Synthetically Accessible Virtual Inventory (SAVI) : Reaction generation and handling at the billion compound scale

Hitesh Patel, Yuri Pevzner, Wolf D. Ihlenfeldt, Marc Nicklaus

Chemical Biology Laboratory
Center for Cancer Research
National Cancer Institute
National Institutes of Health

Idea of SAVI

What can I make easily, reliably, safely and cheaply?

Problems of de novo design

Not in screening databases

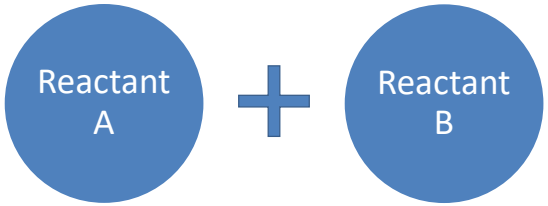
Many steps and very costly

Potentially unsafe

Syntheses often unsuccessful

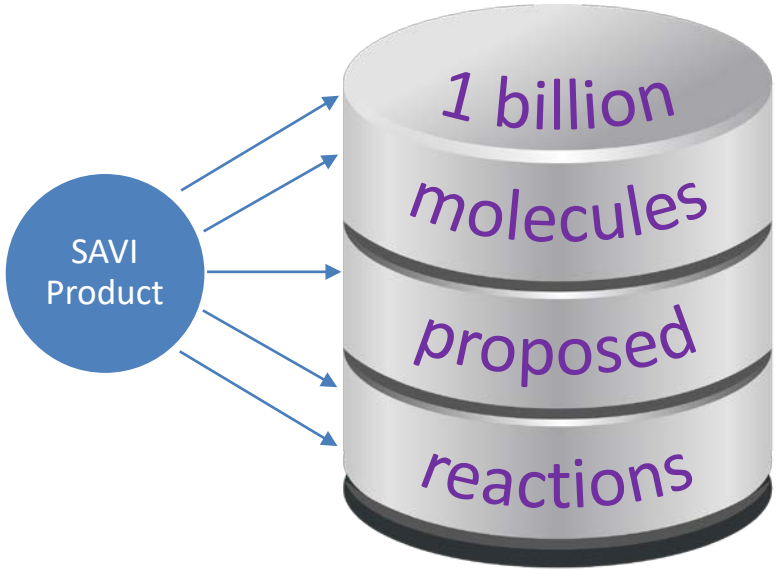
Goal of SAVI

you can buy the starting materials



1-step reaction

no exotic equipment or dangerous procedures

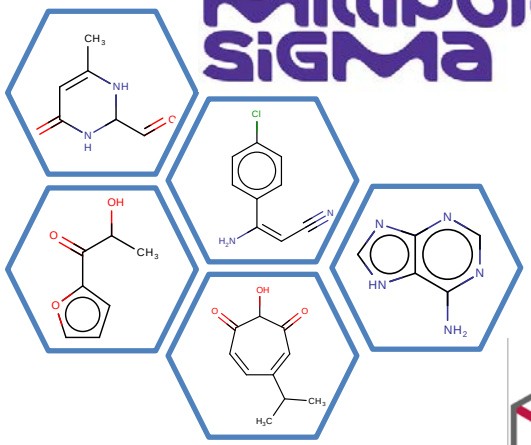


annotate them with computed properties important for CADD

SAVI components

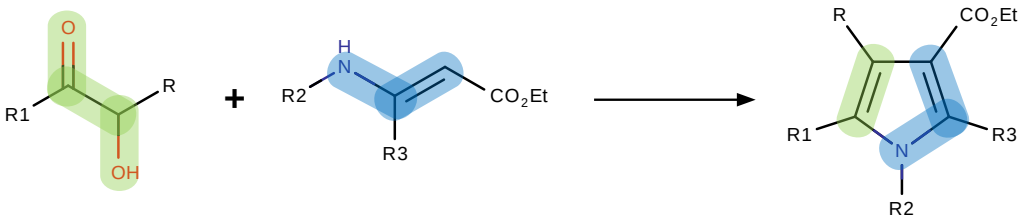
Building Blocks

MilliporeSigma



Transforms

Lhasa Ltd (UK), LHASA LLC (US)



LHASA transforms written in CHMTRN/PATRAN

SAVI

Chemoinformatics engine

CACTVS

Xemistry chemoinformatics

Reaction enumeration tools & libraries based on them

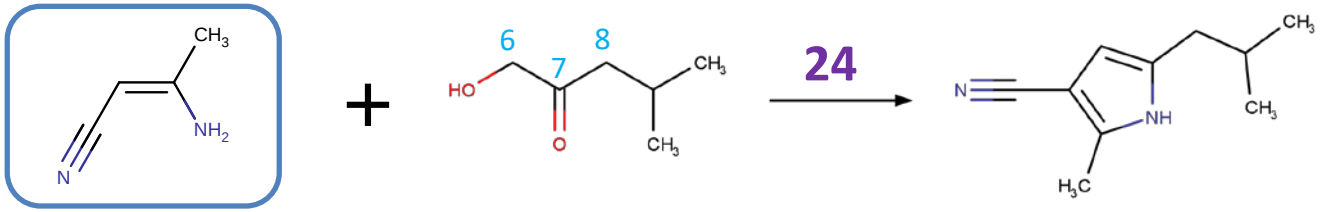
- REAL database – Enamine
  - 145k building blocks X 120 transforms
  - REAL Space Navigator: BioSolveIT
  - http://www.enamine.net/index.php?option=com_content&task=view&id=254%3f
- CHIPMUNK dataset – TU Dortmund
  - Hartenfeller[1], Reaxys and multicomponent reactions
  - http://www.ccb.tu-dortmund.de/koch
- Scaffold design – ChemPass
  - https://www.chempassltd.com/synspace-general-scaffold-design
- Reactor – ChemAxon
  - https://chemaxon.com/products/reactor
- Other in house tools in industry
  - Proximal Lilly Collection
  - Pfizer Global Virtual Library (PGVL)
  - Boehringer Ingelheim - BI CLAIM

1. Hartenfeller M, Eberle M, Meier P, et al (2011) A Collection of Robust Organic Synthesis Reactions for In Silico Molecule Design. J Chem Inf Model 51:3093–3098. doi: 10.1021/ci200379p

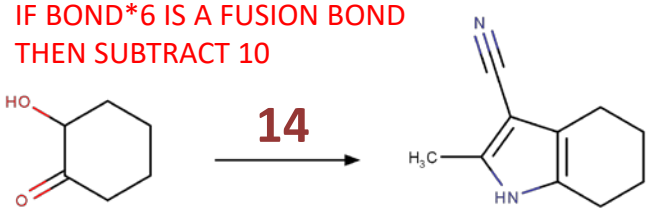
CHMTRN/PATRAN transform example

```
TRANSFORM 1039
NAME Feist Synthesis of Pyrroles

REFERENCES
J. Org. Chem. [53], 2084 (1988);
D.M. McKinnon, Can. J. Chem. [43], 2628 (1965).

                COOR
          _______/
          II      II
          II      II  ===>  AR-COCH-C(H)   +    (H)C-NHC=CHCO2R
          /\      /\              I                        I
   (AR) C   N     C              OH                       CH3


END*REFERENCES
...
TYPICAL*YIELD              POOR
RELIABILITY               FAIR
REPUTATION                FAIR
HOMOSELECTIVITY           GOOD
HETEROSELECTIVITY         FAIR
ORIENTATIONAL*SELECTIVITY      FAIR
CONDITION*FLEXIBILITY     FAIR
THERMODYNAMICS            FAIR
1D*PATTERN
N%C(-C[FGNOT=LEAVING])%C(-
C[FGS=ESTER,KETONE,NITRILE,AMIDE])%
C[HETS=0;FGNOT=LEAVING]%C(-
C[FGNOT=ACID,LEAVING,XWITHDRAWING])%@1
2D*PATTERN
      C  COR,CHO,CN,COOR,CON   H  C COR,CHO,COOR,CN,CON H  O
      |  |                        |  |  |                  |   "
  N%C%C%C%C%@1           =>   N-C=C                      C- -C
                 |                 |                    |   |
                 C                 H                   OH C
END*PATTERNS

DISCONNECTIVE
SUBGOAL*ALLOWED
BROKEN*BONDS BOND*5 BOND*8 BOND*1
...
IF CARBON ON ALPHA TO ATOM*1 OFFPATH THEN SAVE IT AS 1
KILL IF NOT ARYL ON SAVED*ATOM 1 AND:IF  &
  SAVED*ATOM 1 IS MULTIPLY BONDED
...Possible elimination.
...Would eliminate.
KILL IF WITHDRAWING BOND ON ATOM*6 OFFPATH
KILL IF WITHDRAWING BOND ON ATOM*8 OFFPATH
KILL IF AROMATIC BOND ON ATOM*6 OFFPATH
IF BOND*6 IS A FUSION BOND THEN SUBTRACT 10 AND*THEN KILL IF  &
  NOT IN A RING OF SIZE 6
SUBTRACT 10 IF SECONDARY*CENTRE ON SAVED*ATOM 1
SUBTRACT 10 IF TERTIARY*CENTRE ON SAVED*ATOM 1
...Steric hindrance.
IF FEWER THAN TWO HYDROGENS ON ATOM*3 THEN SUBTRACT 15
...Works best with acetoacetic ester.
SUBTRACT 10 IF NOT HYDROGEN ON ATOM*1
ADD 15 IF ARYL ON ALPHA TO ATOM*6 OFFPATH
ADD 10 IF ARYL ON ATOM*8 OFFPATH
...Higher yields.
IF THERE IS A FUNCTIONAL GROUP ON ATOM*5 THEN DESIGNATE  &
  IT AS PARTICIPATING
CONDITIONS SnCl4/25
ACTUAL*CONDITIONS 419: ZnCl2
....
  BREAK BOND*5
  BREAK BOND*8
  SINGLE BOND*6
  ATTACH A KETONE ON ATOM*7
  ATTACH AN ALCOHOL ON ATOM*6
  BOND*1 IS DEFINED*SYN TO BOND*4
...
```

CHMTRN/PATRAN transforms: smarter than SMARTS



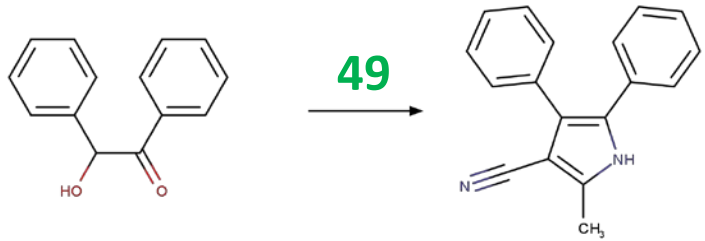
IF BOND*6 IS A FUSION BOND
THEN SUBTRACT 10



ADD 10 IF ARYL ON ATOM*8

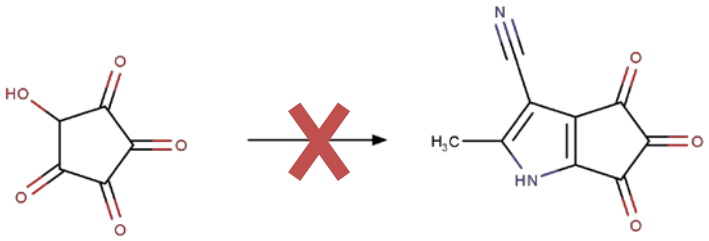
IF BOND*6 IS A FUSION BOND
THEN SUBTRACT 10
ADD 10 IF ARYL ON ATOM*8

ADD 10 IF ARYL ON ATOM*8
ADD 15 IF ARYL ON ALPHA TO ATOM*6

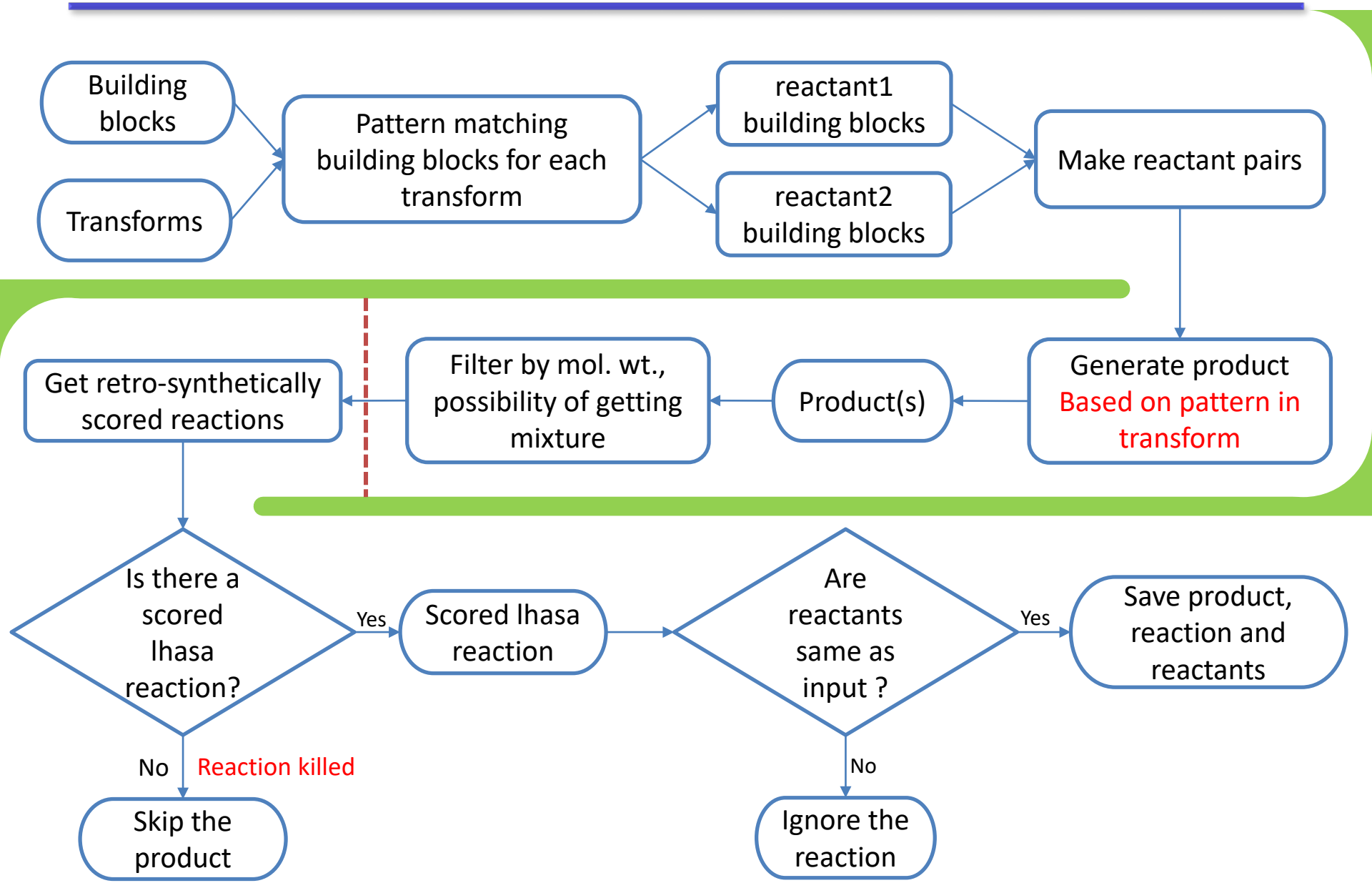


KILL IF WITHDRAWING BOND ON ATOM*6 OFFPATH

IF BOND*6 IS A FUSION BOND THEN SUBTRACT 10 AND*THEN
KILL IF & NOT IN A RING OF SIZE 6



Workflow


Building blocks
Transforms
Pattern matching building blocks for each transform
reactant1 building blocks
reactant2 building blocks
Make reactant pairs
Generate product Based on pattern in transform
Product(s)
Filter by mol. wt., possibility of getting mixture
Get retro-synthetically scored reactions
Is there a scored lhasa reaction?
Yes
Scored lhasa reaction
Are reactants same as input ?
Yes
Save product, reaction and reactants
No
Reaction killed
Skip the product
No
Ignore the reaction

First SAVI production run (2016)

351,298,493

14 transforms X 377,474 building blocks

305,775,262

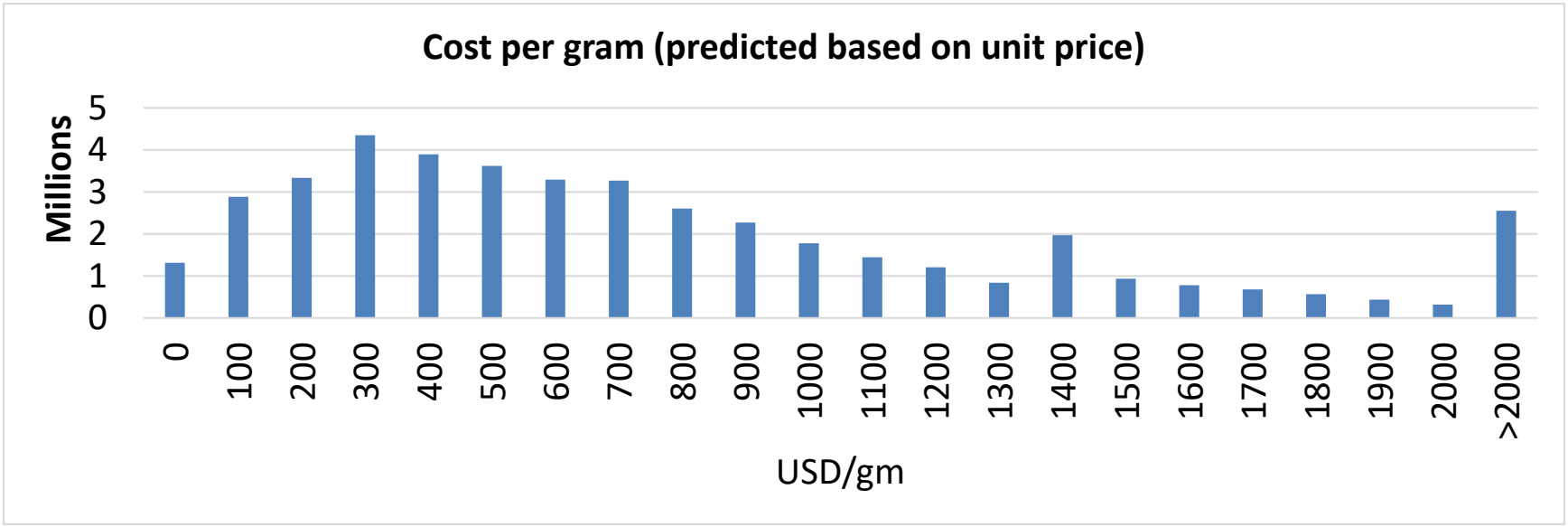
- CHMTRN/PATRAN transform logic disqualified 17% of reactions

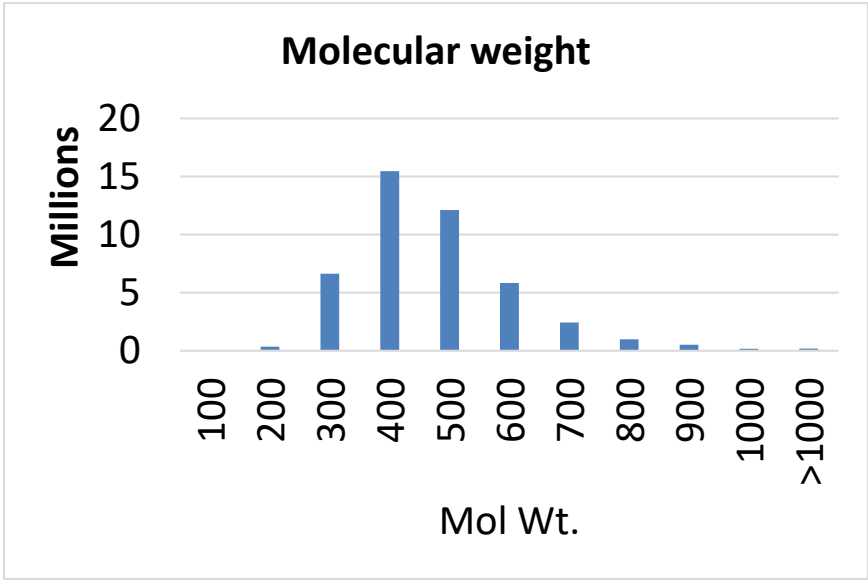
283,194,312

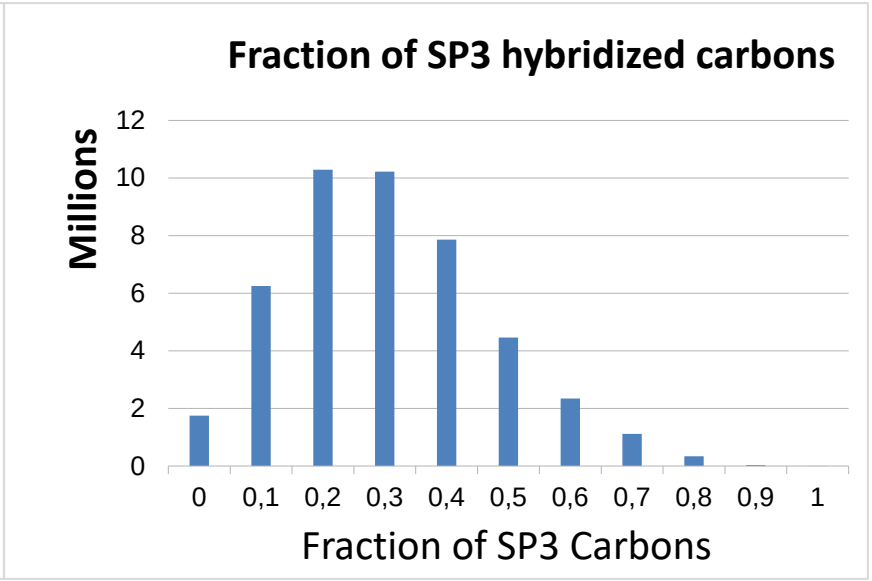
- Unique products

Downloadable at: https://cactus.nci.nih.gov/download/savi_download/

Product property distributions



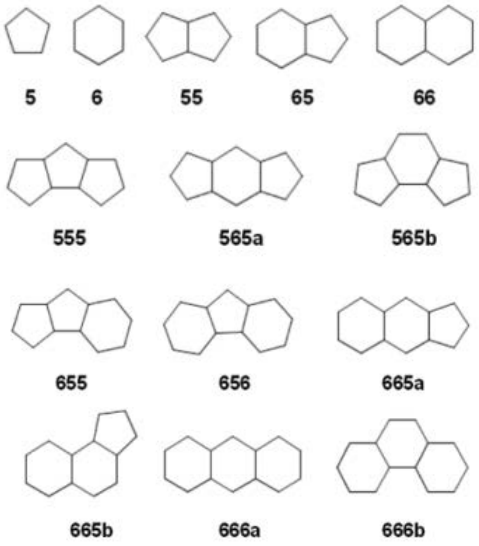


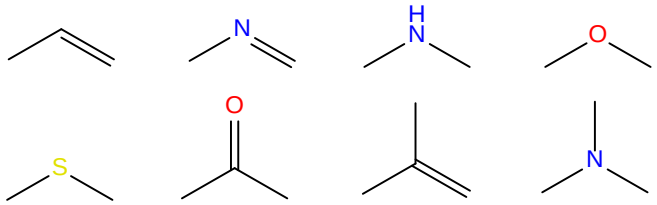


SAVI produces novel ring systems

Simple Aromatic (SA) "Ertl type"[1] novel ring/scaffold systems

SA: Systems up to tricyclic, ≥ 1 aromatic ring, with hetero atoms O, N, S containing the moieties:



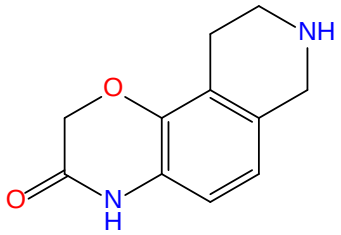


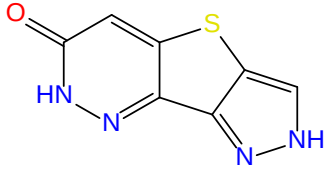
(1) Ertl *et al.*, *J. Med. Chem*. 49, 4568–4573 (2006)

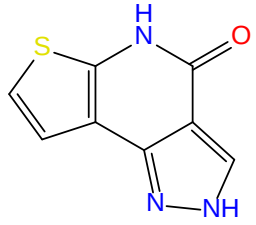
"Novel": not found in CADD Group aggregated database of 152 million structures

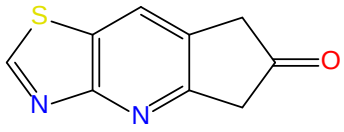
Total novel SA rings/scaffolds created: 999

Novelty confirmed in SciFinder, few examples:









Annotations for drug design

- "Drug design" properties:
  - RO5 violations, PAINS matches, fsp^3, log P…
  - All 275 Bruns & Watson rules[1] - combined as a demerit score
- Sigma catalog numbers and unit price of building blocks

> <SAVI_BUILDING_BLOCK_A_ORDER_LINK>
http://www.aldrichmarketselect.com/listBrowser.asp?structure_id=65139341

> <SAVI_BUILDING_BLOCK_A_COST_GRAM>
482.0

- Proposed reaction

> <SAVI_REACTION>
Paal-Knorr Pyrrole Synthesis

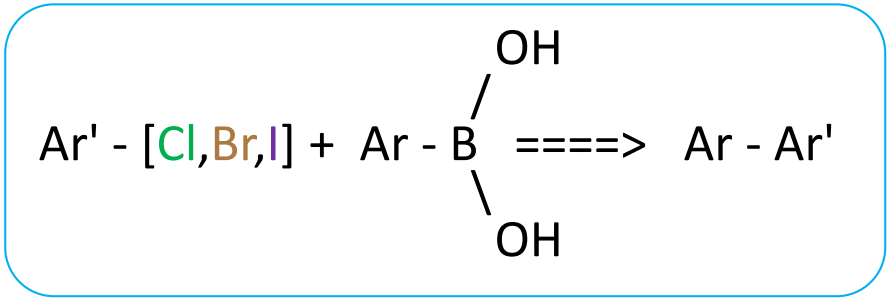
> <SAVI_LHASA_SCORE>
45

> <SAVI_PREDICTED_YIELD>
VERY*GOOD

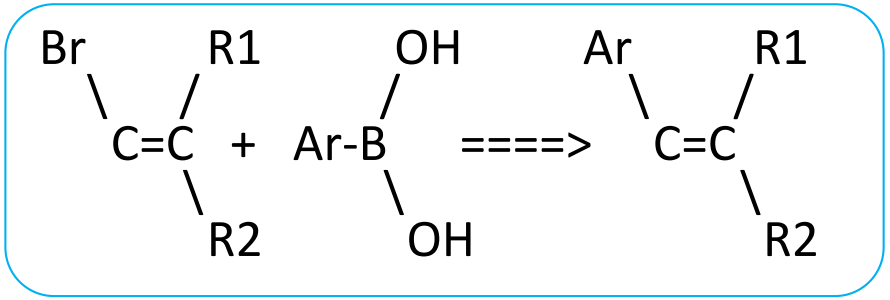
1. Bruns RF, Watson IA (2012) Rules for identifying potentially reactive or promiscuous compounds. J Med Chem 55:9763–9772. doi: 10.1021/jm301008n

Next version of SAVI

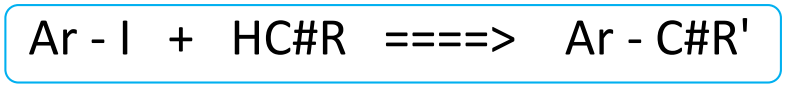
- ~600k building blocks
- ~20 chemistries (~50 transforms) implemented
  - 7 entirely new chemistries
  - example of new chemistry: Suzuki-Miyaura coupling



Suzuki coupling of Chloro, Bromo, Iodo

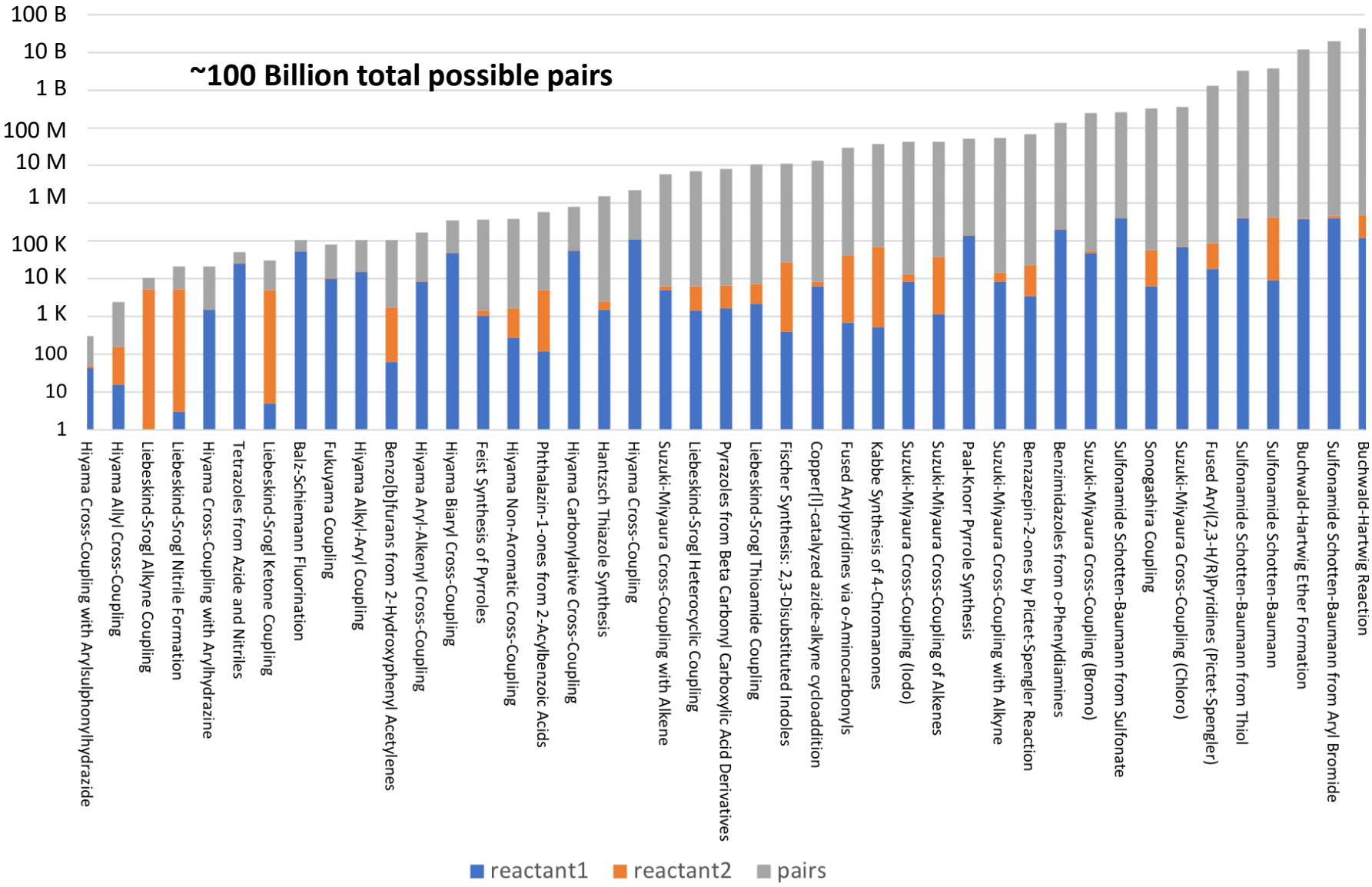


Suzuki coupling of Alkenes



Suzuki coupling with Alkynes

Possible reactant pairs: >> 1 billion



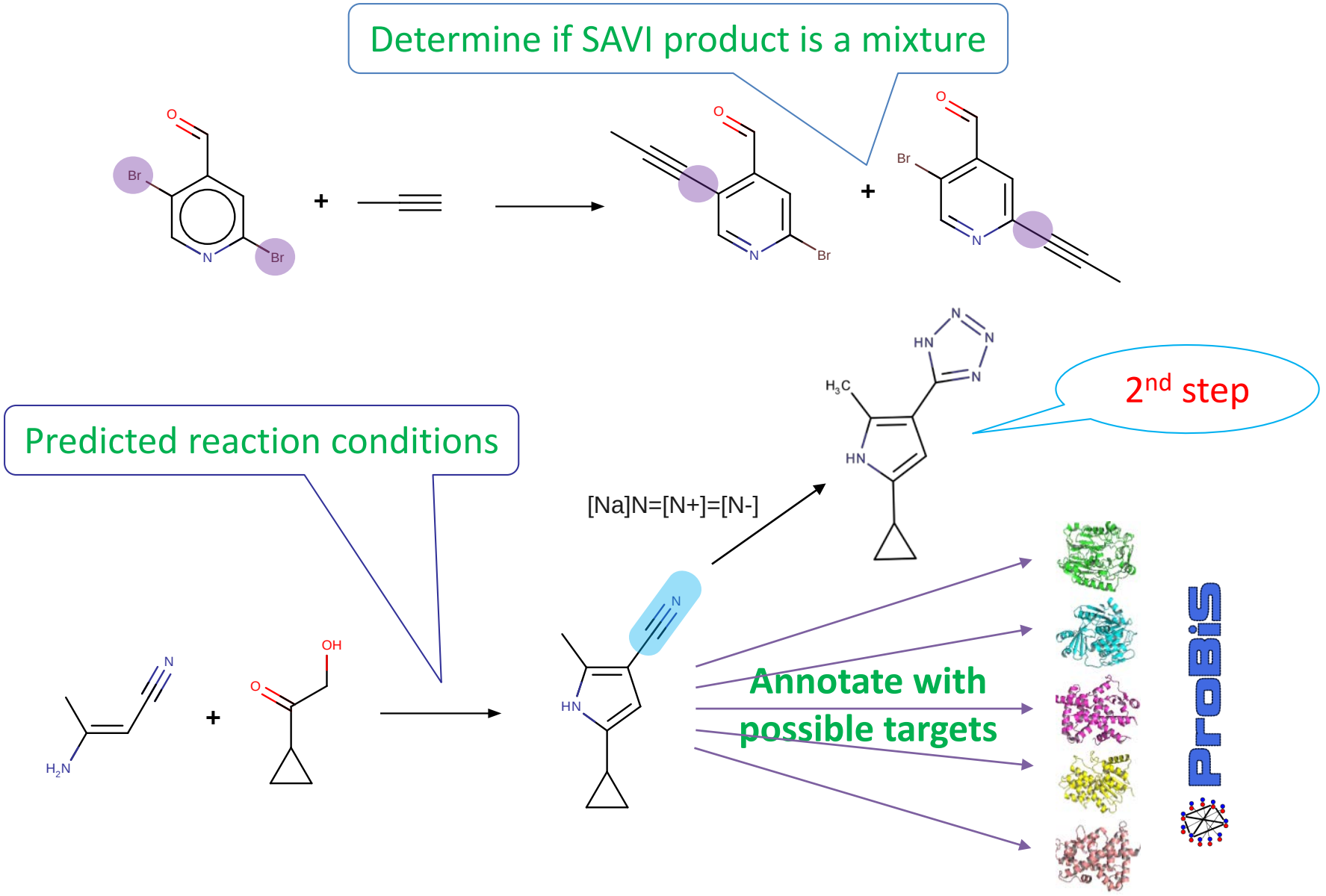
Challenges

- To generate comparable number of products for each transform
- Select 1 billion truly “interesting” compounds from 100 billions reactions
- What filter criteria? Possible:
  - Molecular weight, QED score[1] of products
  - Availability, reliability, purity and price of BBs
  - Strict scoring by SUBTRACT -> KILL
- How to handle possible mixtures of products
- Handling and representation of 1 billion reactions

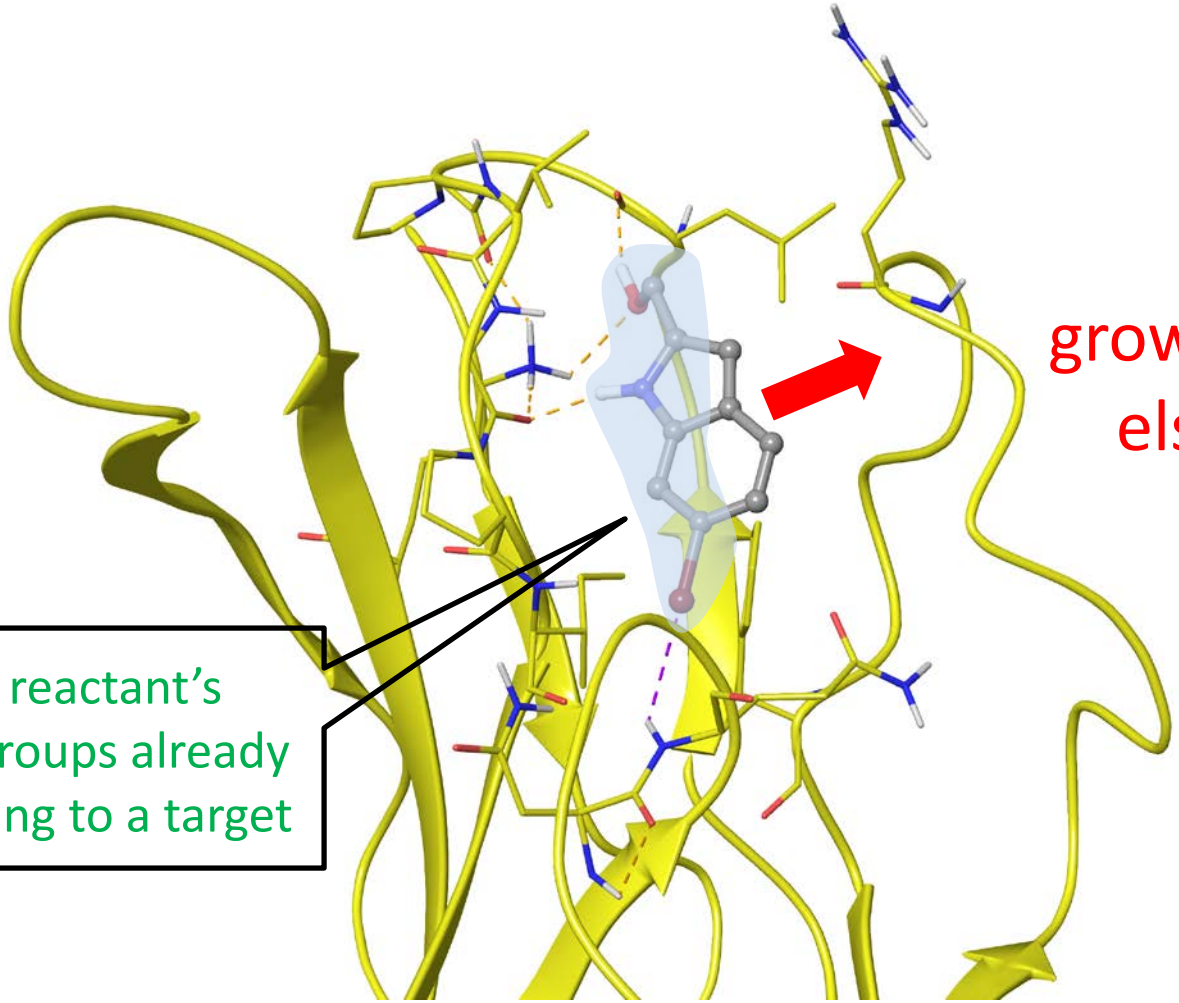
  Poster (P 62) - The need for comprehensive reaction handling in SAVI and beyond, Marc Nicklaus

1. Bickerton GR, Paolini G V, Besnard J, et al (2012) Quantifying the chemical beauty of drugs. Nat Chem 4:90–98. doi: 10.1038/nchem.1243

More and better annotations planned



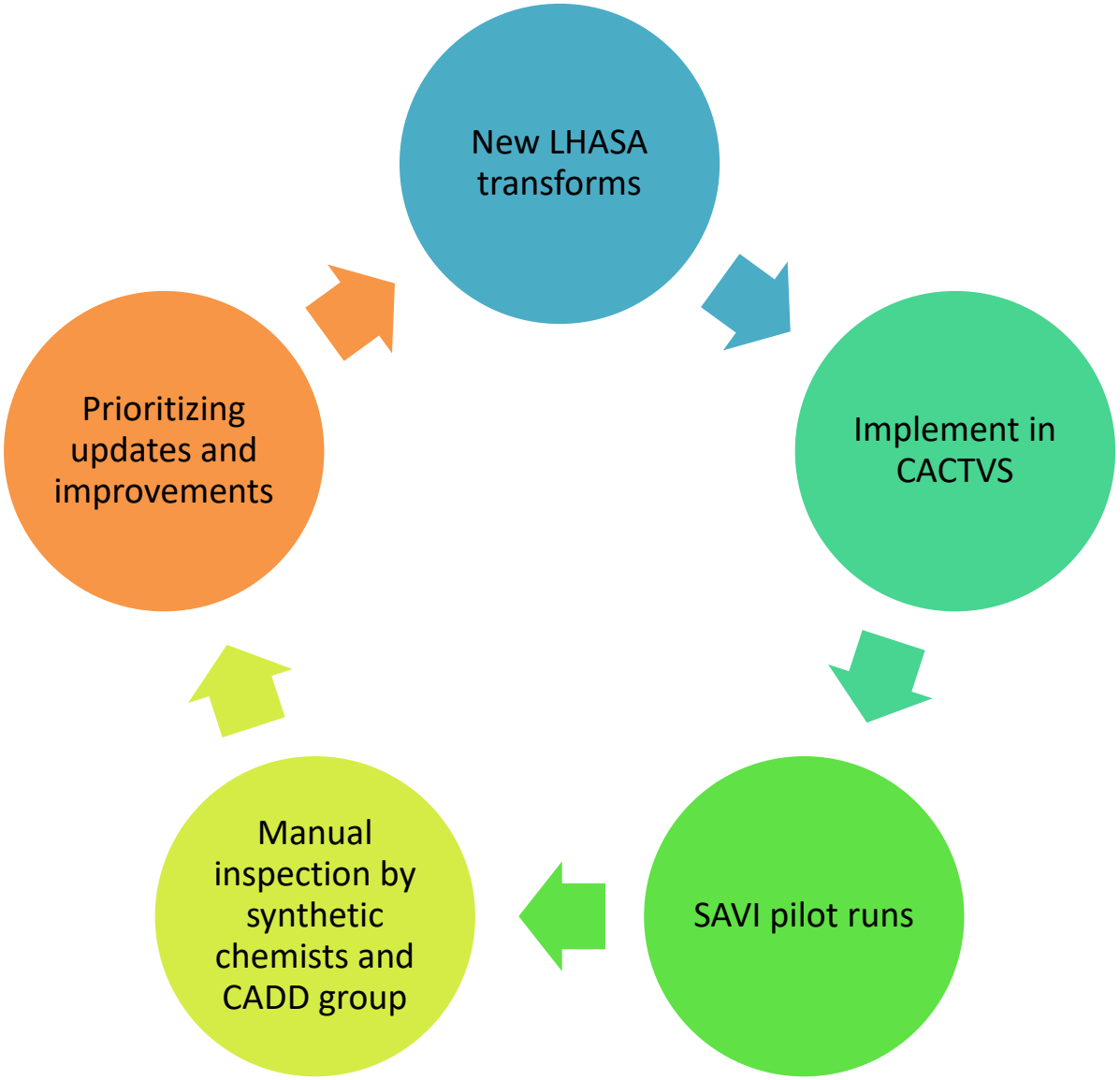
Ongoing work



grow molecule elsewhere

Preserve reactant's functional groups already known binding to a target

Iterative cycle for improved & new transforms



Future plans

- Stricter criteria or CADD property to qualify a reaction: less “haystack”
- Develop web-based GUI
- Investigate multi-step reactions
- Combine with robotic chemistry

Acknowledgements

CADD Group
- Devendra Dhaked
- Victorien Delannee

CCR
- Nadya Tarasova

Leidos
- Raul Cachau
- Megan Peach
- Jeff Saxe

Elsevier
- Scott Hutton
- Matt Clark

InfoChem
- Hans Kraut
- Josef Eiblmaier

Merck KGaA
- Friedrich Rippmann
- Mireille Krier
- Tim Knehans

MilliporeSigma
- Bret Daniel
- Chad Hurwitz
- Subir Ghorai
- Daniel Boesch
- Shashi Jasty

Lhasa Limited, UK
- Philip Judson
- Martin Ott

LHASA LLC, U.S.
- Alan Long
- Alex Sukharevsky

Novartis
- Peter Ertl

National Institute of Chemistry, Slovenia
- Dušanka Janežič
- Janez Konc