Designing a natural products library

Jasna Klicic Badoux

Nestlé Institute of Health Sciences

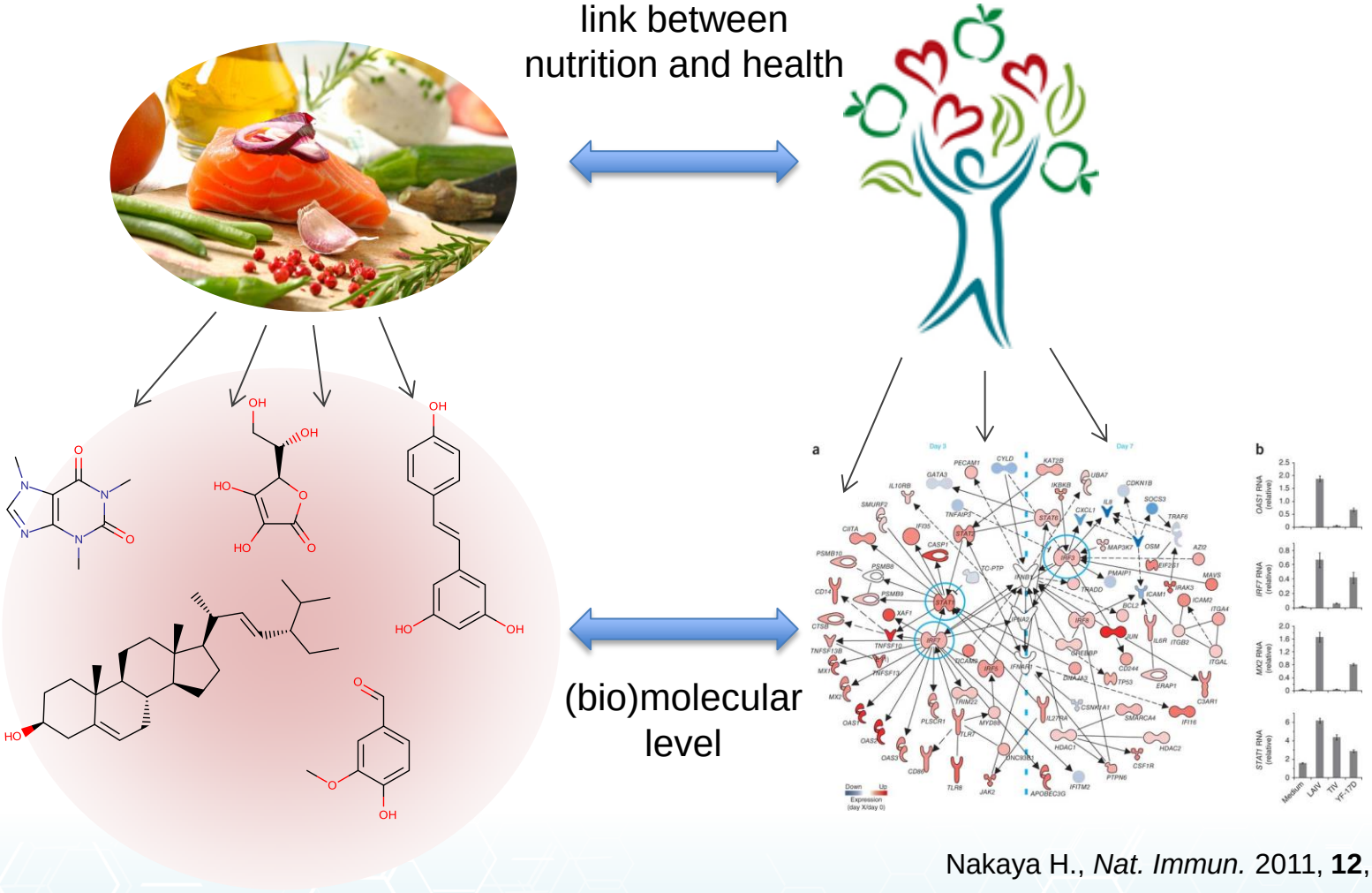
Nestlé Institute of Health Sciences (NIHS)

Developing the scientific foundation of how nutrition can be used to empower people to improve and maintain their health

- founded in 2011
- based near Lausanne, Switzerland
- over 150 scientists

NIHS objective

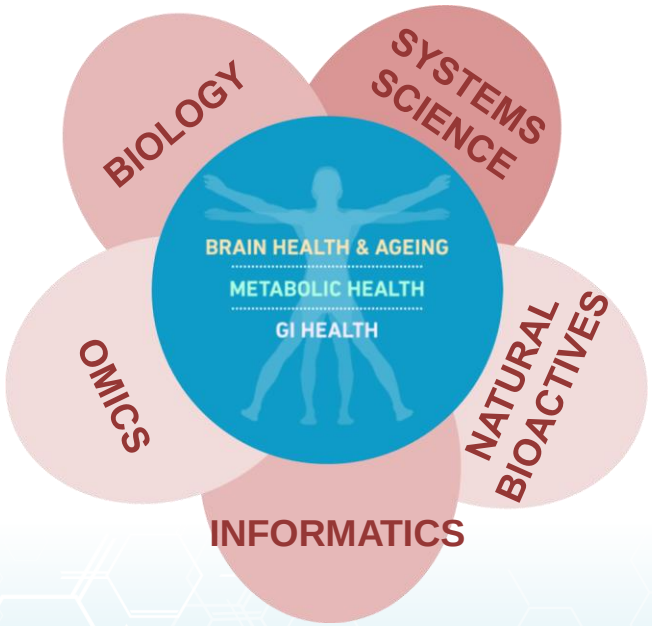
NIHS approach



Nakaya H., *Nat. Immun.* 2011, **12**, 786

Outline

- Natural bioactives
- Chemical space of natural products and drug-like molecules
- Natural Product Library - current status and further development



Natural Bioactives and Screening

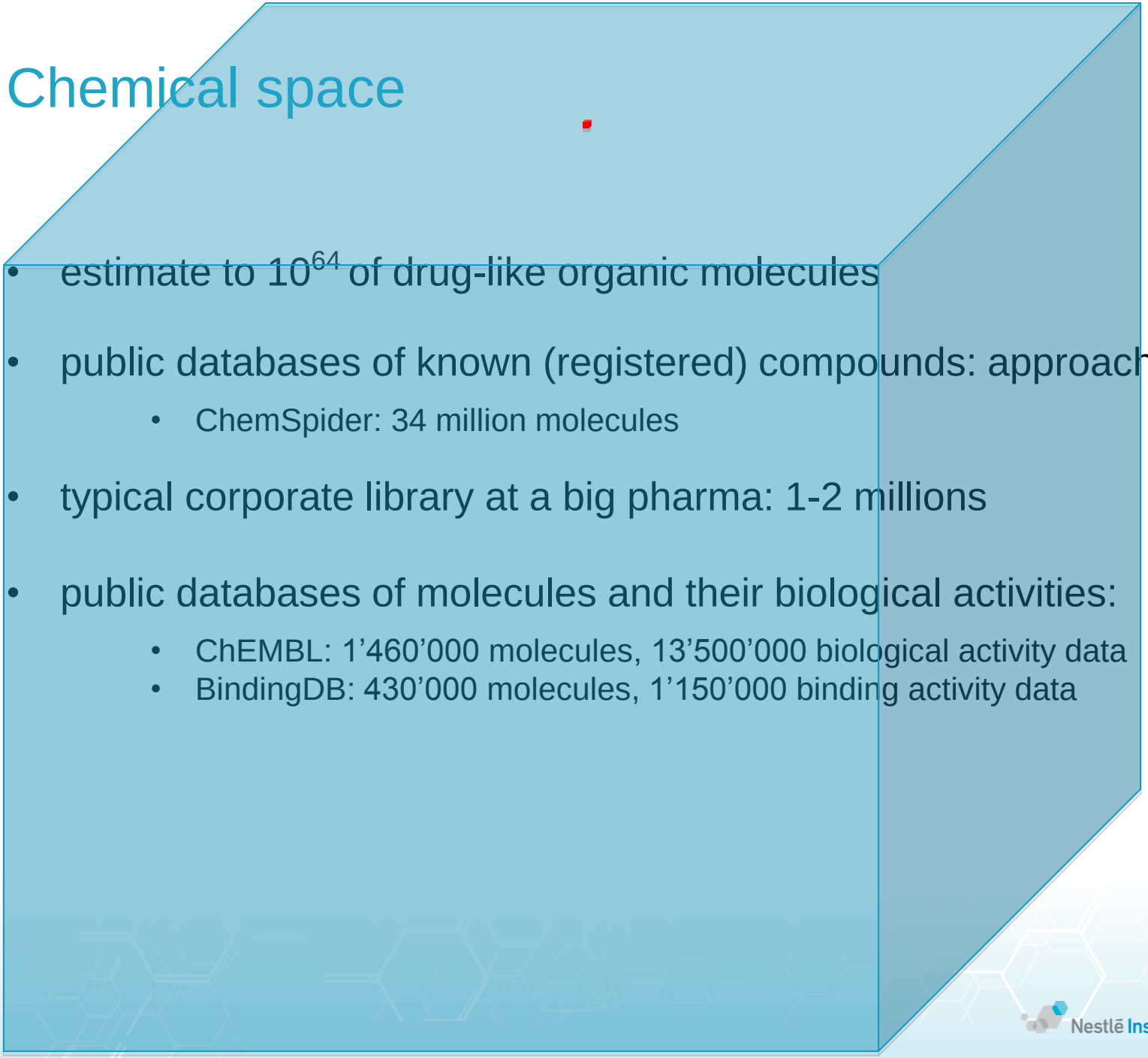
OBJECTIVE: identify natural bioactive molecules by high-throughput screening

1) compounds and extracts screening library
2) HTS platform

Library **quality** is of crucial importance

→ **no bioactives can be identified by screening if they are not in the library !**

Chemical space



- estimate to 10^{64} of drug-like organic molecules
- public databases of known (registered) compounds: approaching 10^{8}
  - ChemSpider: 34 million molecules
- typical corporate library at a big pharma: 1-2 millions
- public databases of molecules and their biological activities:
  - ChEMBL: 1’460’000 molecules, 13’500’000 biological activity data
  - BindingDB: 430’000 molecules, 1’150’000 binding activity data

Natural products (NP)

- *defined as* chemical compounds produced by living organisms
- *traditionally*: substances extracted from plants, microorganisms, fungi and marine organisms
- *enlarged* to include endogeneous metabolites
- known natural products:
  - CRC Dictionary of Natural Products (DNP): 265’000
  - metabolites (HMDB, Metabolights): ~ 40’000
  - lipids (LipidMaps): ~ 40’000

Chemical space of natural products

- how does NP chemical space compare to the chemical space of drug-like molecules?

- concept of drug-likeness:
  → a range of physicochemical and molecular properties that are linked with higher **bioavailability**

- concept of similarity metric
  → required in order to compare two objects
    - physicochemical and molecular properties
    - structural elements: fingerprints, molecular shape..
    - scaffolds: molecular graphs

Natural products vs. druglike molecules: properties

	DNP	DNP deglycosylated 17%	Drugs/druglike
Aromatic atoms	5.2 (7.2)	5.2 (7.0)	11.5 (5.9)
Chiral atoms	5.8 (6.9)	4.0 (4.1)	1.0 (1.6)
Heavy atoms	31.5 (18)	27.6 (13.9)	26.5 (6.6)
Oxygen atoms	6.8 (6.2)	5.1 (4.1)	2.9 (1.9)
Nitrogen atoms	0.6 (1.6)	0.7 (1.7)	2.7 (1.7)
Hydrogen-bond acceptors	6.2 (5.7)	4.6 (3.5)	3.8 (1.7)
Hydrogen-bond donors	3.3 (3.9)	2.4 (2.9)	1.8 (1.5)
Rule-of-five violations	26%	20%	
Aqueous solubility (log *S*)	4.7 (3.3)	4.7 (3.3)	4.6 (1.8)
Molecular weight	444 (257)	388 (195)	375 (91)
Rigid bonds	19.1 (13.0)	16.8 (10.9)	17.9 (6.5)
Rotatable bonds	6.7 (7.1)	5.8 (6.7)	4.3 (2.5)
Rings	3.4 (2.4)	3.0 (2.1)	3.2 (1.3)
Lipophilicity (*S*log *P*)	2.8 (3.2)	3.2 (2.9)	3.5 (1.8)
Topological polar surface area (Å^2)	113 (102)	88 (74)	75 (35)

Natural Products

- higher property variability
- fewer aromatic atoms (1/6)
- more chiral centers: 6 per mol.
- more HB donors & acceptors
- more O, fewer N atoms
- 74% comply with «rule-of-five»

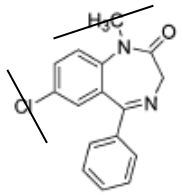
Drugs & Druglike

- more aromatic atoms (1/2)
- 1 chiral center per molecule
- fewer HB donors & acceptors
- 3 O/N atoms per mol.

Grabowski K., *Nat. Prod. Rep.* 2008, 25, 892

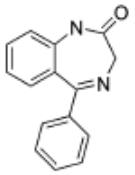
Natural products vs. druglike molecules: scaffolds

a)



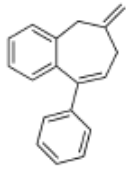
Molecule

b)



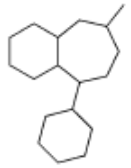
Atomic Scaffold

c)



Carbon Scaffold

d)



Graph Scaffold

	Drugs/druglike	DNP	DNP (deglycosylated)
Molecules, total	121 593	128 600	111 474
Acyclic	1826 (2%)	11 448 (9%)	11 507 (10%)
Atomic scaffold (AS)	51 338 (43%)	31 050 (27%)	24 788 (25%)
AS singletons	35 519 (30%)	18 450 (16%)	14 491 (14%)
Carbon scaffold (CS)	32 469 (27%)	28 144 (24%)	22 172 (22%)
CS singletons	19 373 (16%)	16 372 (14%)	12 633 (13%)
Graph scaffold (GS)	21 296 (18%)	19 962 (17%)	15 039 (15%)
GS singletons	11 594 (10%)	10 905 (9%)	7973 (8%)

→ overall broader scaffold diversity for druglike molecules than for natural products !

Number of scaffold that account for 50% of the database:

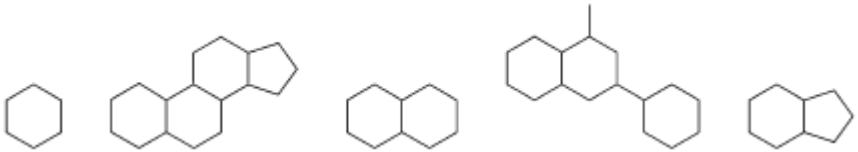
Scaffold	Drugs/druglike	DNP	DNP deglyc.
atomic	5492	1188	795
graph	458	278	159

Grabowski K., *Nat. Prod. Rep.* 2008, 25, 892

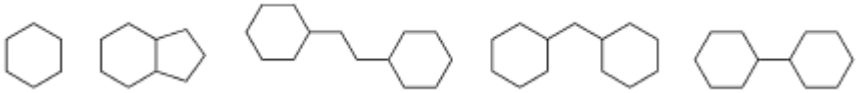
Most common graph scaffolds

OVERALL

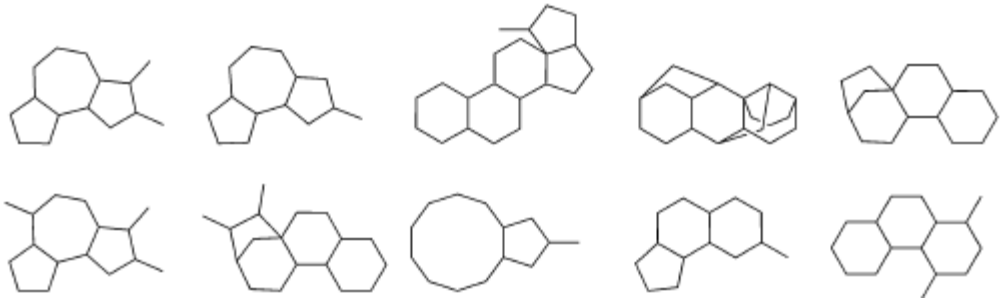
natural products, deglycosylated



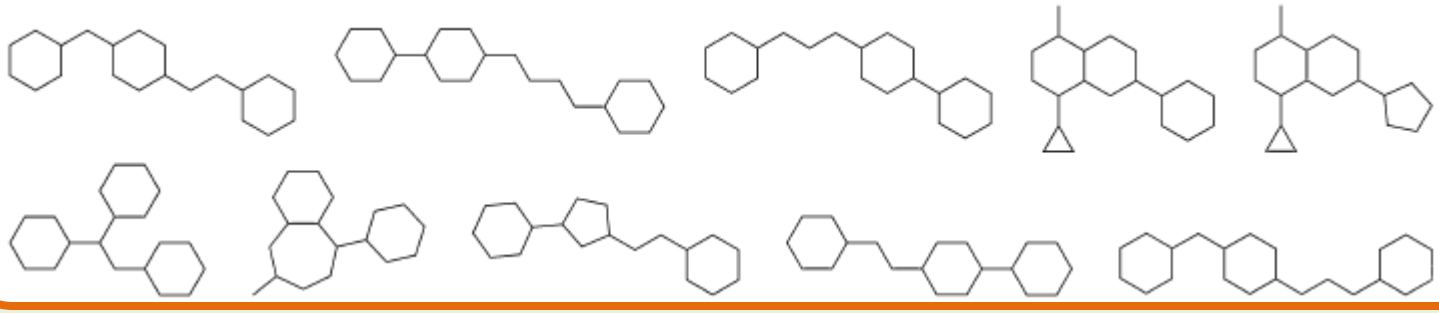
synthetic drugs



EXCLUSIVE



natural products, deglycosylated

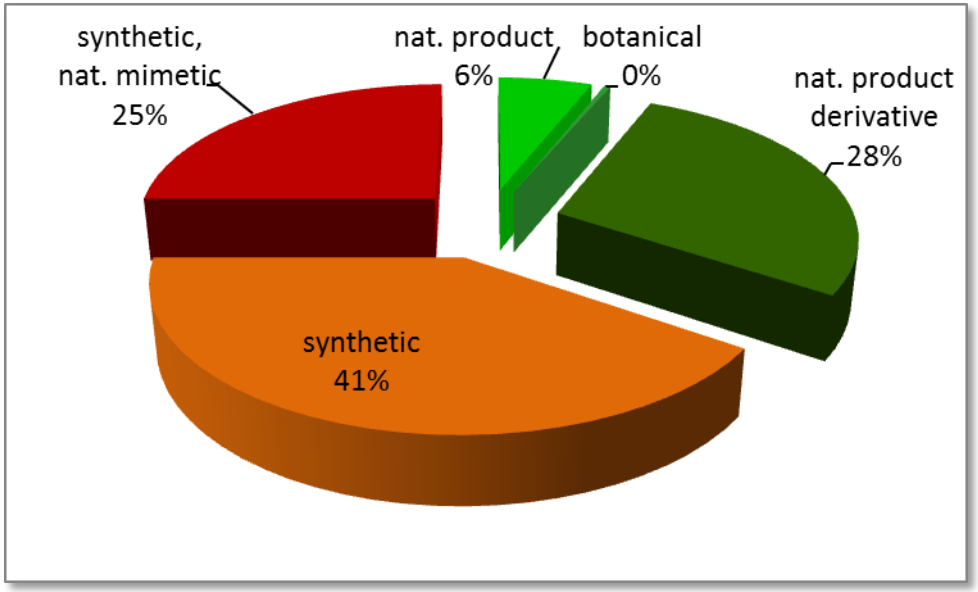


synthetic drugs

Grabowski K., *Nat. Prod. Rep.* 2008, 25, 892

Origins of drugs (30 year period)

Drugs are compounds with validated pharmacological activity → bioactives



origins of all new approved drugs (N=1355) from 1981-2010

modified data from Newman, D., *J. Nat. Prod.* 2012, **75**, 311

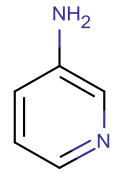
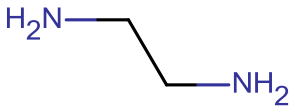
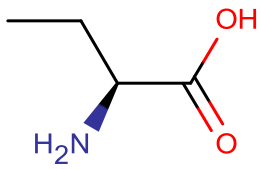
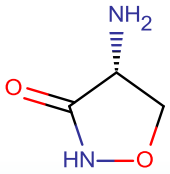
→ Natural products evolved to interact with multiple proteins

NIHS compound library – current status

→ physchem and molecular properties very similar to DNP averages

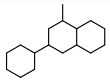
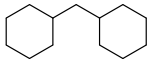
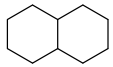
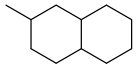
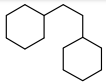
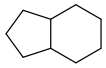
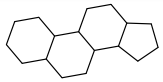
	JChem (16794 mols)				RDKit (17418 mols)			
Property	**Mean**	SD	Min	Max	**Mean**	SD	Min	Max
exact mass	**439.66**	230.05	58.04	4879.60	**437.22**	227.91	58.04	4879.60
heavy atoms	**31.31**	16.05	4	331	**31.15**	15.89	4	331
hetero atoms					**8.14**	5.86	0	151
O atoms	**7.58**	5.93	0	151				
N atoms	**0.53**	1.32	0	23				
halo atoms	**0.03**	0.23	0	6				
chiral atoms	**6.12**	6.86	0	120				
HBA (Lip)	**7.06**	5.62	0	151	**8.07**	5.85	0	151
HBD (Lip)	**3.61**	3.64	0	92	**3.64**	3.66	0	92
HBA	**7.06**	5.62	0	151	**7.63**	5.76	0	151
HBD	**3.61**	3.64	0	92	**3.59**	3.61	0	92
nrotb	**5.99**	5.29	0	118	**5.45**	4.87	0	118
fsp3	**0.55**	0.27	0	1	**0.54**	0.27	0	1
nrings	**3.59**	2.14	0	30	**3.61**	2.15	0	30
aliphatic rings	**2.42**	2.29	0	30	**3.26**	2.29	0	30
saturated rings					**1.83**	2.17	0	30
hetero aliph. rings	**1.28**	1.37	0	30	**0.99**	1.31	0	30
hetero aliph. rings 6	**0.82**	1.17	0	9				
hetero aliph. rings 5	**0.36**	0.67	0	30				
logP	**2.17**	2.65	-43.52	19.90	**1.83**	2.75	-66.27	16.81
TPSA	**125.01**	94.49	0	2405.73	**124.45**	93.88	0	2405.73

Hydrogen bond donors/acceptors - examples

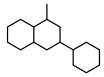
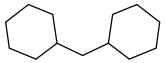
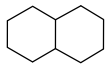
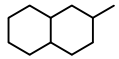
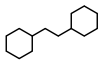
	HBA			HBD		
Structure	JChem	RDKit Lip	RDKit	JChem	RDKit Lip	RDKit
	2	2	2	1	2	1
	2	2	2	2	4	2
	3	3	2	2	3	2
	3	4	3	2	3	2

NIHS cpd lib – most common graph scaffolds

JChem:

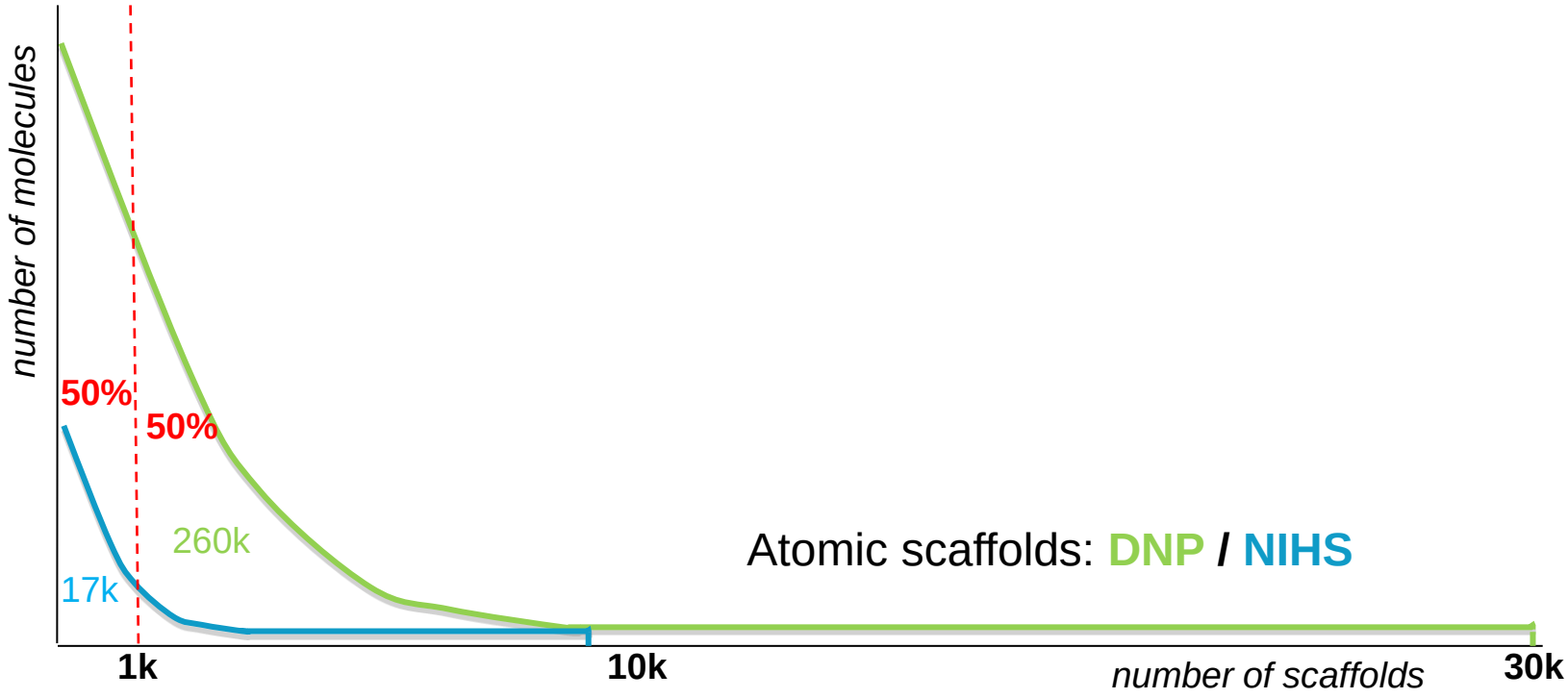
Scaffold	Freq
	954
	344
	280
	263
	261
	194
	165
	158
	137
	134
acyclic	817
No. scaff	4366

RDKit:

Scaffold	Freq
	1005
	325
	270
	267
	256
	153
	151
	140
	132
	124
acyclic	806
No. scaff	6164

NIHS cpd lib: scaffold diversity

atomic scaffold distribution, DNP vs NIHS



Library design strategy

Compound selection strategy:

1) **diversity** at the molecular scaffold level
2) **focus** subsets for selected targets or target families
3) **chemical stability** and **bioavailability** calculated into the selection
4) preferrence for **clusters** over than singletons

Compound source:

- commercial compounds from special suppliers
- external synthesis
- in-house isolation from extracts

Acknowledgements

Denis Barron
Sofia Moço
Yann Ratinaud
Martine Cabo
Loraine Merimod
Jaroslaw Szymczak
Terry Reilly
Bruce O'Neel

Thank you !