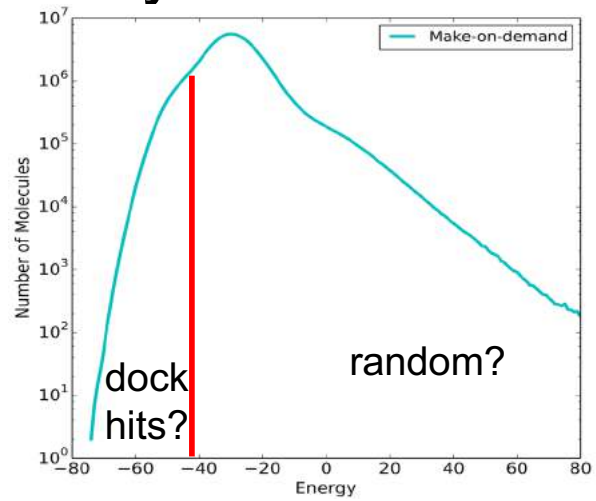
How many ligands *findable* in the library?
138 million vs. dopamine D4

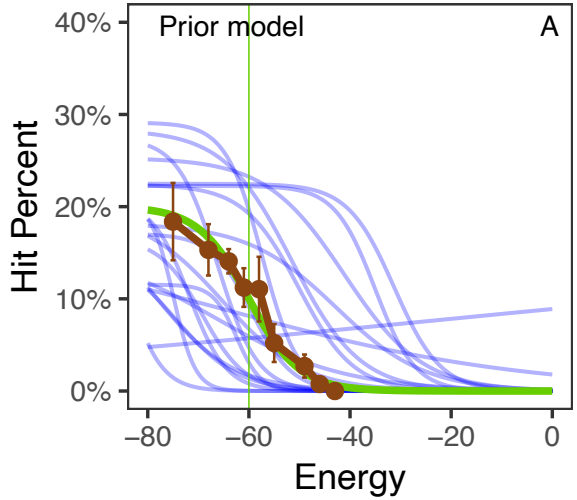
138 million molecules
10^{14} complexes
1 mol/sec;10^3 cores
1.8 cluster days



~~Test 30~~

Test 549





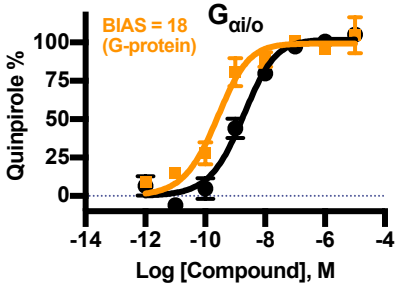
JK Liu, A Levit, T Balius, M. O'Meara, *Nature* 2019

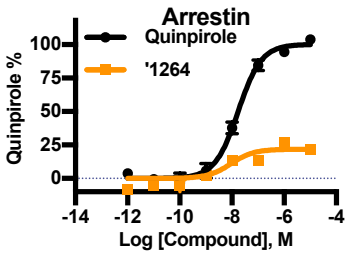
from top ranked: novel, selective, potent D4 ligands

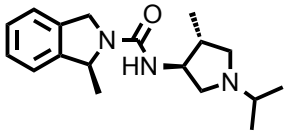
Sheng Wang
Bryan Roth

ZINC361131264, 18x bias

Rank: 1067

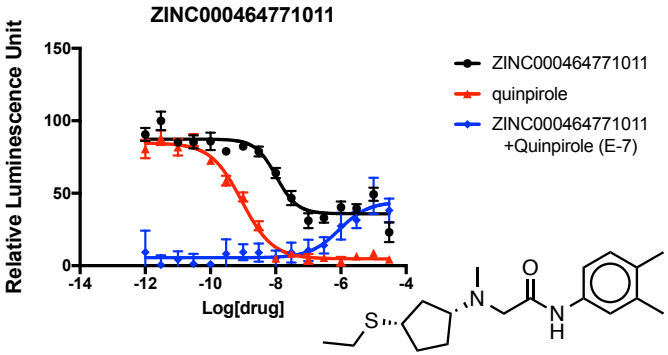




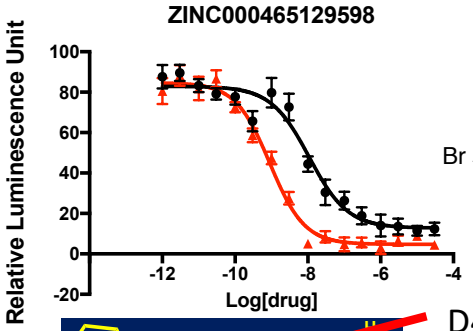


D4 = 0.18 nM
D2 > 10,000 nM
D3 > 10,000 nM

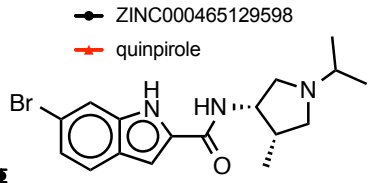
Rank: 936



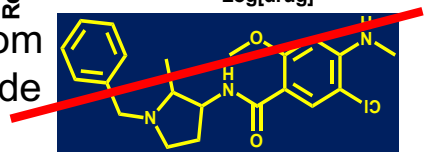
D4 = 10 nM
D2 > 10,000 nM
D3 > 10,000 nM



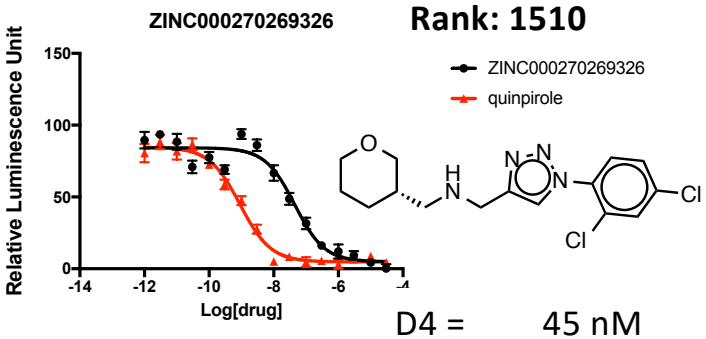
Rank: 1565



Different from Nemonapride



D4 = 11 nM
D2 > 10,000 nM
D3 > 10,000 nM

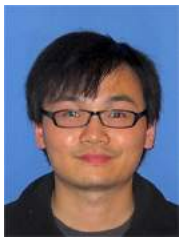


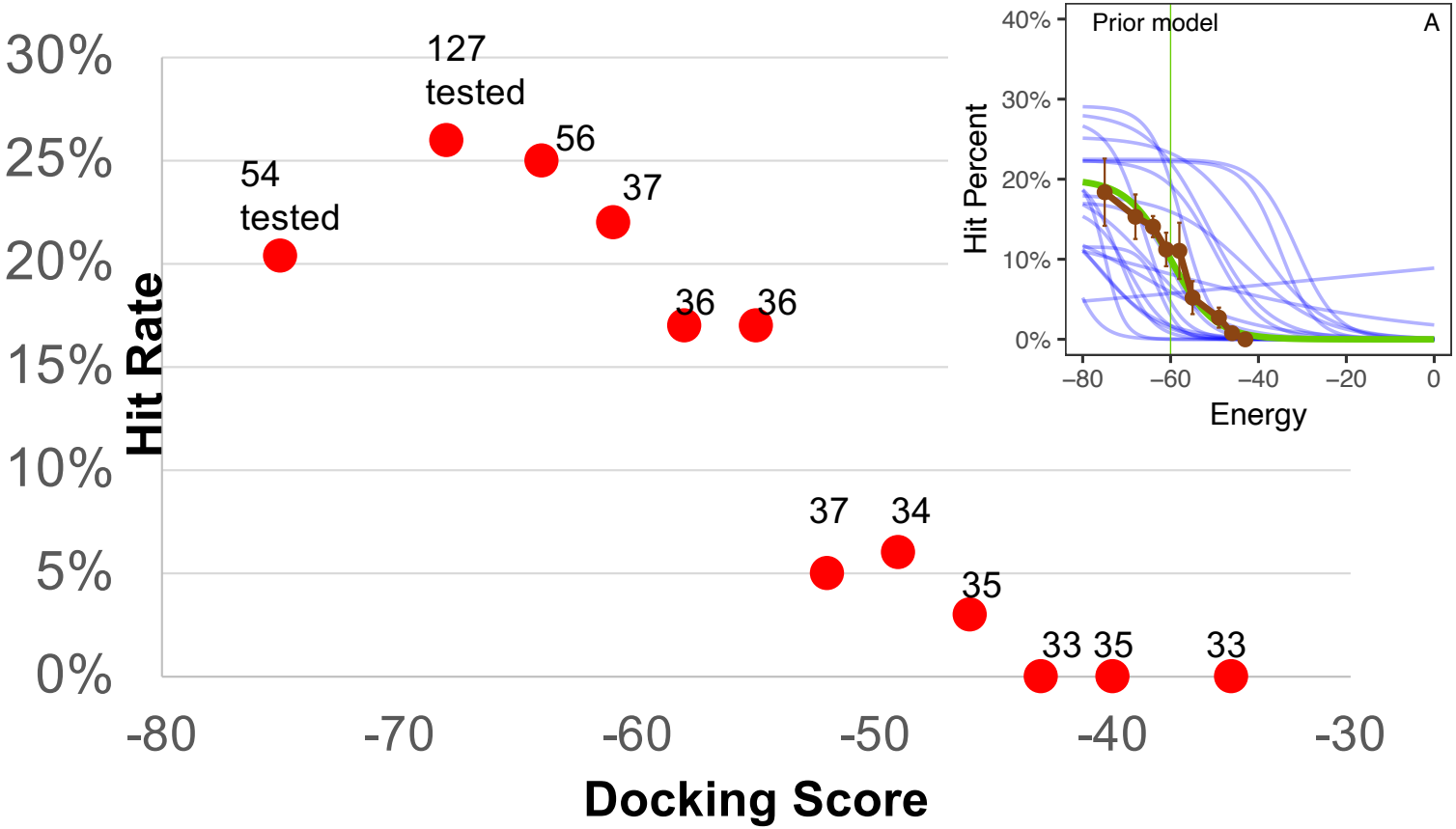
Rank: 1510

	ZINC000270269326
IC50	4.496e-008

D4 = 45 nM
D2 > 10,000 nM
D3 > 10,000 nM

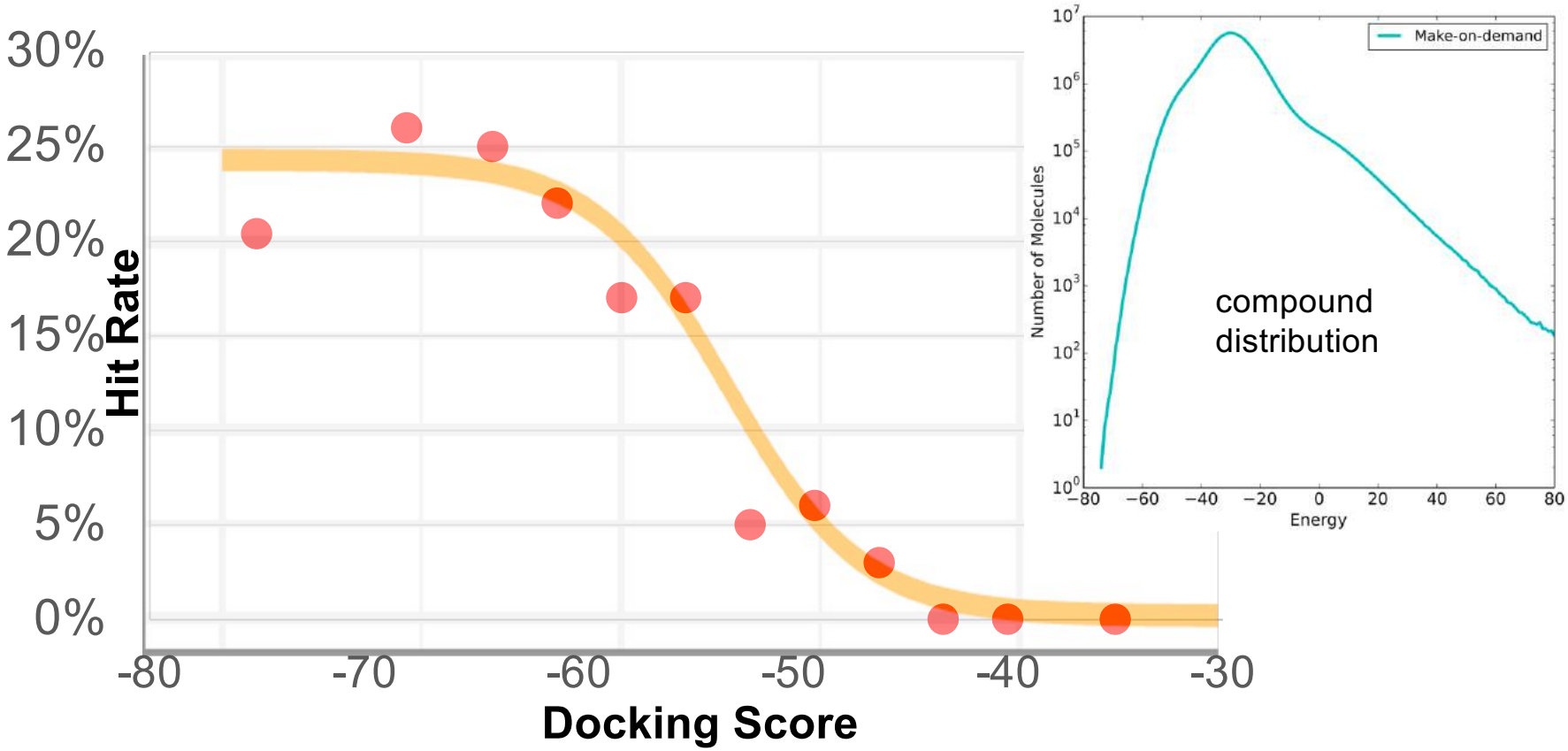
hit rates correlate with docking rank





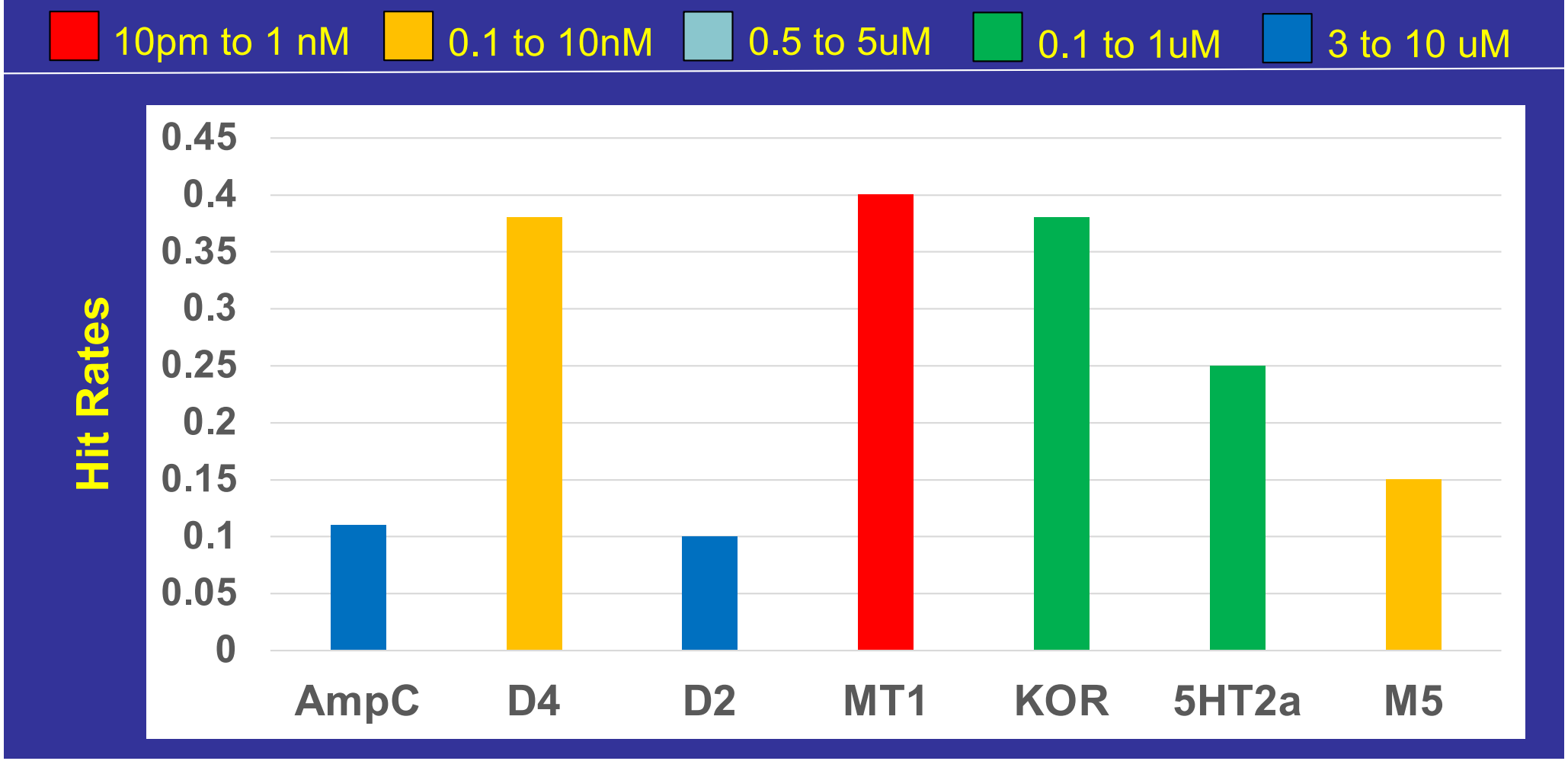
Sheng Wang, Bryan Roth, *Nature* 2019

AUC predicts 479,000 (±40%) DRD4 actives (10 uM) in 72,000 (±30%) scaffold families



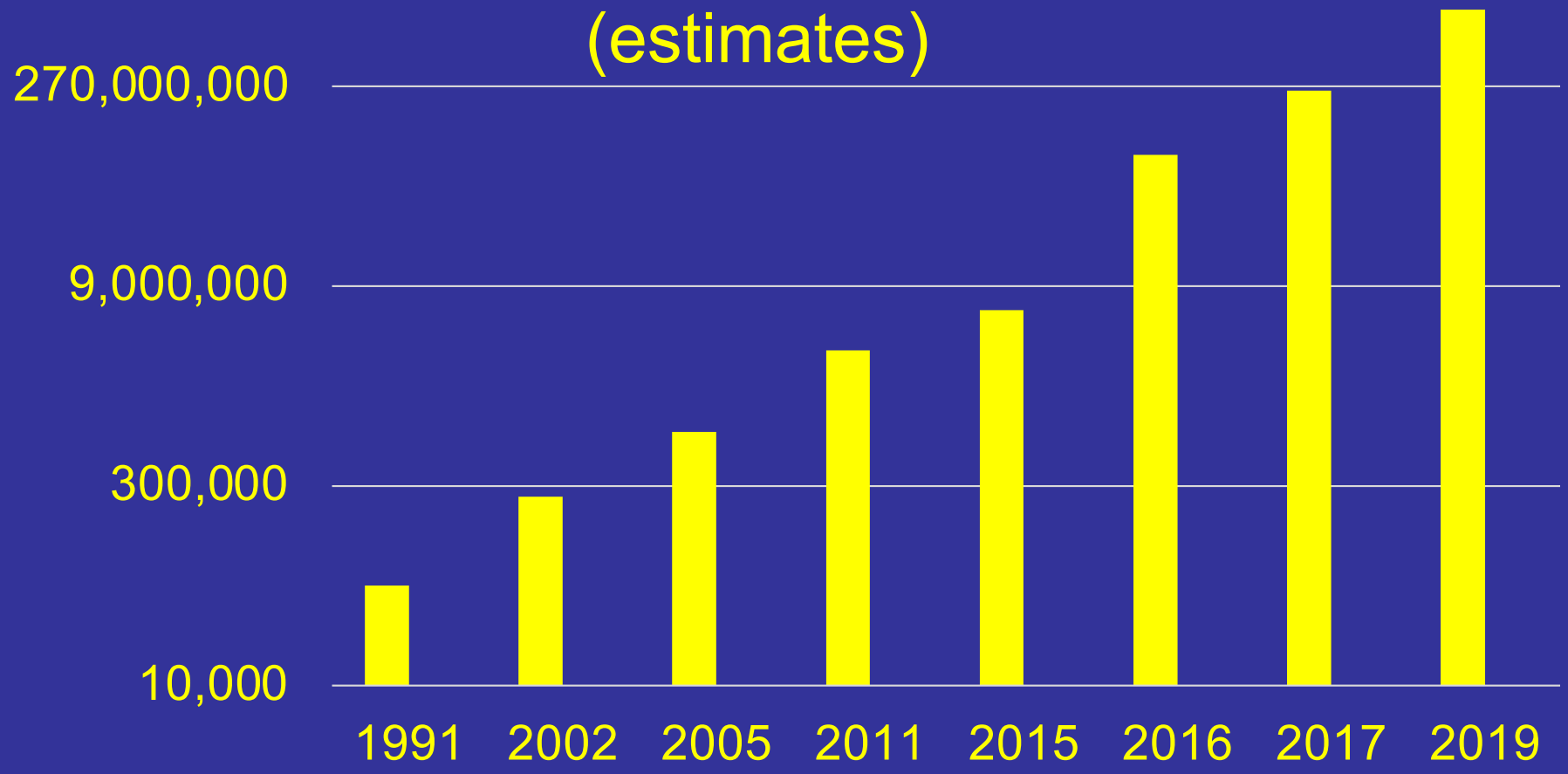
Matt O'Meara, *Nature* 2019

high hit rates in LSD campaigns vs over a dozen targets



Dockable libraries keep growing

End 2020: 1 billion, 2021: 4 billion, 2022 10 billion (estimates)



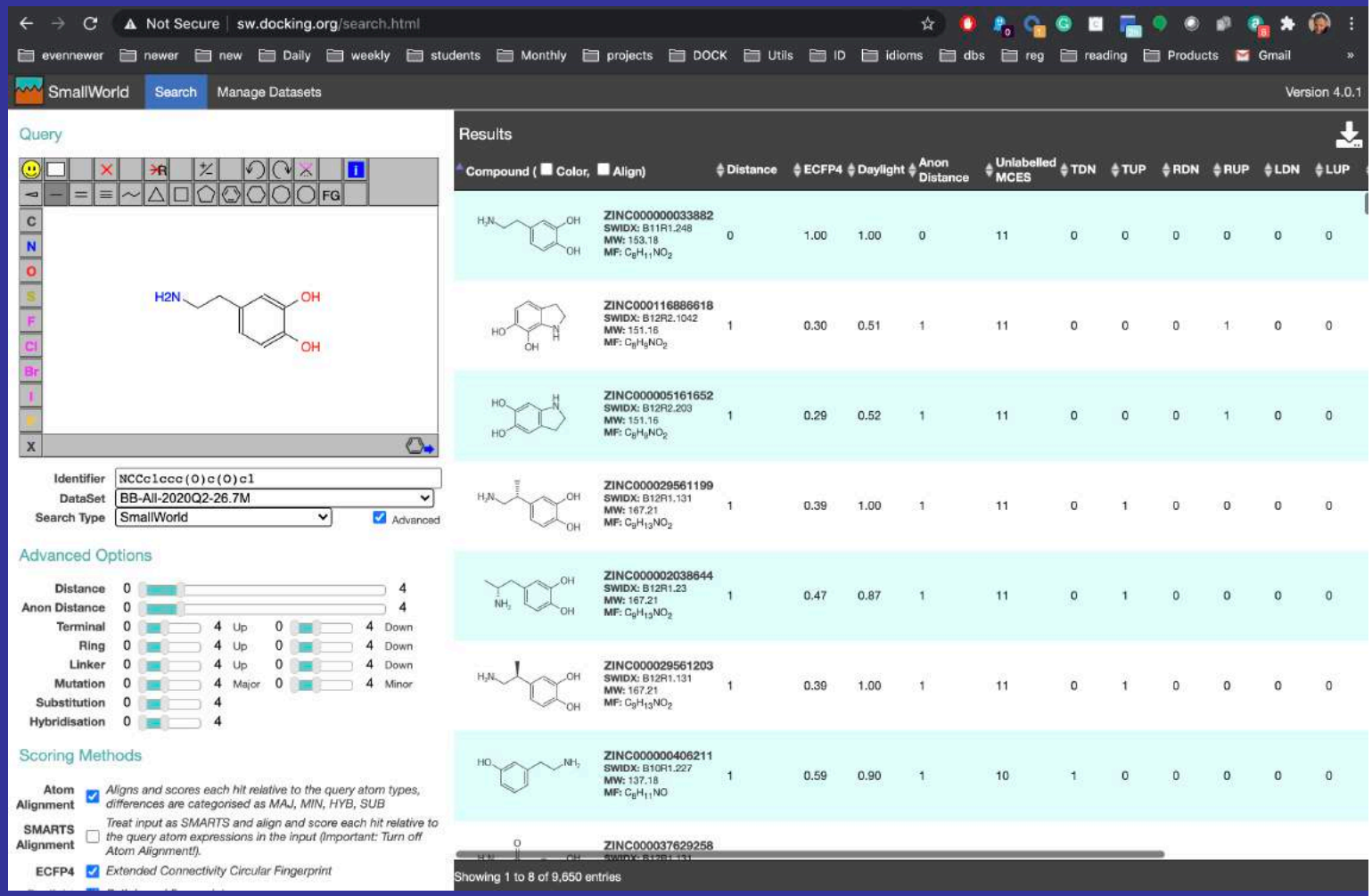
Free Public Resources and how to use them

Smallworld

for

similarity

sw.docking.org



NextMove Software

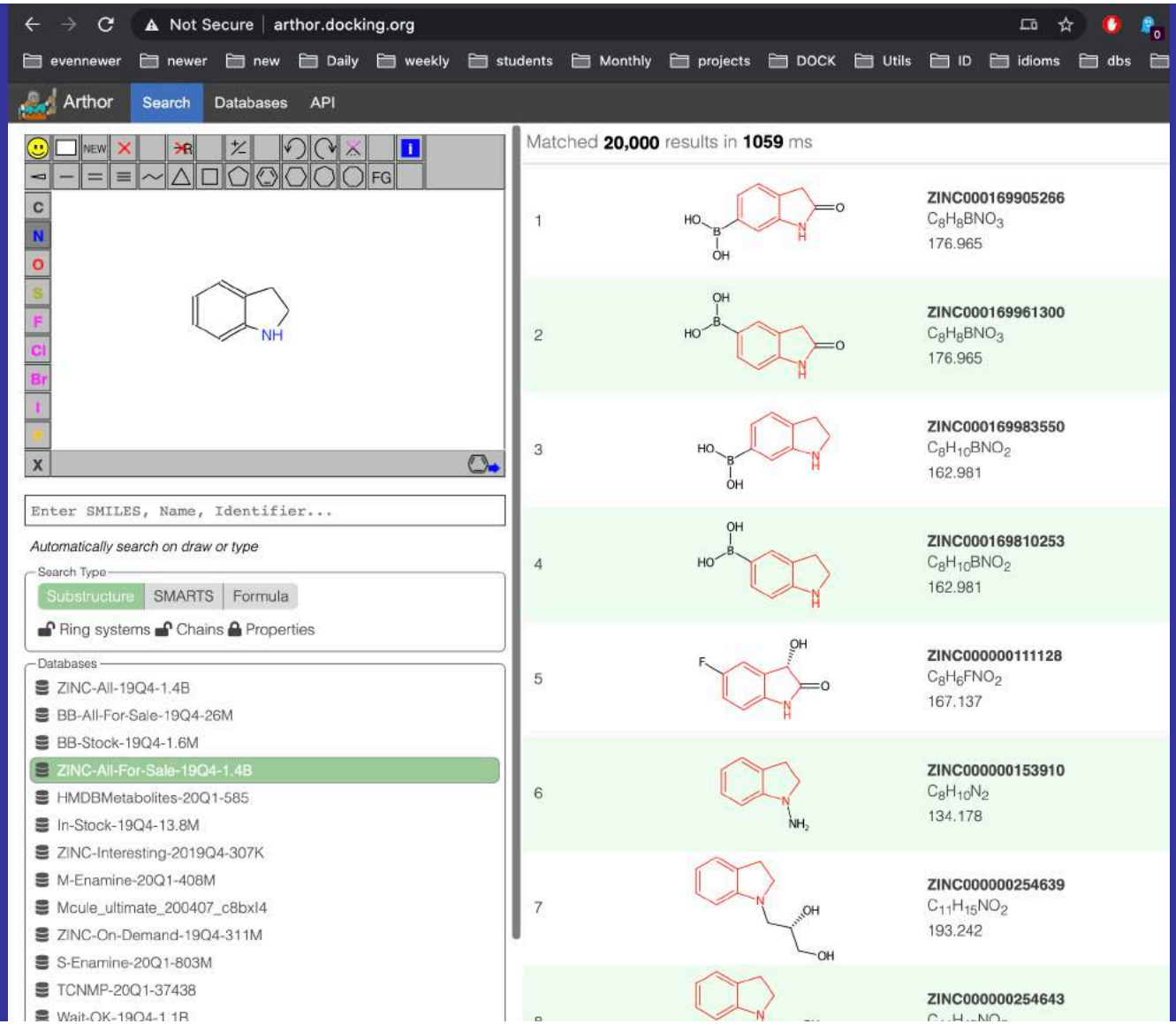
Roger A. Sayle

Arthor
for substructure

arthor.docking.org

NextMove Software

Roger A. Sayle

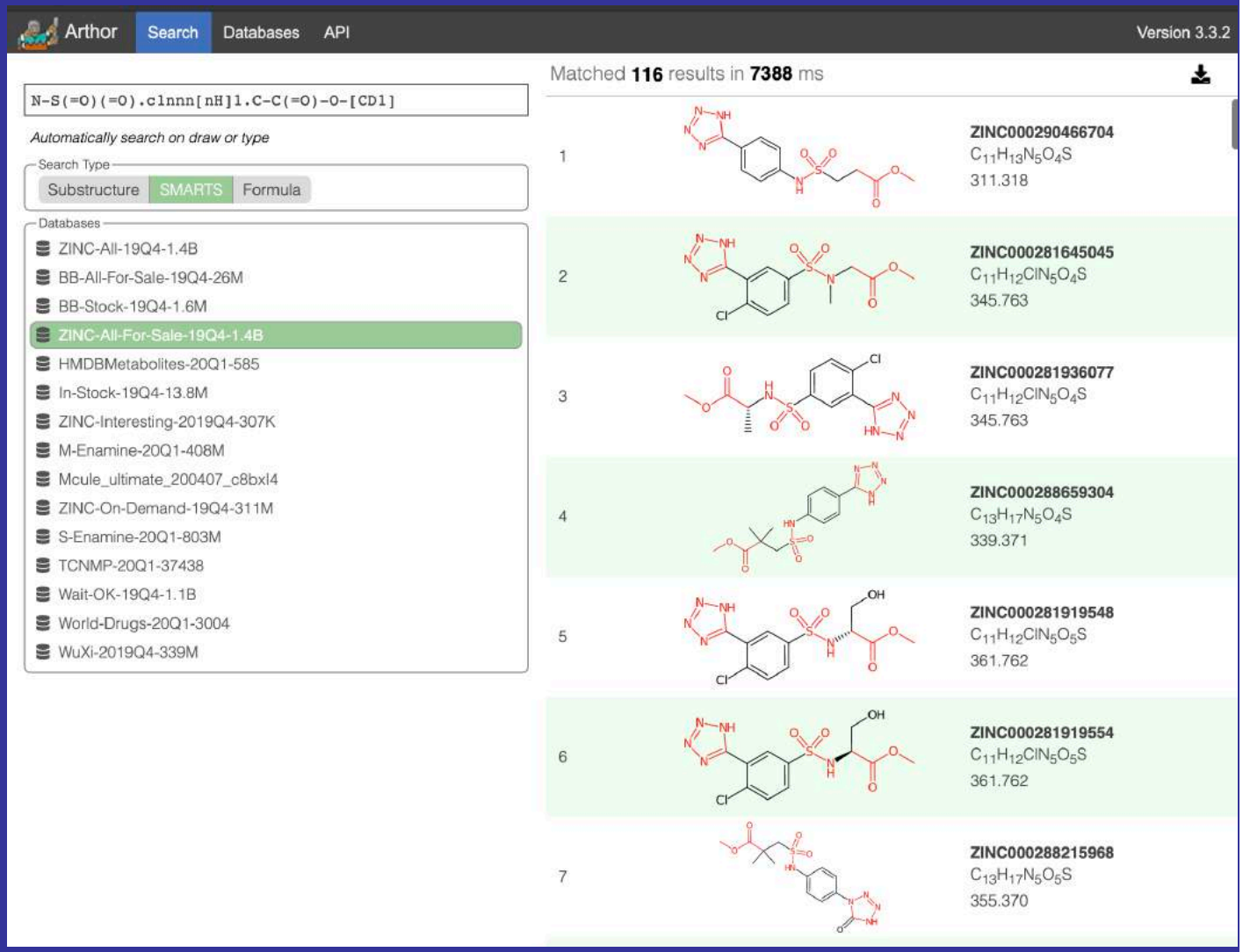


Arthor for patterns

arthor.docking.org

NextMove Software

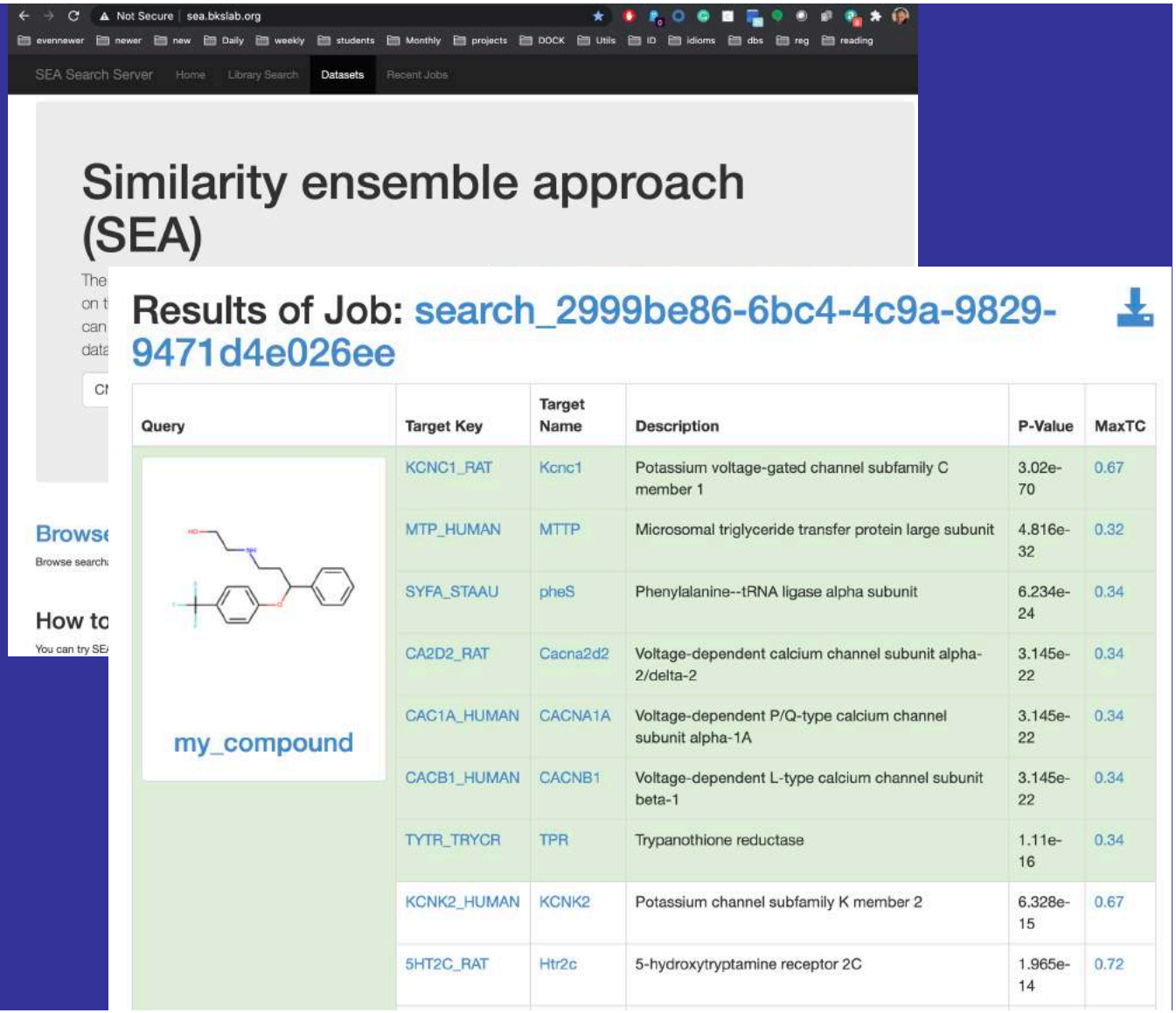
Roger A. Sayle



SEA: Similarity Ensemble Approach

for drug targets of small molecules

sea.bkslab.org



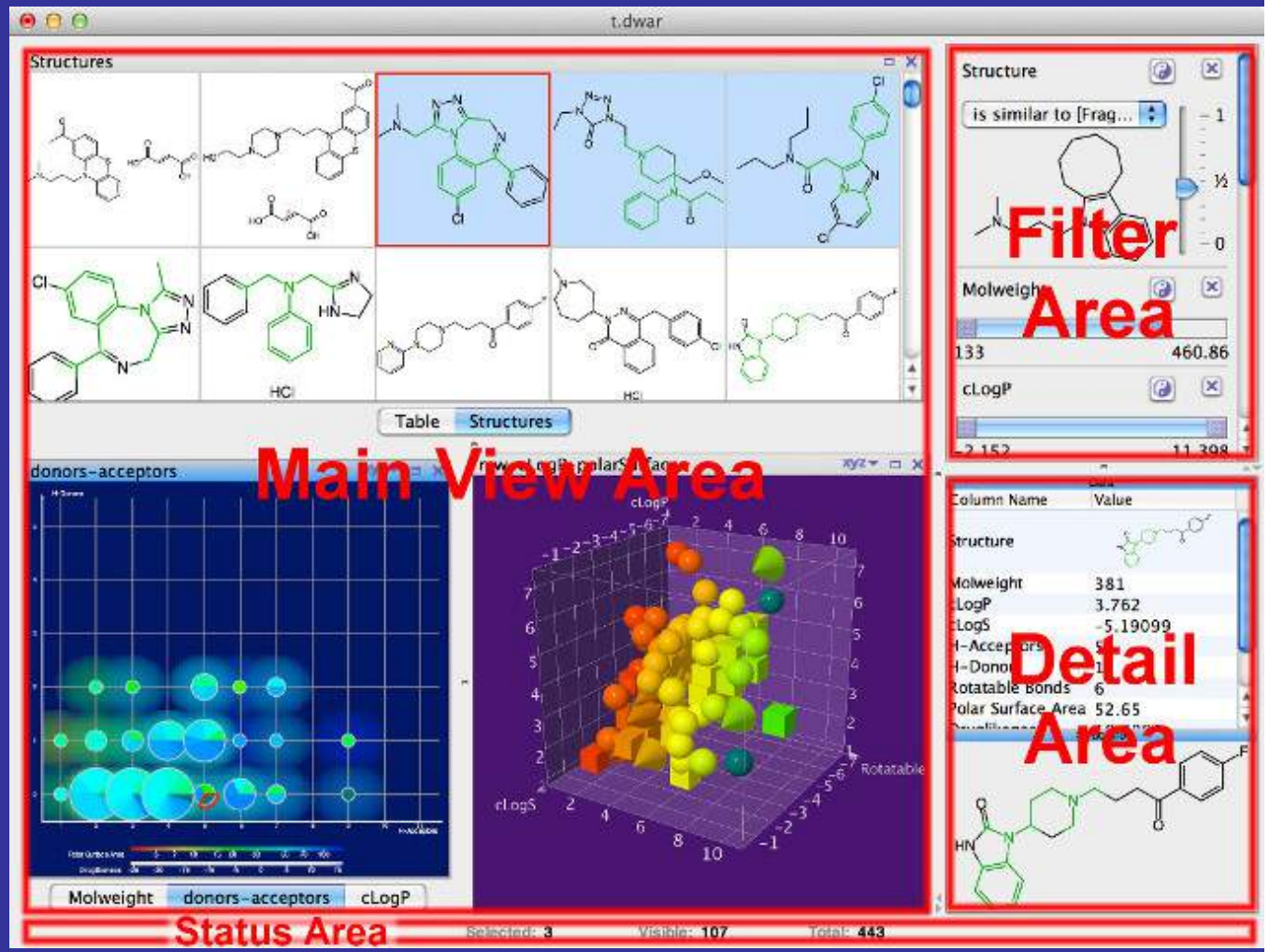
Query	Target Key	Target Name	Description	P-Value	MaxTC
my_compound	KCNC1_RAT	Kcnc1	Potassium voltage-gated channel subfamily C member 1	3.02e-70	0.67
	MTP_HUMAN	MTTP	Microsomal triglyceride transfer protein large subunit	4.816e-32	0.32
	SYFA_STAAU	pheS	Phenylalanine--tRNA ligase alpha subunit	6.234e-24	0.34
	CA2D2_RAT	Cacna2d2	Voltage-dependent calcium channel subunit alpha-2/delta-2	3.145e-22	0.34
	CAC1A_HUMAN	CACNA1A	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	3.145e-22	0.34
	CACB1_HUMAN	CACNB1	Voltage-dependent L-type calcium channel subunit beta-1	3.145e-22	0.34
	TYTR_TRYCR	TPR	Trypanothione reductase	1.11e-16	0.34
	KCNK2_HUMAN	KCNK2	Potassium channel subfamily K member 2	6.328e-15	0.67
	5HT2C_RAT	Htr2c	5-hydroxytryptamine receptor 2C	1.965e-14	0.72

Datawarrior

free from

openmolecules.org

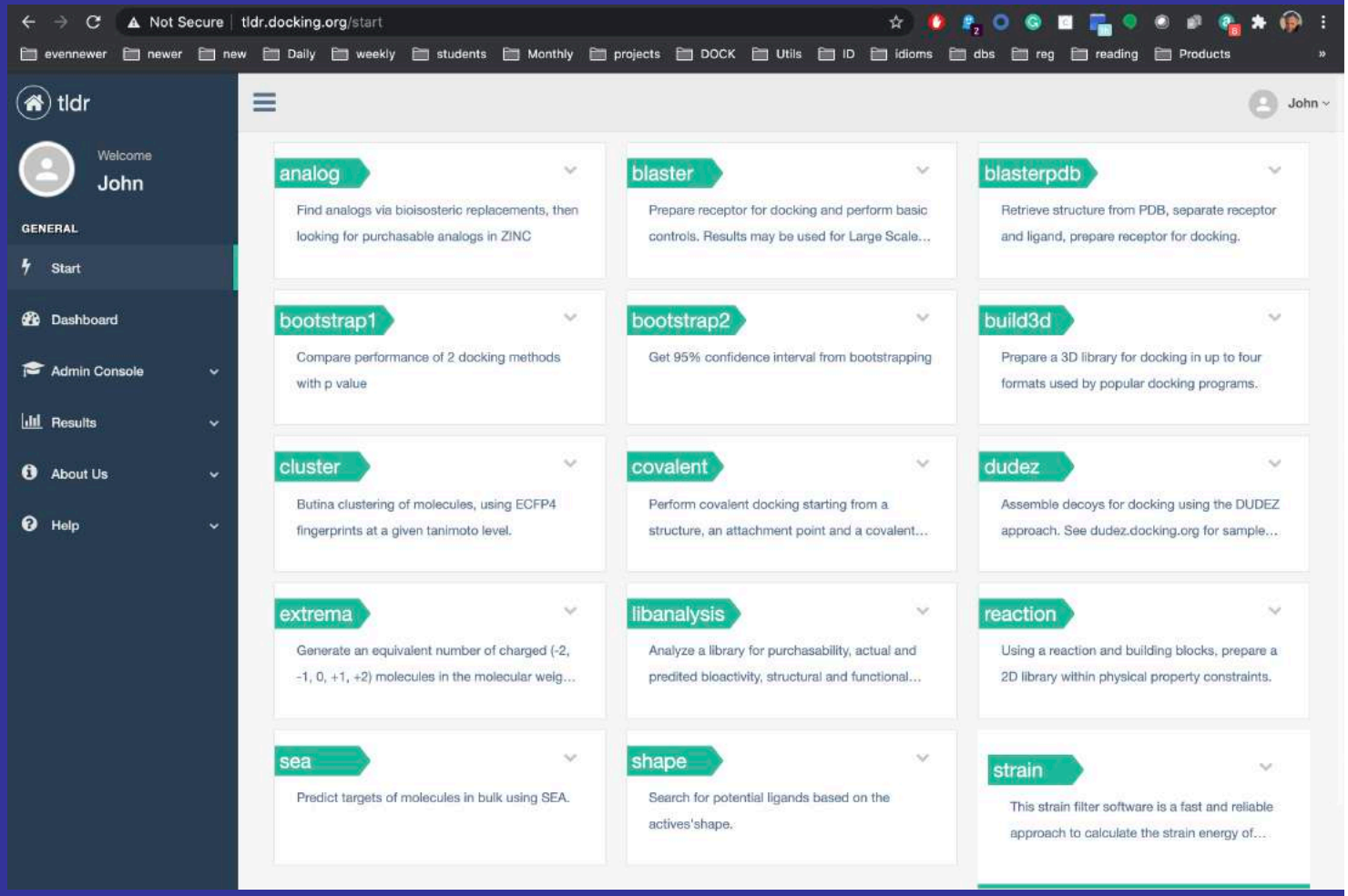
Thomas Sander
of idorsia.com



TLDR's a Ligand Discovery Resource

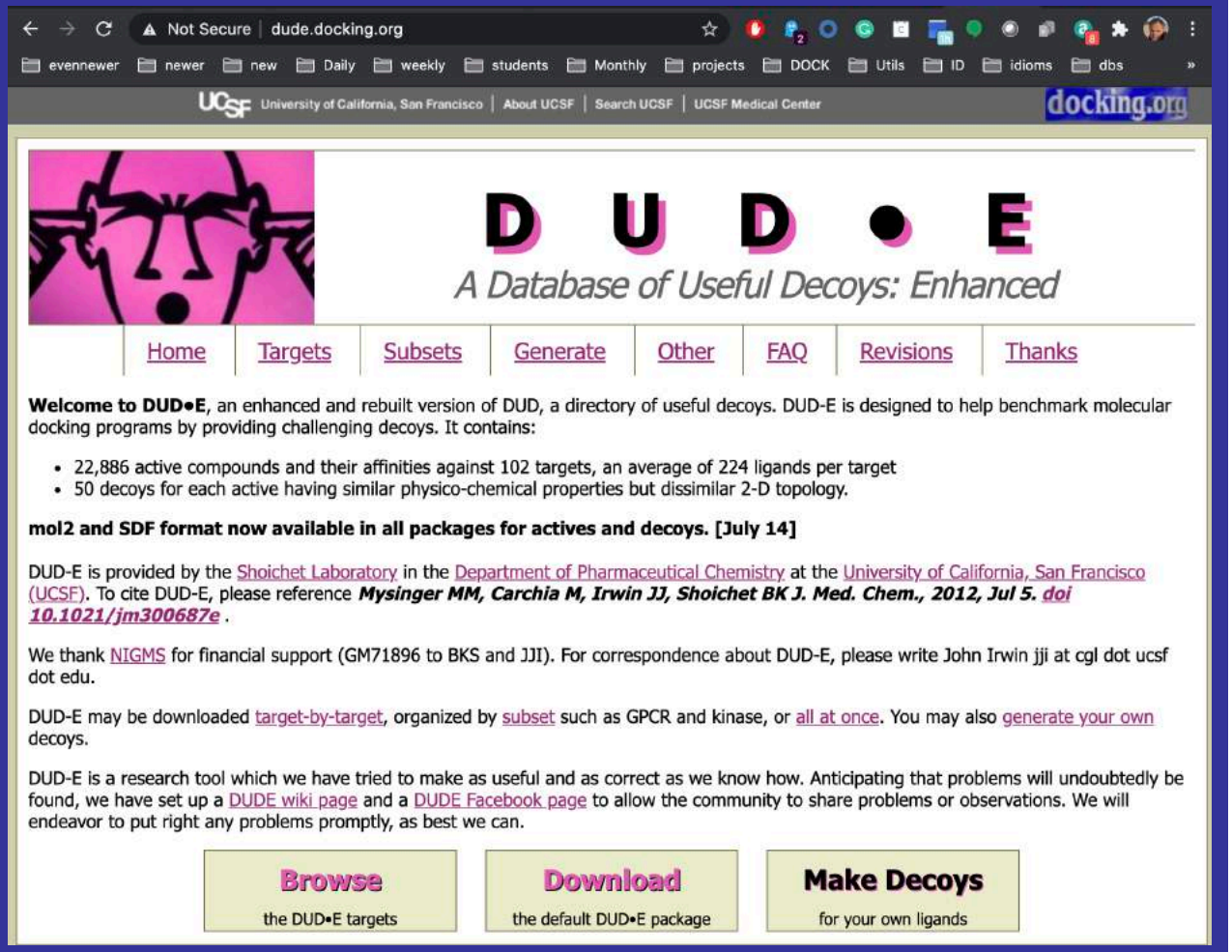
for free web-based tools

tldr.docking.org



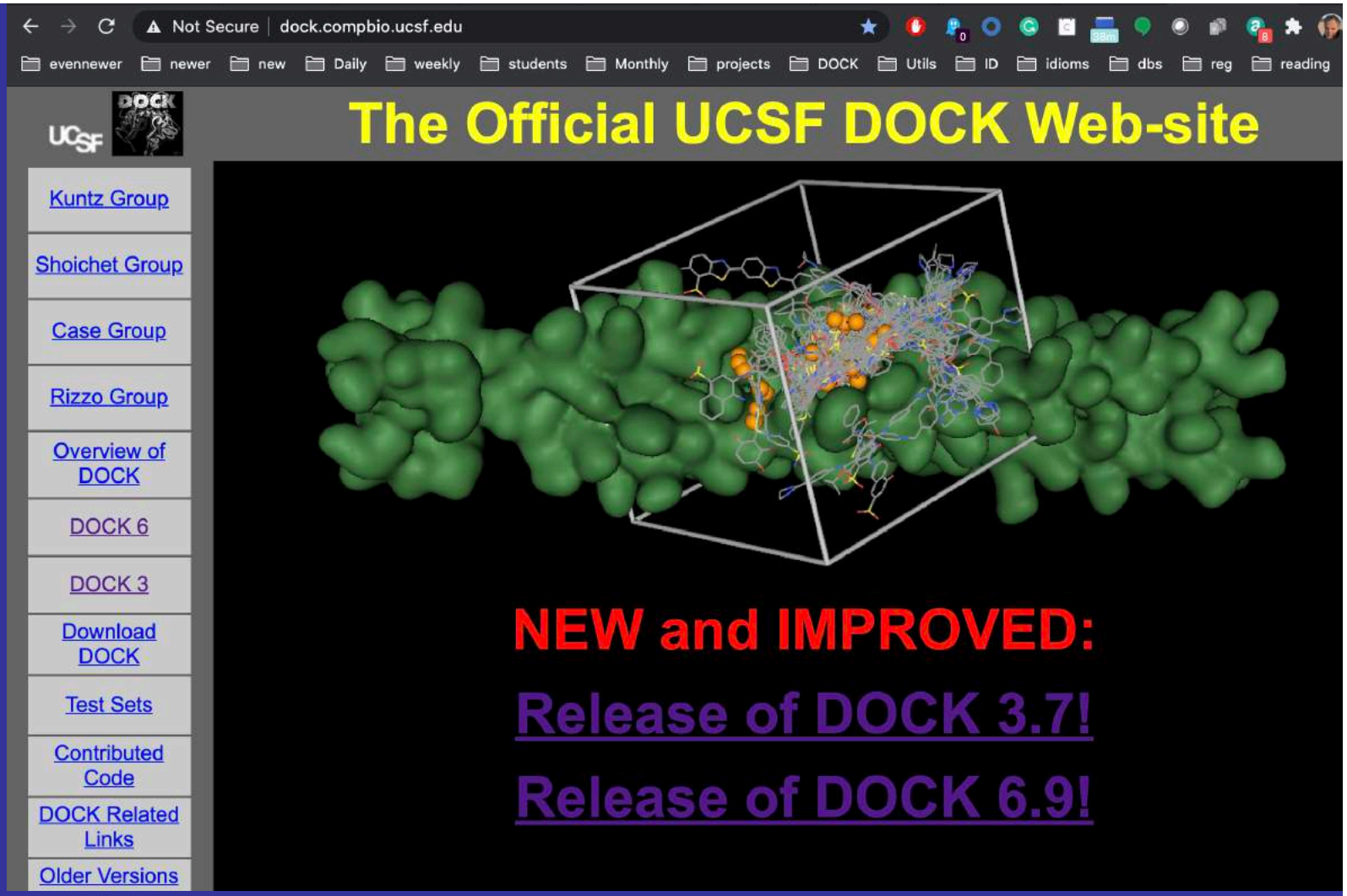
DUD
DUD-E
DUD-EZ

for
docking
benchmarks



UCSF DOCK

molecular docking software

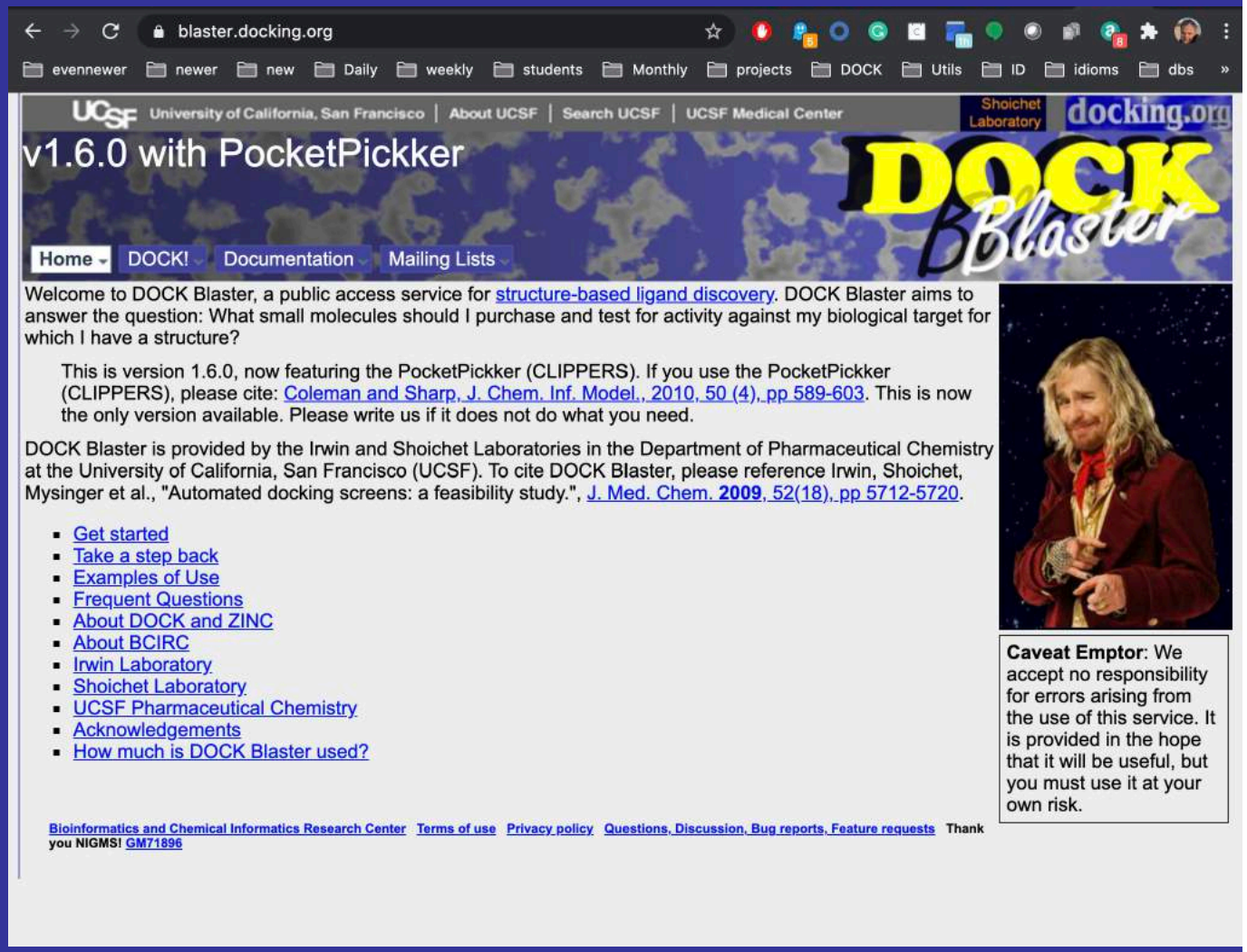


dock.compbio.ucsf.edu

DOCK Blaster

web interface to docking

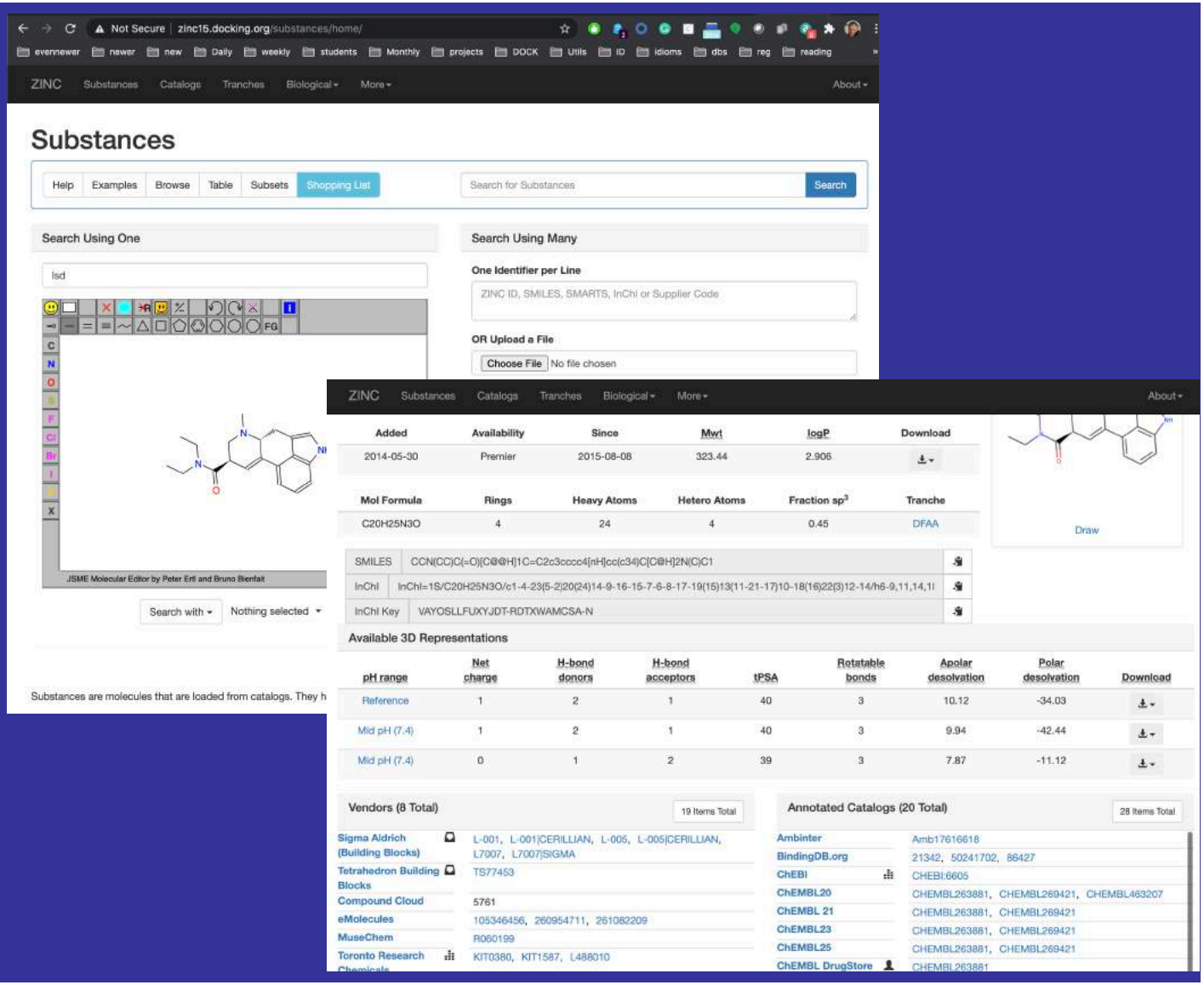
blaster.docking.org



ZINC

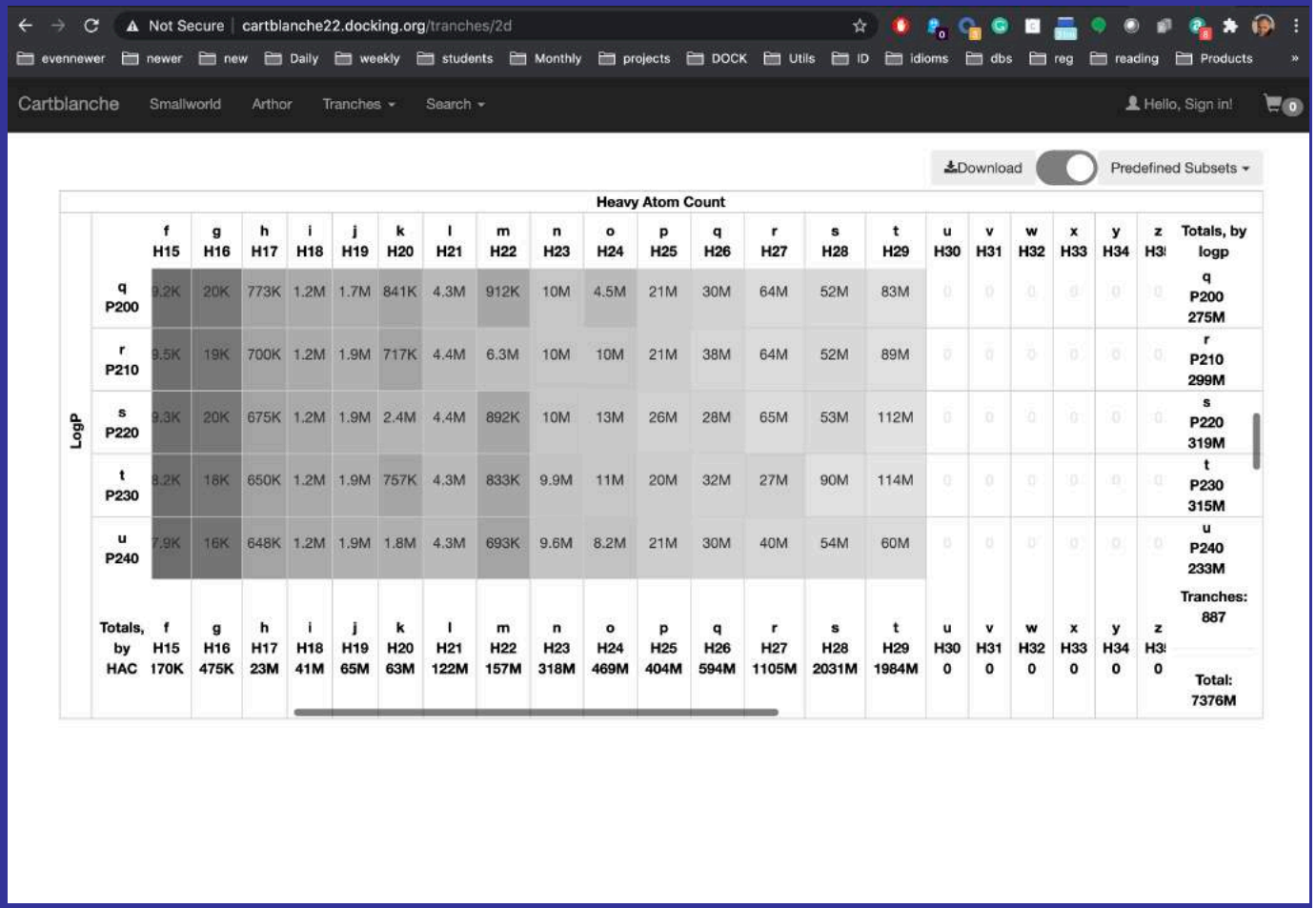
Database of small molecules for ligand discovery

zinc.docking.org



ZINC22 via cartblanche22

Next generation database towards hundreds of billions of molecules

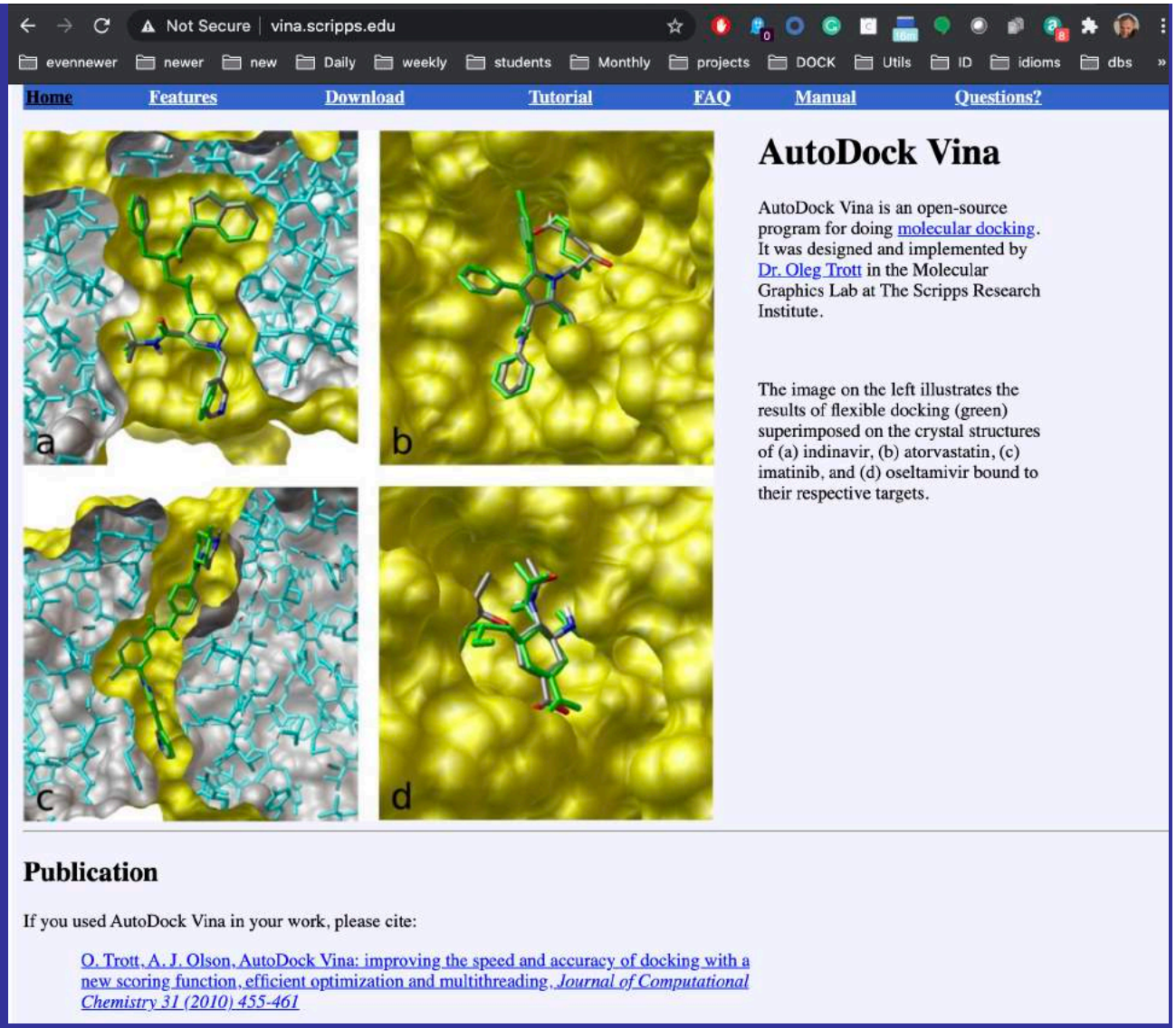


	f H15	g H16	h H17	i H18	j H19	k H20	l H21	m H22	n H23	o H24	p H25	q H26	r H27	s H28	t H29	u H30	v H31	w H32	x H33	y H34	z H3	Totals, by logp
q P200	9.2K	20K	773K	1.2M	1.7M	841K	4.3M	912K	10M	4.5M	21M	30M	64M	52M	83M	0	0	0	0	0	0	q P200 275M
r P210	9.5K	19K	700K	1.2M	1.9M	717K	4.4M	6.3M	10M	10M	21M	38M	64M	52M	89M	0	0	0	0	0	0	r P210 299M
s P220	9.3K	20K	675K	1.2M	1.9M	2.4M	4.4M	892K	10M	13M	26M	28M	65M	53M	112M	0	0	0	0	0	0	s P220 319M
t P230	8.2K	18K	650K	1.2M	1.9M	757K	4.3M	833K	9.9M	11M	20M	32M	27M	90M	114M	0	0	0	0	0	0	t P230 315M
u P240	7.9K	16K	648K	1.2M	1.9M	1.8M	4.3M	693K	9.6M	8.2M	21M	30M	40M	54M	60M	0	0	0	0	0	0	u P240 233M
Totals, by HAC	f H15 170K	g H16 475K	h H17 23M	i H18 41M	j H19 65M	k H20 63M	l H21 122M	m H22 157M	n H23 318M	o H24 469M	p H25 404M	q H26 594M	r H27 1105M	s H28 2031M	t H29 1984M	u H30 0	v H31 0	w H32 0	x H33 0	y H34 0	z H3 0	Tranches: 887 Total: 7376M

Vina

other free molecular docking software

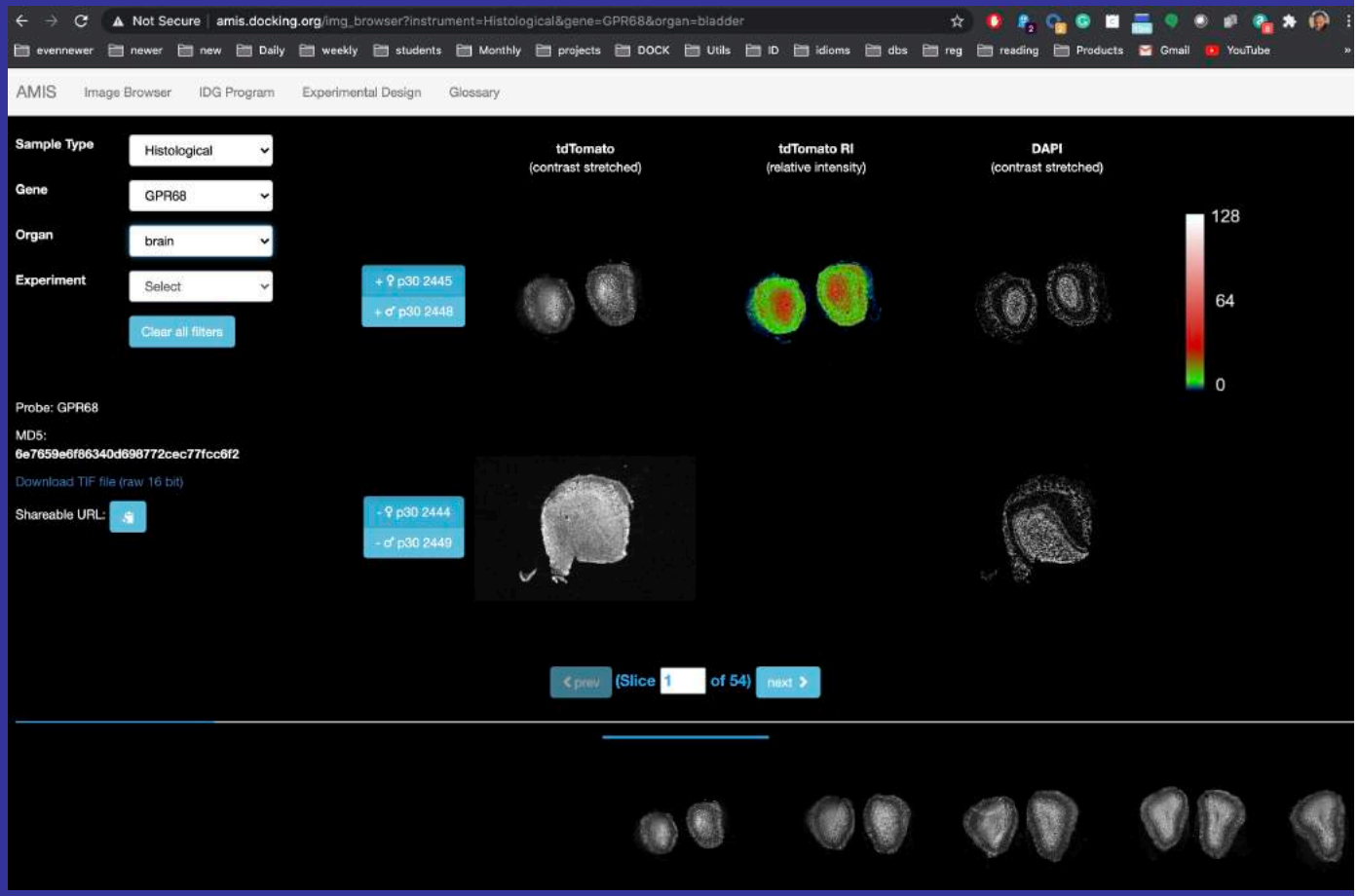
vina.scripps.edu



AMIS

A Mouse Image Server

Gene expression locations



amis.docking.org

Summary

- The problem of exploring chemical space
- Molecular docking as a pragmatic tool for ligand discovery using protein structure
- Inexpensive access to billions of purchasable compounds is changing how ligands are discovered
- Many free software tools you can use