AstraZeneca
IMED Biotech Unit

reening collection enhancement through open-innovation

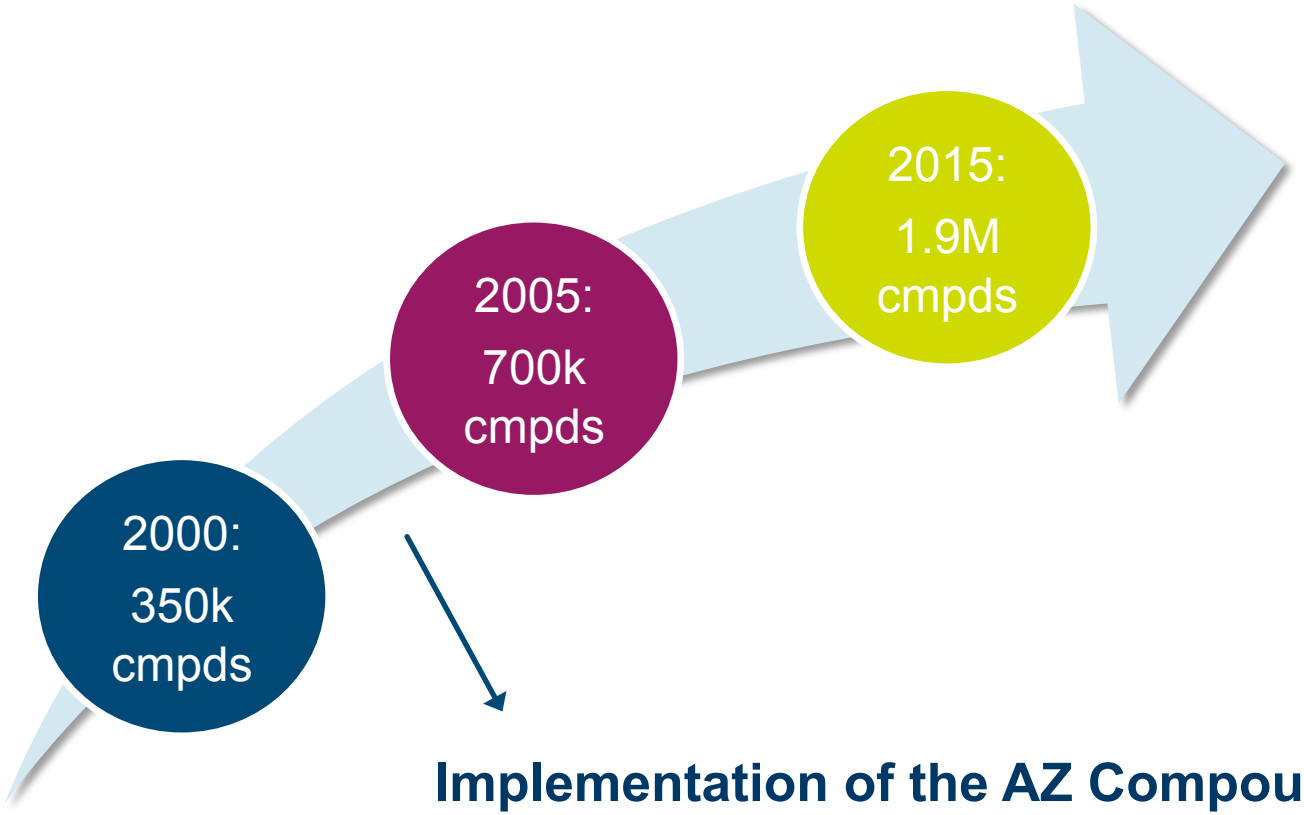
ry Kogej

nal Sciences, Computational Chemistry

8th March 2017

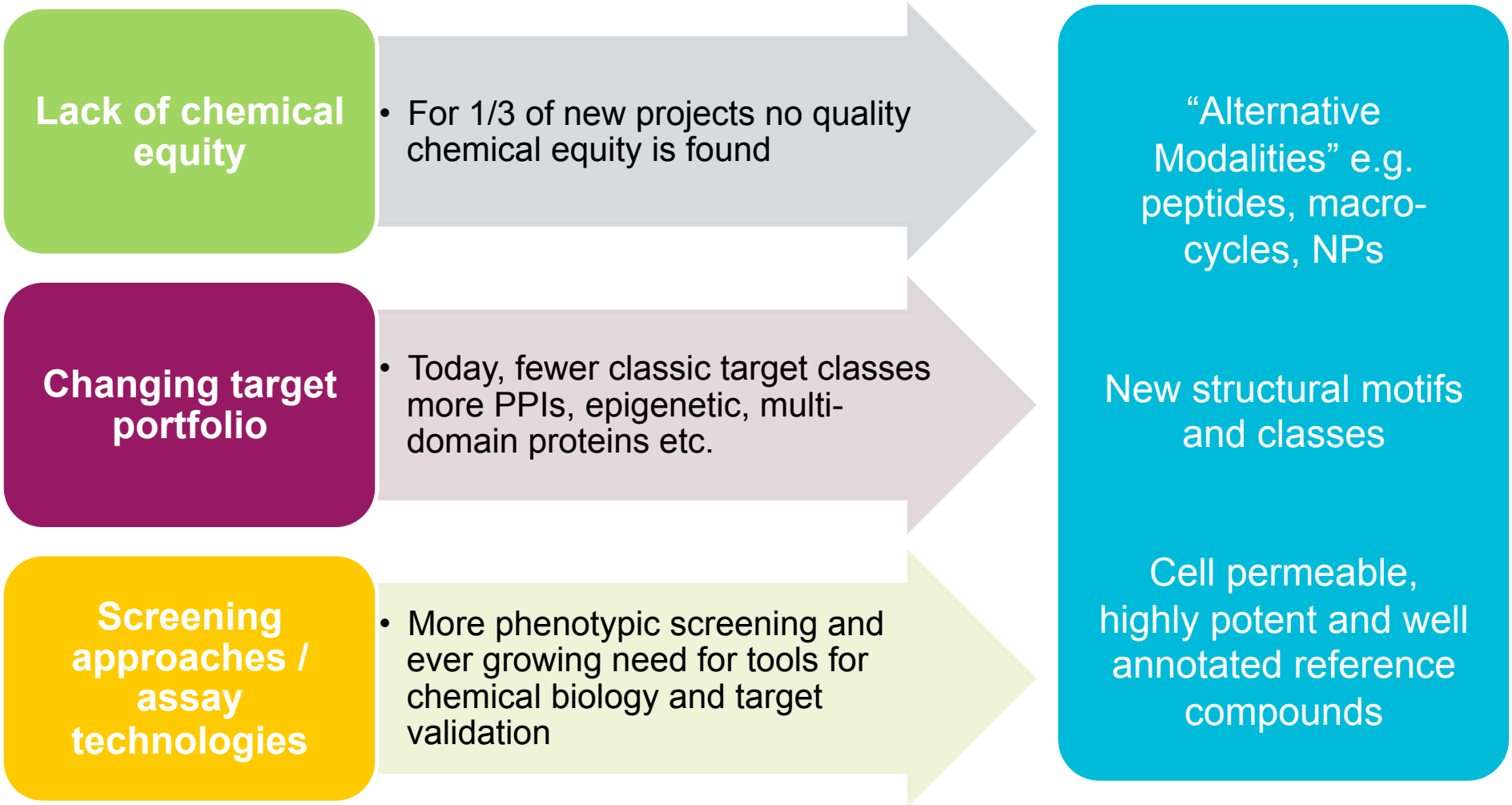
gh Throughput Screening

S remains one of the few methods to **discover novel chemical equity** and remains the **benchmark for lead generation**
S success rates are typically between 50 and 60%

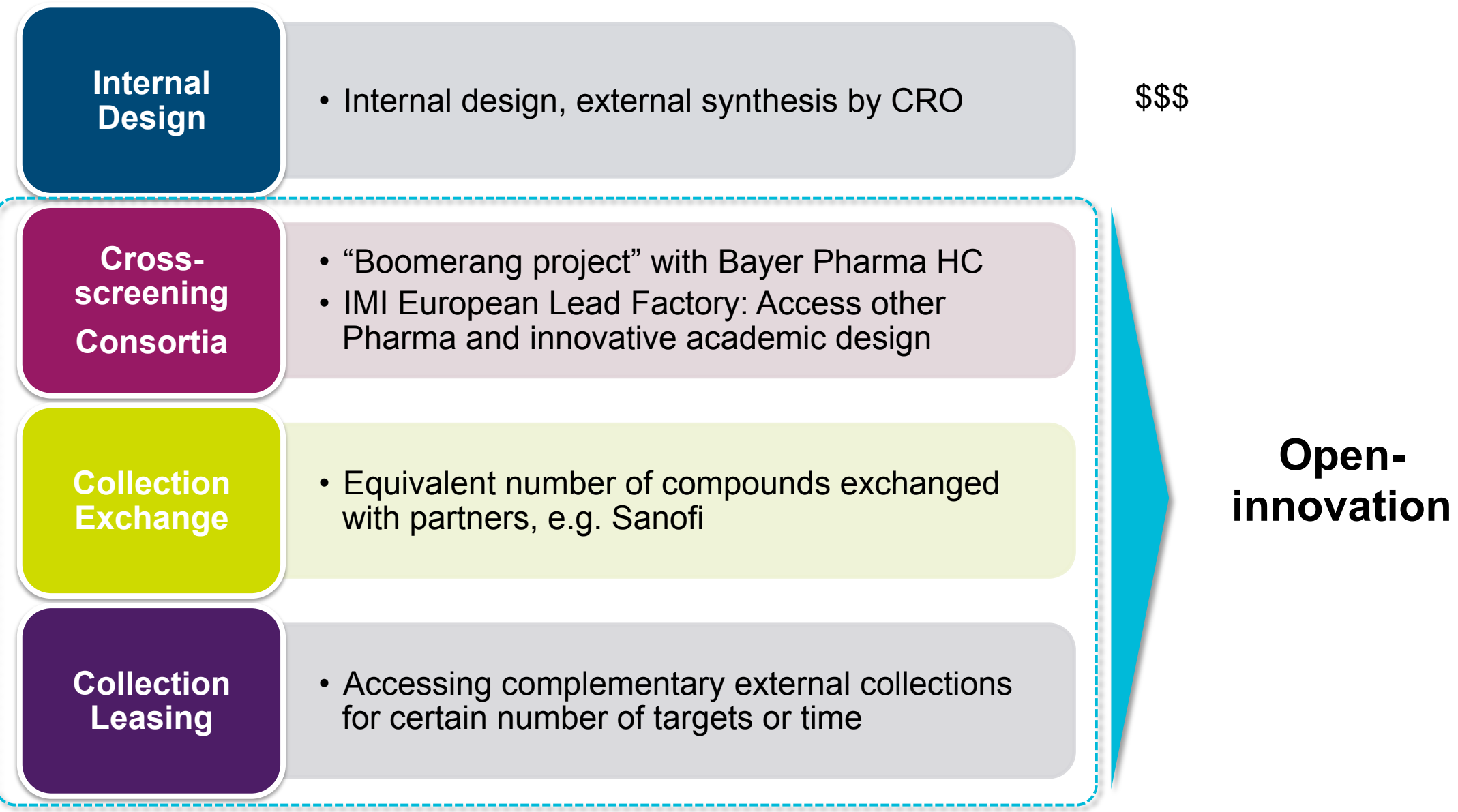


**Implementation of the AZ Compoun
Collection Enhancement programm
(CCE programme)**

o we still need to enhance our collection?

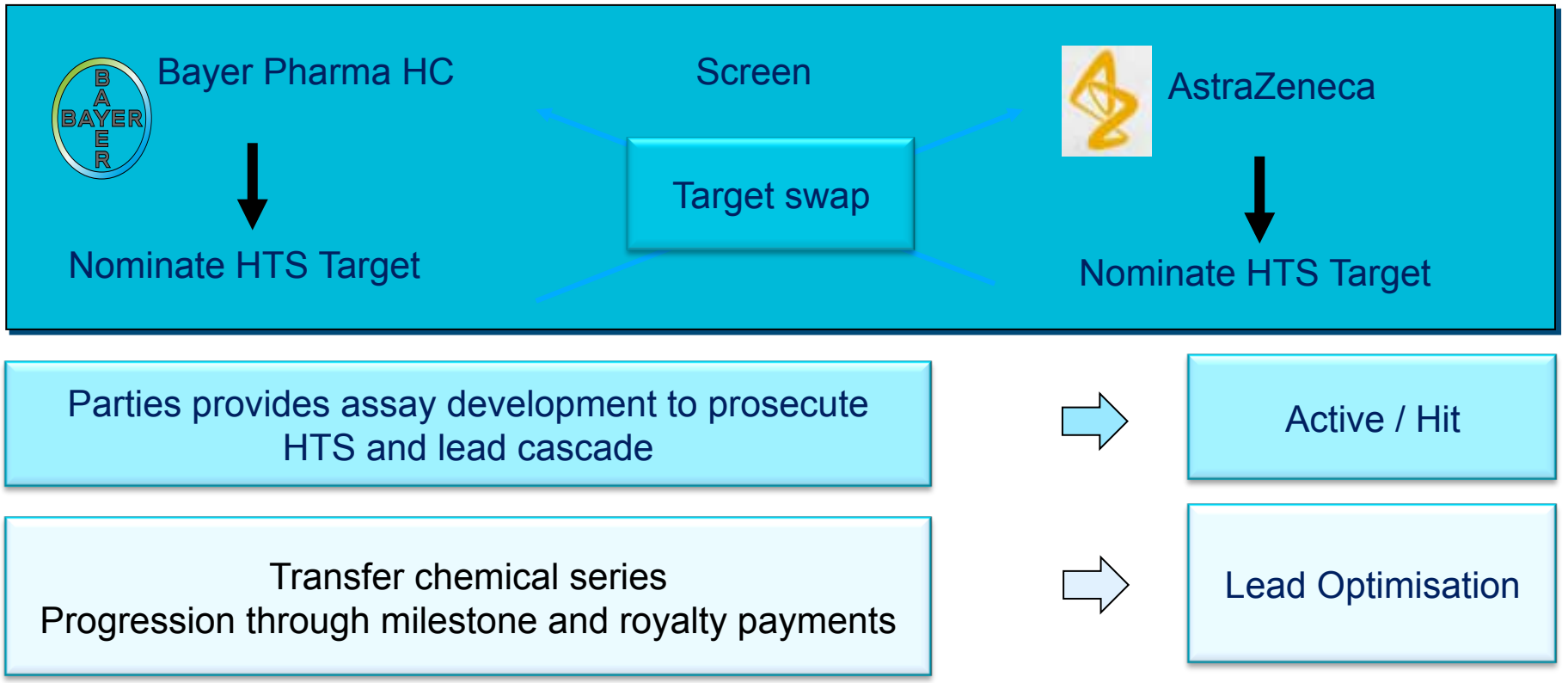


lany ways to access novel chemistry



ayer Pharma HC - AstraZeneca ‘Boomerang’ project
a successful example of peer-peer collaboration

Pioneering Joint initiative established between AstraZeneca and Bayer in 2010 (alliance xtended until 2016), based on mutual trust and shared values
Enables both parties to seek chemistry starting points not available in their internal collections



ayer and AstraZeneca - Origin of libraries

Bayer Collection

2.7M structures

- SCHERING part of the collection (875.000 structures) was cleaned and expanded library with mainly purchasable external compounds (2003-2005)
- Huge investments at BAYER between 2000 and 2007 to expand library based on proprietary building blocks
- Compound design based on favorable PhysChem properties and undesirable groups filtering
- Realization through *external* collaborations and internal combinatorial chemistry
- 1/3 classical medchem structures from optimization projects
- 2/3 combichem compounds

AstraZeneca Collection

1.4M structures

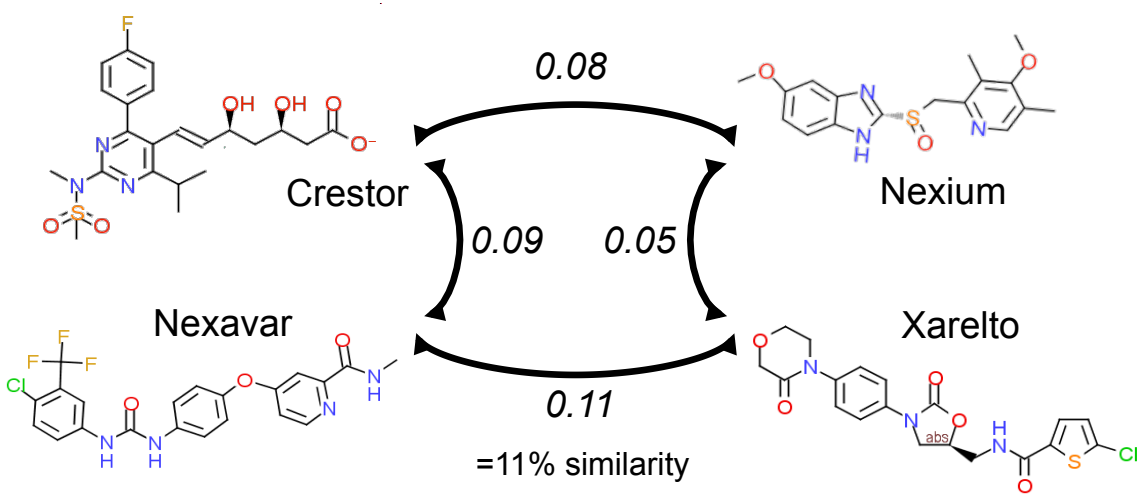
- AstraZeneca screeing collection underwent major review (structural and sample quality) in 2001/2002
- Strict classification on Phys-Chem and structural features
- Major acquisition campaign in 2002
- Three consecutive *Compound Collection Enhancement* programmes (2003-2005, 2006-2008, and 2008-2011)
- Internal design from Lead Generation chemistry, outsourced production of small libraries, no combichem
- >80% proprietary compounds

- The Bayer collection consists of MedChem designed proprietary classical and combinatorial compounds
- The Astra collection consists of MedChem designed project and enhancement compounds and acquisitions

omparison of Bayer and AstraZeneca collection

On-site “workathon” at AstraZeneca Sweden with Bayer + AstraZeneca scientists (in 2011)
Collaborative calculations and analysis effort of pre-prepared fingerprint and data-files
3 days of compute time
1920 cores / Infiniband QDR 40Gbit interconnect
...resulting in 150 GB of data

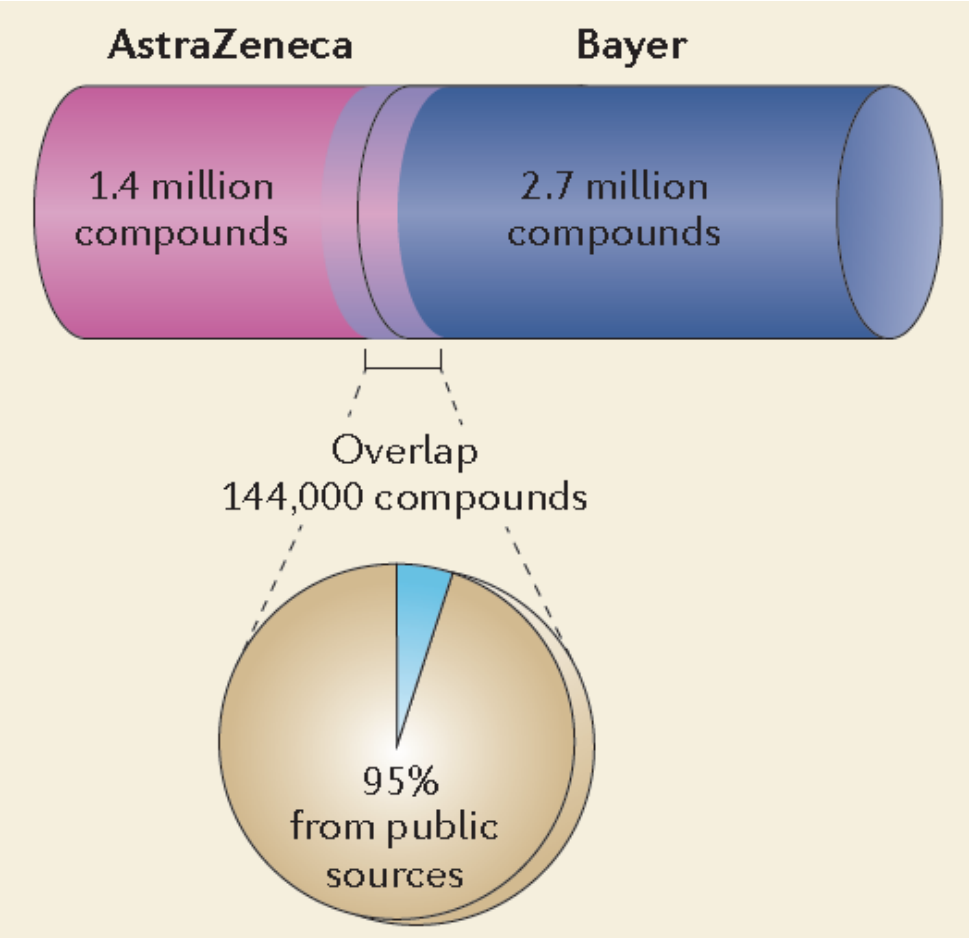
our laptops and >>50 cups of coffee...
ery fruitful scientific discussions on ethods and results!



milarity is based on comparison of fingerprints using the *tanimoto distance*
CFP4 Fingerprints are derived with Pipeline Pilot.
uplicate fingerprints for each collection were eliminated prior to computation

- Diverse research areas and projects led to structural diverse libraries
- Fingerprint and Tanimoto Index (ranging [0-1]) map this diversity

verlap of Bayer Pharma HC and AstraZeneca collection dentical fingerprints



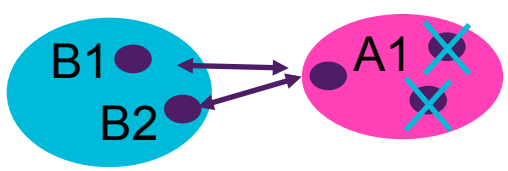
- 3.3% of the total collection (Bayer + AZ is overlapping*)
- 95% of the overlap are public domain compounds

⇒ Screening of > 4.2 Millions unique cmpds

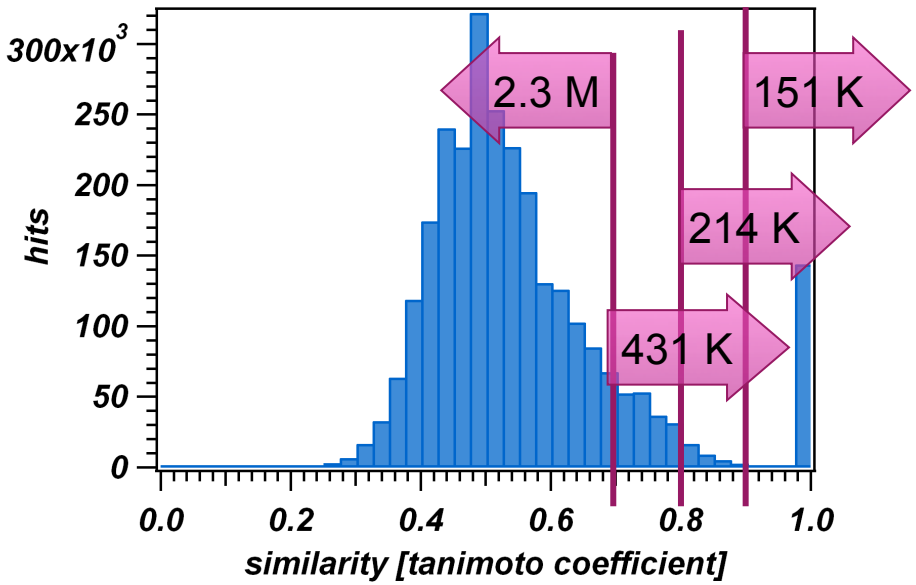
*) As we are not sharing structures for analysis the overlap is based on exact match of molecular fingerprints (ECFP4). This is an overestimate of identity as a small fraction of non-identical compounds will have the same fingerprint

Big pharma screening collections: more of the same or unique libraries? The AstraZeneca-Bayer Pharma AG case
Kogej T, Blomberg N, Greasley PJ, Mundt S, Vainio MJ, Schamberger J, Schmidt G, Huser J. Drug discovery today (2013).

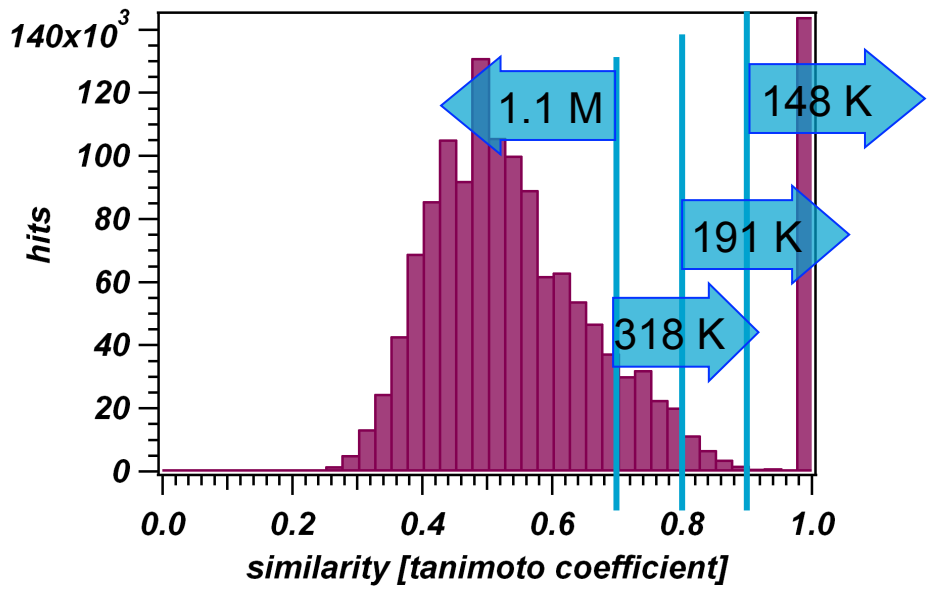
milarity of Bayer and AstraZeneca collection
earest Neighbour Distribution



Bayer collection



AstraZeneca collection

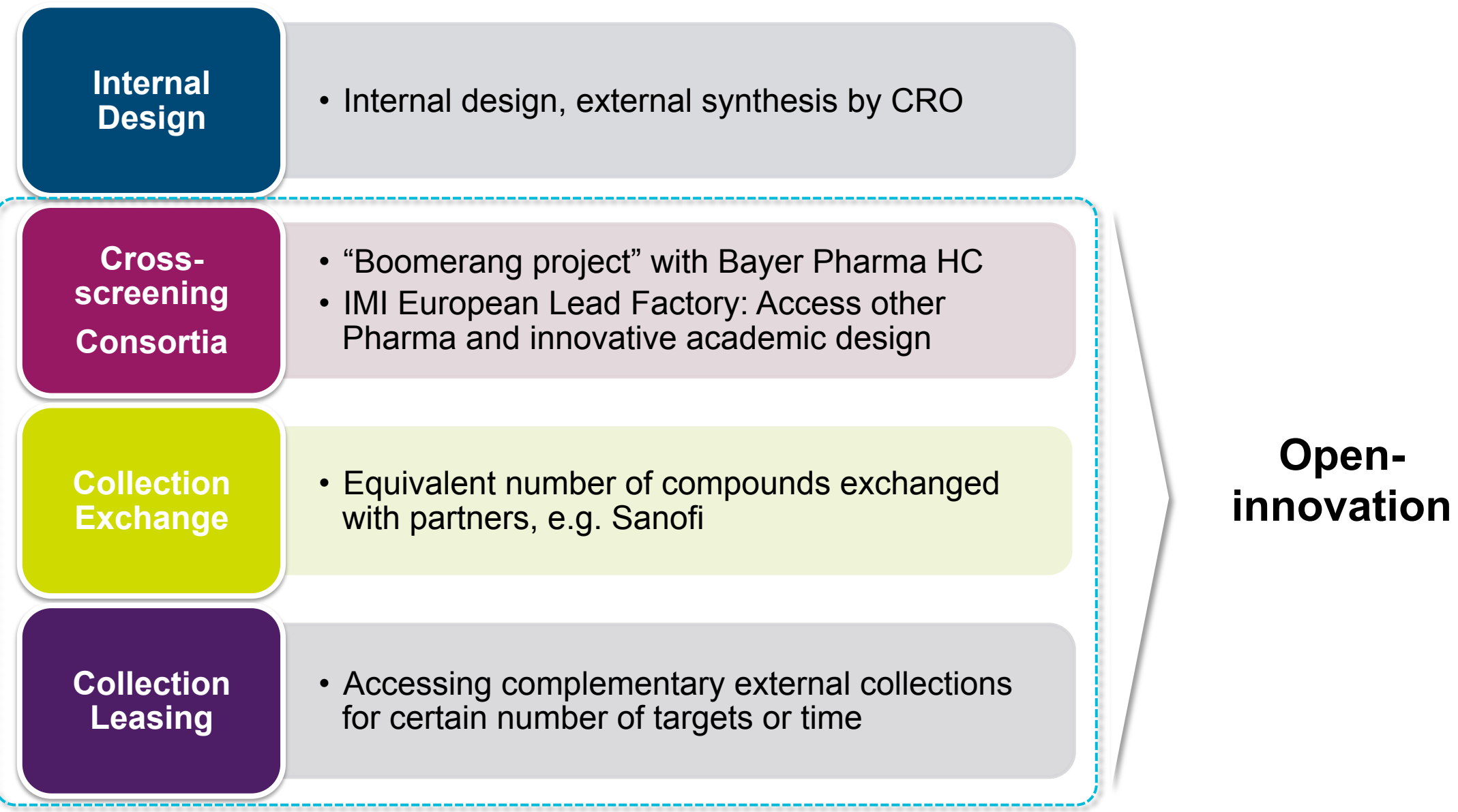


”The AZ business case”:
2.3 M Compounds are ”new” to Astra

”The Bayer business case”:
1.1 M Compounds are ”new” to Bayer

*A ECFP4 similarity cutoff of 0.70 seems high, but the main objective of this study was to access that the two collections had complementary or alternative chemical series that might not contain necessarily totally novel chemotypes or being in a different part of the chemical space. We should also consider that some subtle changes in the similarity value can dramatically change the target space of the compound. e.g. slightly reduce/increase nitrogen basicity.

lany ways to access novel chemistry



Collection exchanges

rinciple: *quid pro quo* exchange of compounds
o cash payments
o royalties

Identify set of compounds suitable for sharing

ot AZ patents
ot restriced by other alliances
ot in active projects
gh quality physical sample and
fficient amount (no stock depletion)

Join meeting

2 ICore7, 8 threads, SSD
2 days (including travel)

Physically exchange t samples

2 days
Standalone computer (not network, clusters)
Isolated room
Sharing structures

straZeneca: Sanofi exchange

210k novel screening compounds
200 µL of 10 mM solution (>10y HTS)

Strong relationship foundation for future collaboration

Positive example of cross Pharma collaboration

FINANCIAL TIMES

Sanofi-AstraZeneca chemicals swap takes open route to R&D

The deal — the biggest of its kind among major pharmaceuticals groups — highlights an increased openness to co-operation in an industry criticised in the past for secrecy and beggar-thy-neighbour practices.

The Daily Telegraph

AstraZeneca and Sanofi agree novel drugs-sharing deal

Britain's AstraZeneca and French pharmaceutical company Sanofi have agreed a landmark deal to share data - for free - in the hope that it could lead to breakthrough treatments for disease.

SCRIP Intelligence

Is AstraZeneca/Sanofi Library Exchange Open Innovation's Future?

THE WALL STREET JOURNAL.

Drug Giants Seek Edge By Sharing Secrets

AstraZeneca PLC and Sanofi SA have agreed to share thousands of their proprietary chemical compounds with each other, an unusual deal that shows the creative lengths to which pharmaceutical companies will go to pursue new drugs.

REUTERS

Sanofi, AstraZeneca swap compounds in new twist on open drug R&D

hat we were/are interested to ?

structurally **diverse organic scaffolds** from low to high molecular complexity

diverse heterocycles including ones from known privilidged structures

fragment screening set of small heterocycles, small undecorated polycycles, plus various common organic linkers and branching groups

near and cyclic **peptides** with established bioactivity

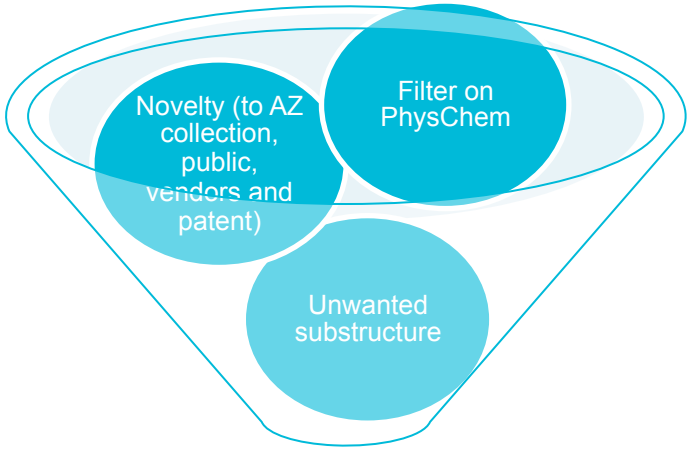
novel oligomers based on **helix, strand, turn and coil structures**

known drugs and clinical candidates

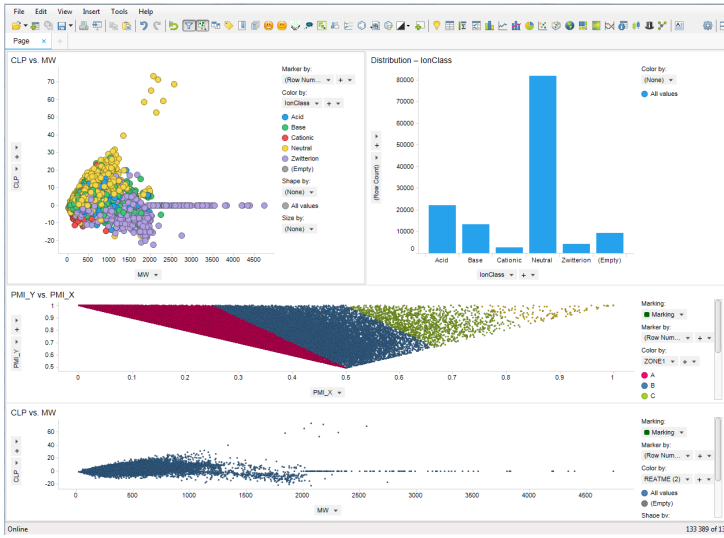
diverse collection of isolated **natural products**

lection process

”Filtering OUT”



Result visualization



lue annotation”

- Fingerprint

imilarity to iterature, atent active cmpds

iversity to Z collection

Intrinsic molecular properties

Desirable Physchem (lead-likeness,…)

“3D” (FSp3, PMI)

Modelling

Model likeness (Bayesian) - Natural-Product

2D substructures

Macrocycles

Kinase cores

Specific scaffolds

Virtual Screening

β-turn

α-Helix

PPI

Clustering, grouping

visual inspection

Final selection

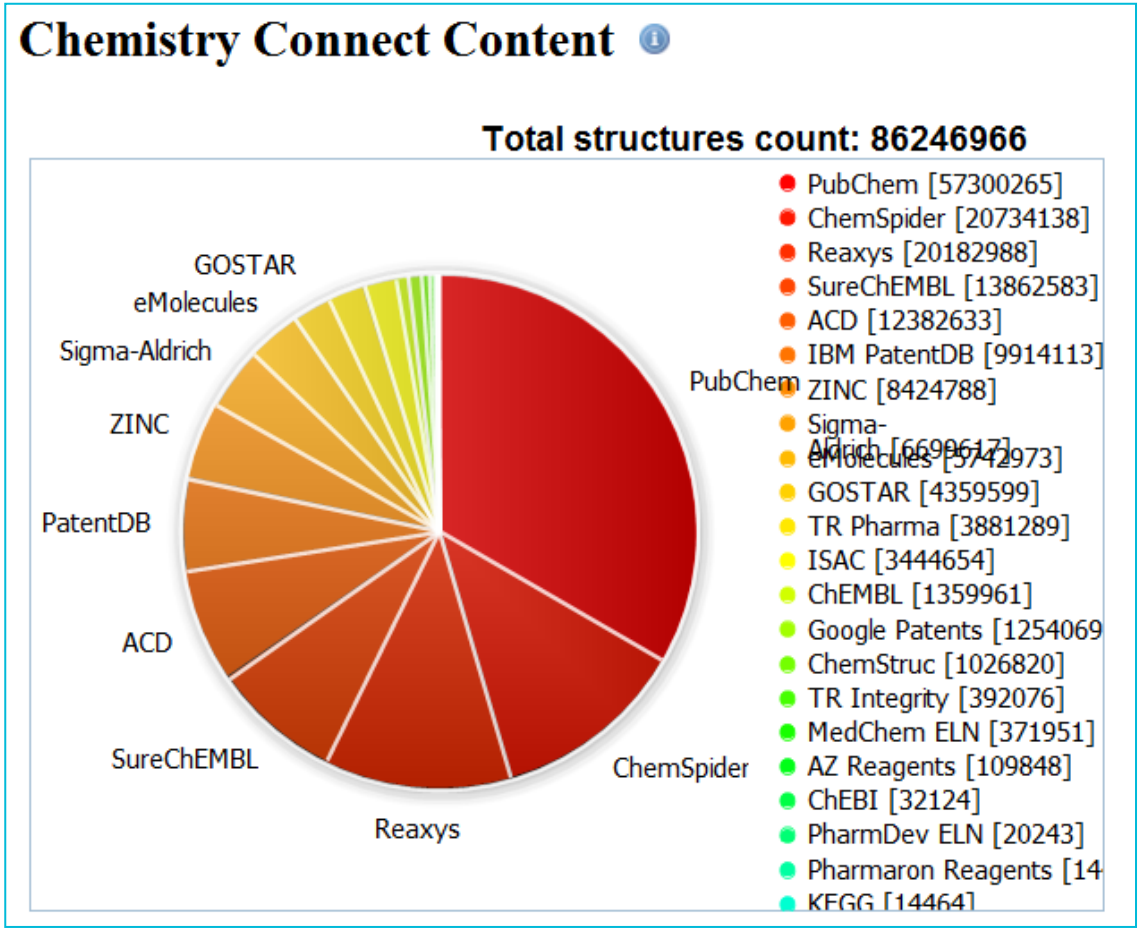
heminformatics assessment of external collections

Novelty to AstraZeneca

gh interest in Sanofi
opriatory compounds

w in interested in
blically available compounds

gh interest in ”tool compounds”



Making every SAR point count: the development of Chemistry Connect for the large-scale integration of structure and bioactivity data.
Muresan S, Petrov P, Southan C, Kjellberg MJ, Kogej T, Tyrchan C, Várkonyi P, Xie PH. Drug Discov Today. 2011, 16, 1019-30

heminformatics assessment of external collections

Novelty to AstraZeneca

Compound filters and alerts

AZFilters®

Property filters

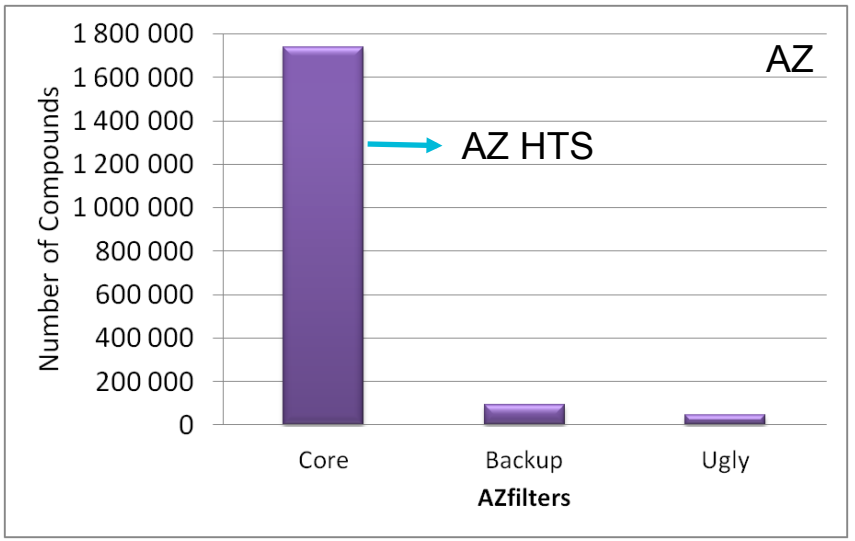
- 100 ≤ MW ≤ 550
- -2 ≤ ClogP ≤ 6
- 1 ≤ PSA ≤ 160

Chemical filters

- ~150 chemistry alerts (ca 500 SMARTS)
- 11 classes (reactives, unwanted structures,…)

	Prop. Filters	Chem. Filters
Core	No violation	No match
Back-Up	1 violation	No match
Ugly	no matter	>= 1 match
Ugly	>= 2 violations	no matter

Chemical predictive modelling to improve compound quality
Cumming JG, Davis AM, Muresan S, Haeberlein M, Chen H. Nature Review Drug Discovery, 2013, 12, 948-62



lerts: - Genotoxicity via Bursi alert
- Reactive Metabolite
- PAINS structures

heminformatics assessment of external collections

Novelty to AstraZeneca

Compound
filters and alerts

PhysChem

heminformatics assessment of external collections

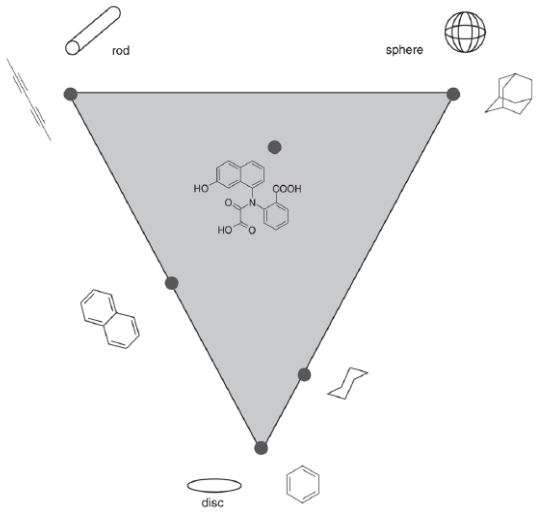
- Novelty to AstraZeneca
- Compound filters and alerts
- Physchem
- Molecular Complexity

Bioactive Molecules: Perfectly Shaped for Their Target?

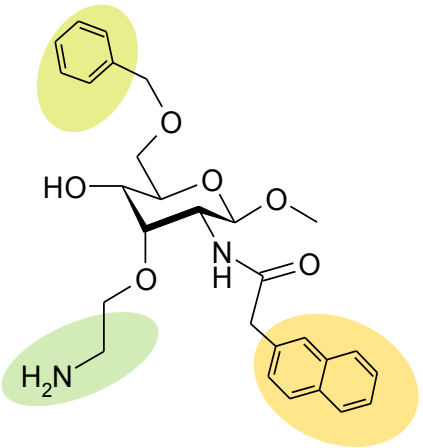
Matthias Wirth*[a] and Wolfgang H. B. Sauer[a]

Mol. Inf. 2011, 30, 677 – 688

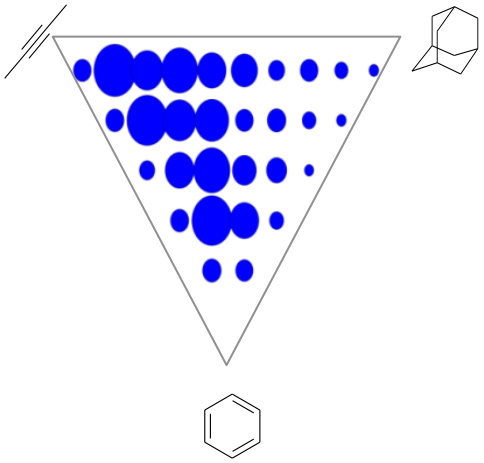
Principal moment of inertia



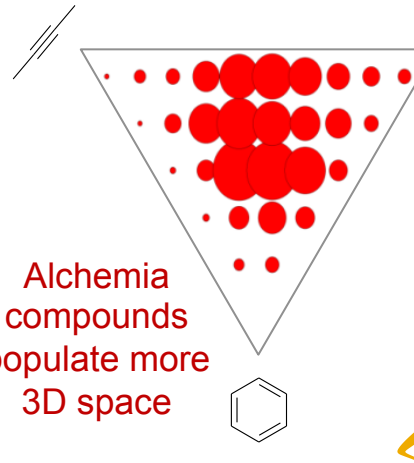
Alchemia



AZ collection



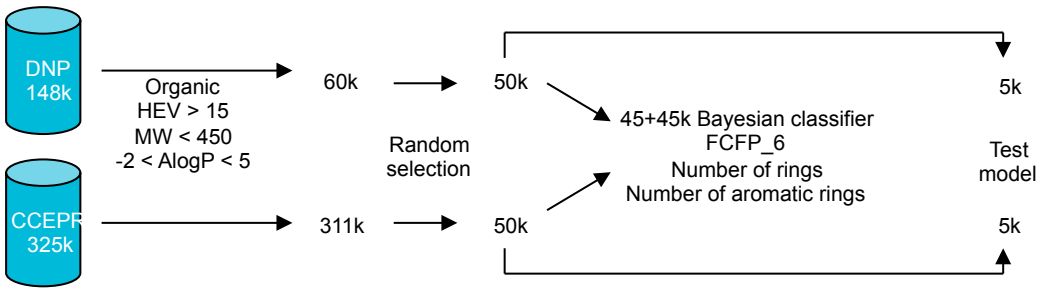
Alchemia 'sub-library



Alchemia compounds populate more 3D space

heminformatics assessment of external collections

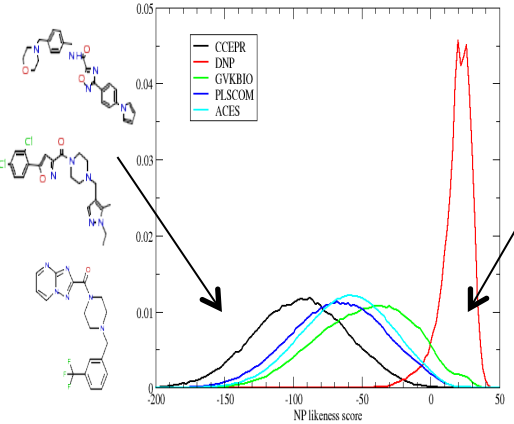
- Novelty to AstraZeneca
- Compound filters and alerts
- Physchem
- Molecular Complexity

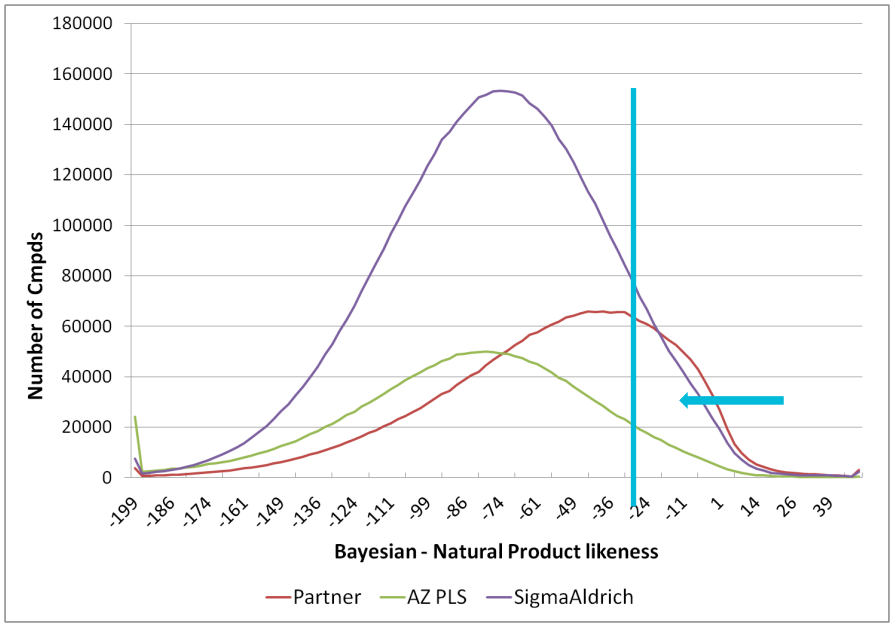


[Have tried different variants]

Ertl P, et al. J Chem Inf Model. 2008, 48, 68

AUC for ROC (TPR v. FPR): 0.999
Best split: -22.85





Score > -25 is associated with "natural product like" cmpd

heminformatics assessment of external collections

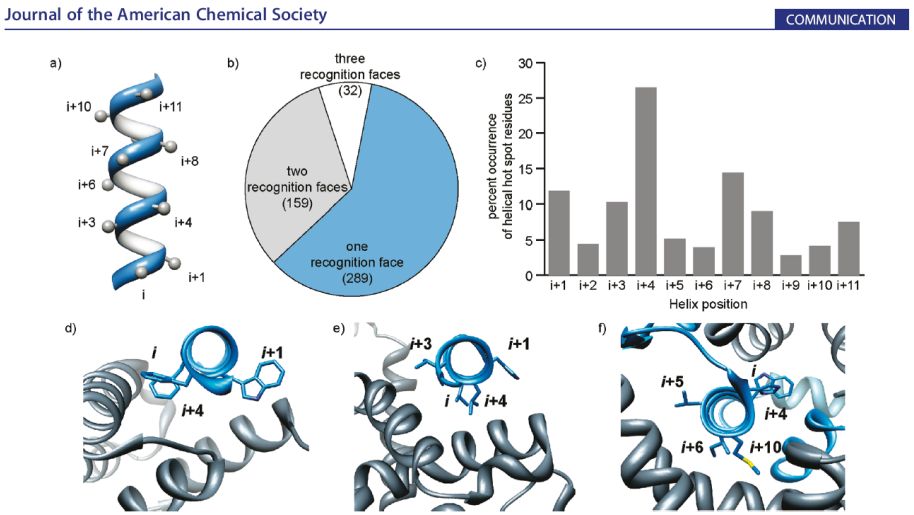
- Novelty to AstraZeneca
- Compound filters and alerts
- Physchem
- Molecular Complexity
- D shape/2D search

2D structures:
Macrocycles, spiro scaffold, “ring of the futures”, privileged structures

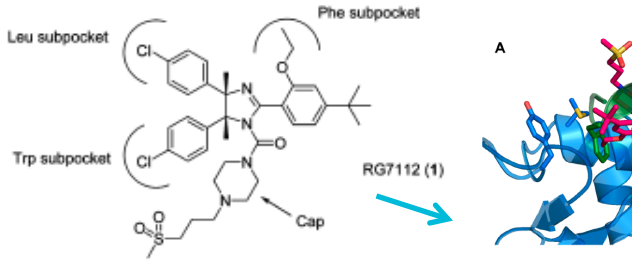
3D Know mimetics:
-helix i.e. nutlins
β-strand → i.e. peptidomimetics
β-hairpin → macrocycles
PPI Co-crystelize X-ray structure (ref:)

OpenEye tools →

- Filter database to include only compounds with 25 < HAC < 40
- Generate omega db (max 100 conf
- Run ROCS with different tanimoto depending on the query



JACS 2011,133,14220



ACS Med. Chem. Lett. 2013, 4, 660

heminformatics assessment of external collections

- Novelty to AstraZeneca
- Compound filters and alerts
- Physchem
- Molecular Complexity
- D shape/2D search
- Fingerprint-based ilarity/diversity
- Clustering
- Visual inspection

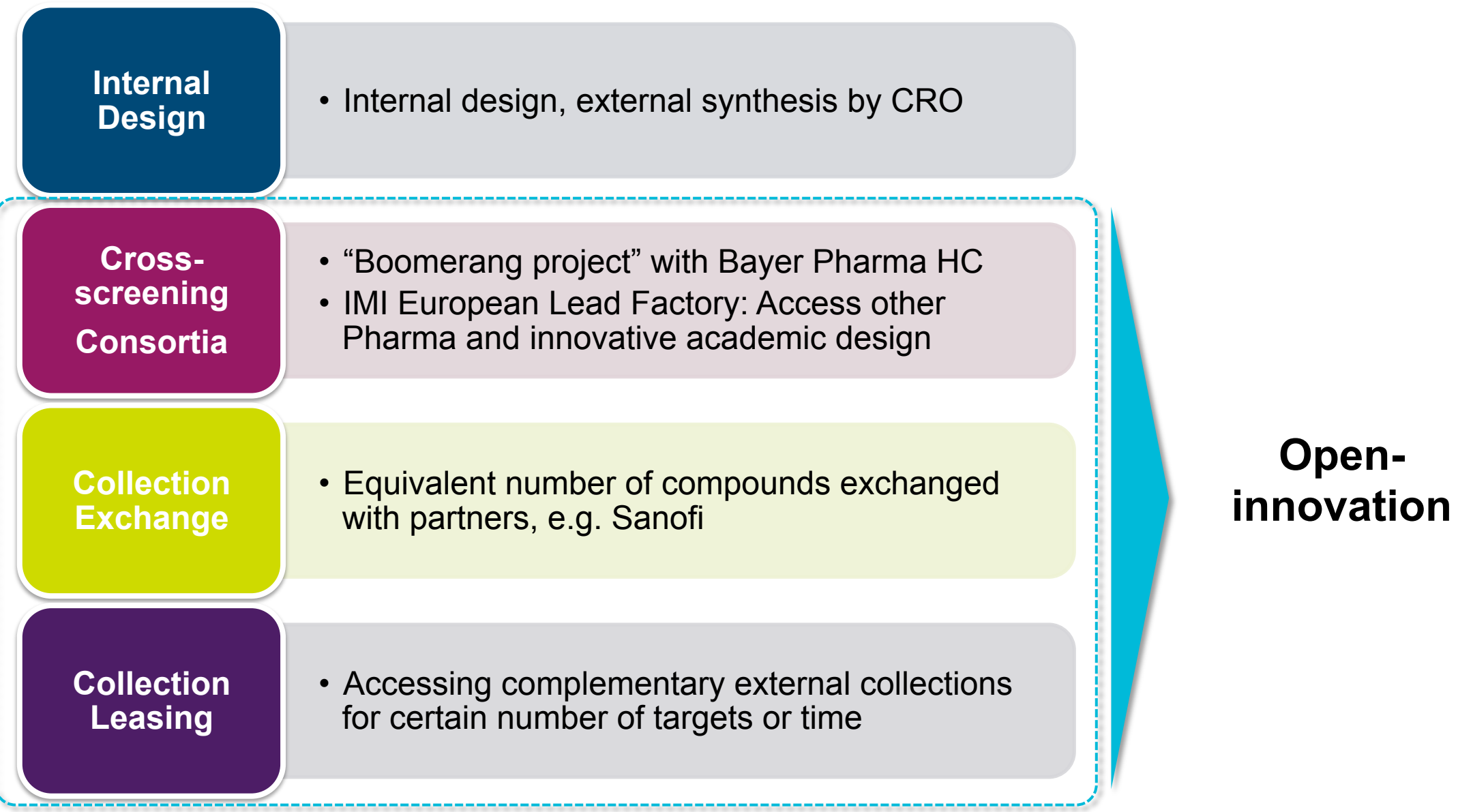
Assessment based on 2D fingerprint based (Tanimoto – ECFP4):

- Similarity to reference active compounds from ChEMBl, GVKBio to enrich AZ biological active space
- Diversity to AZ collection

- Ensure maximum coverage of external collection
- Reduce singleton risk
- Use of 2D fingerprint approach and scaffold analysis (Murcko)

Last but not least, review from medicinal chemists

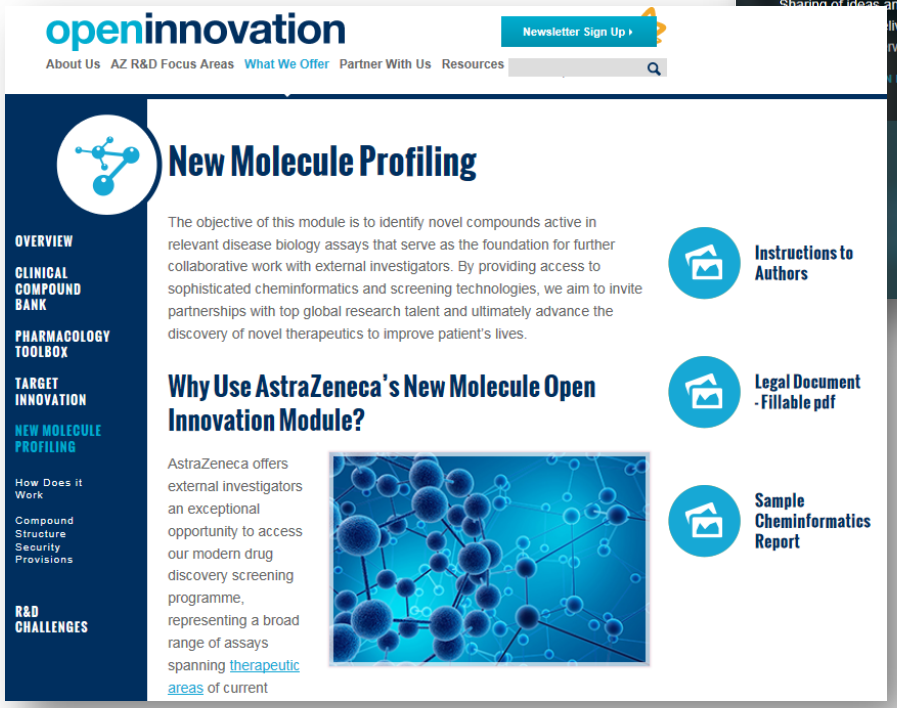
lany ways to access novel chemistry

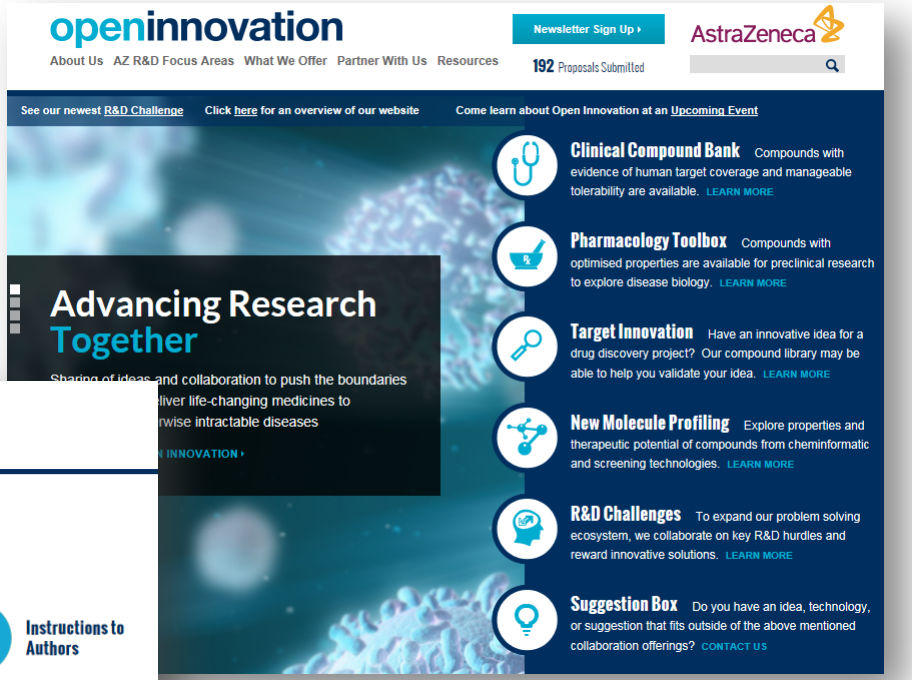


://openinnovation.astrazeneca.com

AZ Open Innovation: New Molecule Profiling

Your molecules are welcome!



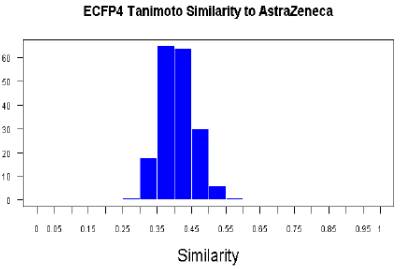


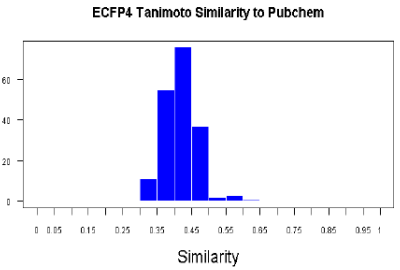
More under

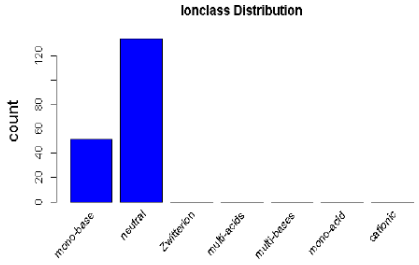
http://openinnovation.astrazeneca.com

heminformatic report

Similarity

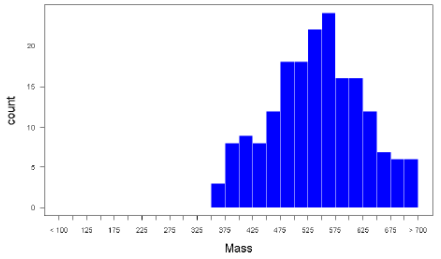


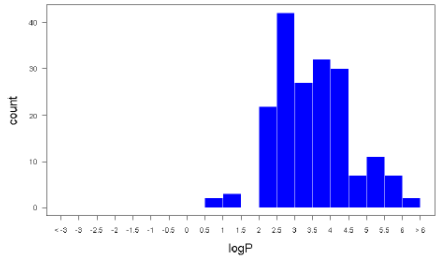


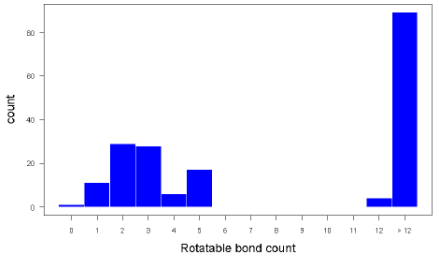


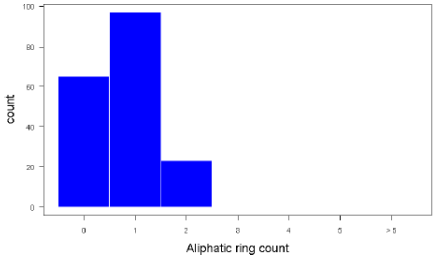
Main considerations for recommendation:
- Novelty
- Molecular complexity
- "Drug discovery friendly"

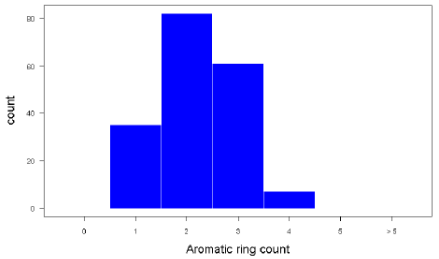
Physchem properties

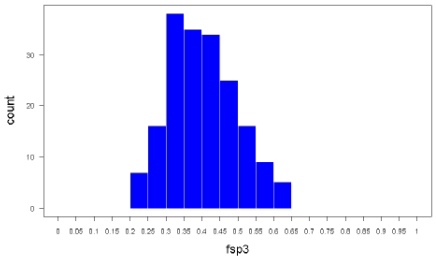


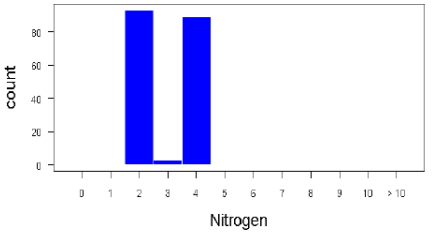


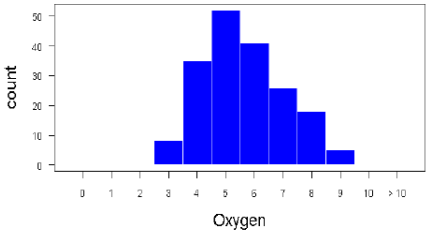












Structure filters

GenoTox	Count	Percentage
No	185	100.00%
Yes	0	0.00%

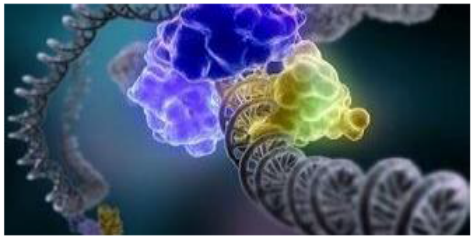
Reactive metabolite	Count	Percentage
No	185	100.00%
Yes	0	0.00%

Risk level	Count	Percentage
No Risk	74	40.00%
Risk	111	60.00%
All Known explosives	0	0.00%

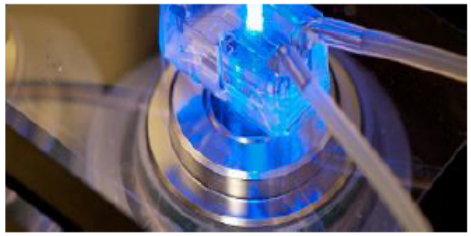
Controlled	Count	Percentage
No	185	100.00%
Yes	0	0.00%

When	Metals and metalloids
After standardization	[ ]
Before standardization	[ ]

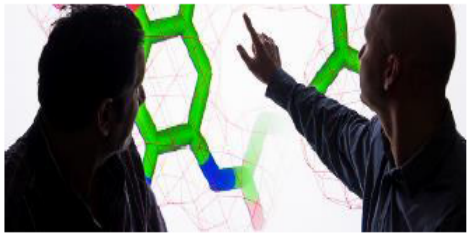
Being open for collaboration…



…creating an environment where science thrives



…and challenge conventional thinking



…will enable us to change the way we treat disease and transform lives

What science can do

With special thanks to…

HTS:

Mark Wigglesworth

Martina Fitzek

Marian Preston

Carolyn Blackett

Dave Murray

Kirsty Rich

Matt Collier

OI:

Craig Wegner

Pam Hill

Hitesh

Discovery Sciences:

Mike Snowden

Steve Rees

Dave Smith

Michael Kossenjans

Selmi Nidhal

David Andrews

Clive Green

Kevin Cross

Ian Sinclair

John Cuff

Phil Spencer

Abdul Ingar

Rick Davies

,

Computational Chemistry:

Ola Engkvist

Hongming Chen

Isabella Feieberg

Thierry Kogej

BD:

Iain Comely

Terry Reed

Duncan Young

Vicki Foster

IMED

Fred Goldberg,

Lena Ripa

ackup

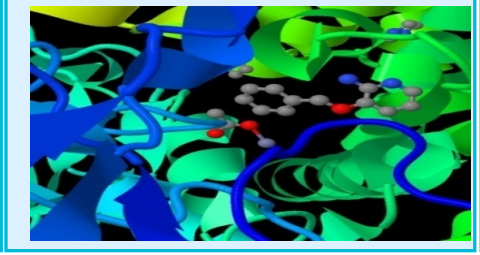
Z hit finding strategy: multiple approaches

Diversity Screening

- High-throughput screening (HTS)
- Broad set of assay technologies for all types of targets
- Strong capacity for cherry picking

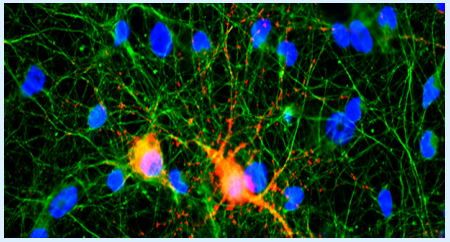
Structural Biology & Fragments

- Large experience in FBLG by NMR & SPR, and assays
- Dedicated team to perform chemistry
- Broad range of biophyscial technologies
- Specific fragment library for FBLG



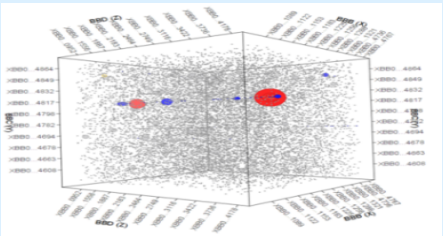
Phenotypic Drug Discovery

- Hit discovery using primary / stem cell assays for finding of pathway specific hits
- iPSC alliance with external partner
- Strategy for target deconvolution



DNA Encoded Library Technologies

- Screening of very large libraries of compounds (>10^{10} molecules)
- Collaboration with external partner
- Build-in selectivity profiling in hit finding strategy
- Proprietary libraries



Public-Private & Private Private Partnerships

- Hit discovery using academic and industri partners
- Novel chemical space
- Novel biological space
- Synergies between partners

Key to all approaches is access to high quality screening compounds!