



SAVI: Billions of easily synthesizable compounds generated through expert-system type rules

Marc C. Nicklaus, Nadya I. Tarasova

CADD Group
Chemical Biology Laboratory
Center for Cancer Research
National Cancer Institute
National Institutes of Health
Frederick, MD, USA

Synthetic Biologics and Drug Discovery Group
Laboratory of Cancer Immunometabolism
Center for Cancer Research
National Cancer Institute
National Institutes of Health
Frederick, MD, USA

**A. Synthetically Accessible Virtual Inventory (SAVI) –
A Billion Compound Database for *In Silico* Screening**

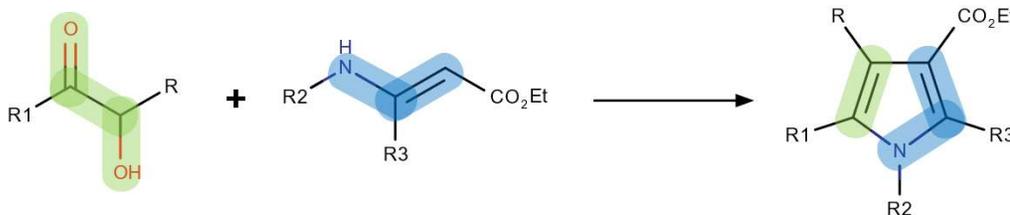
SAVI Components

Building Blocks



Yurii Moroz (Ukraine)

Transforms



LHASA transforms written in CHMTRN/PATRAN

Philip Judson (UK)



SAVI → *Product Generation*

Hitesh Patel (US)

**Chemoinformatics
engine**

Wolf D. Ihlenfeldt (Germany)

Xemistry chemoinformatics

Transform Encoding

- Created in the context of the LHASA project; started in late 1960s/early 1970s
- Language pair CHMTRN & PATRAN¹
- Expert-system type approach
- LHASA: retrosynthetic; SAVI: forward-synthetic
- Allows one to quickly add novel chemistry (no training set needed)
- Contains scoring system:
 - ADD statements
 - SUBTRACT statements
- Has KILL statements
- Original knowledgebase did not contain modern chemistry such as Suzuki coupling
- Have added new transforms
- Parser and execution engine developed for CACTVS

¹Judson *et al.*, *JCIM*, 2020 60 (7), 3336-3341.
DOI: 10.1021/acs.jcim.0c00448

```
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
...
```

Cutout from Transform 1039: Feist Synthesis of Pyrroles

Transforms Used

We focused on coupling reactions, including a number of ring-forming reactions

ID	Name
1031	Paal-Knorr Pyrrole Synthesis
1039	Feist Synthesis of Pyrroles
1171	Hantzsch Thiazole Synthesis
1391	Allene 2+2 Cycloaddition
1439	Pyrazoles from Beta Carbonyl Carboxylic Acid Derivatives
2201	Fused Arylpyridines via o-Aminocarbonyls
2218	Tetrazoles from Azide and Nitriles
2230	Phthalazin-1-ones from 2-Acylbenzoic Acids
2238	Fused Aryl(2,3-H/R)Pyridines (Pictet-Spengler)
2267	Sonogashira Coupling
2269	Kabbe Synthesis of 4-Chromanones
2630	Benzazepin-2-ones by Pictet-Spengler Reaction
2684	Benzo[b]furans from 2-Hydroxyphenyl Acetylenes

Transforms from the old knowledgebase

Newly written transforms

ID	Name
2875	Copper[I]-catalyzed azide-alkyne cycloaddition
6003	Buchwald-Hartwig Ether Formation
6004	Suzuki-Miyaura Cross-Coupling (Bromo)
6005	Suzuki-Miyaura Cross-Coupling (Iodo)
6006	Suzuki-Miyaura Cross-Coupling (Chloro)
6008	Suzuki-Miyaura Cross-Coupling with Alkene
6009	Suzuki-Miyaura Cross-Coupling of Alkenes
6013	Hiyama Aryl-Alkenyl Cross-Coupling
6014	Hiyama Non-Aromatic Cross-Coupling
6015	Hiyama Allyl Cross-Coupling
6016	Hiyama Carbonylative Cross-Coupling
6017	Hiyama Cross-Coupling with Arylhydrazine
6022	Liebeskind-Srogl Thioamide Coupling
6024	Liebeskind-Srogl Nitrile Formation
6025	Liebeskind-Srogl Heterocyclic Coupling
6026	Sulfonamide Schotten-Baumann
6027	Sulfonamide Schotten-Baumann from Sulfonate
6028	Sulfonamide Schotten-Baumann from Thiol
6029	Sulfonamide Schotten-Baumann from Aryl Bromide
6031	Mitsunobu Reaction
6032	Mitsunobu carbon-carbon bond formation
6033	Mitsunobu SN2' Reaction
6034	Mitsunobu Imide Reaction
6035	Mitsunobu Aryl Ether Formation
6036	Mitsunobu Sulfonamide Reaction
6038	Ester or Amide or Thiolester Formation
6039	Williamson Ether Synthesis
6041	Buchwald-Hartwig Reaction
6043	Buchwald-Hartwig Reaction
7005	Benzimidazoles from o-Phenylenediamines
7009	Acylsulfonamide from Sulfonamide and Carboxylic Acid
7013	Benzimidazoles from o-Phenylenediamines and Aldehydes
7014	Benzimidazoles from o-Phenylenediamines and Aldehydes
7015	Sulfonamide from sulfonic acid and amine
7017	Sulfonamide alkylation with a cyclic ether
7018	Sulfonamide acylation
7019	Wittig Reaction
7020	Wittig via Methoxy-Ylide
7021	Horner-Wadsworth-Emmons Olefination
7022	Chan-Lam coupling

SAVI-2020 Product Counts

All reactions: single-step syntheses (A + B -> C)

Class	SAVI products	Unique	Percentage
Plus	1,094,782,440	976,051,945	62.61%
Neg0	609,262	579,532	0.03%
Neg10	54,775,204	48,036,148	3.13%
Neg20	82,180,372	80,366,188	4.7%
Neg30	516,116,725	457,508,945	29.52%
Total	1,748,464,003	1,526,316,392	

Plus class: may not have encountered any SUBTRACT.
KILL rate: 51%

Non-unique structures: due to multiple proposed routes

SAVI-2020 internally stored in PostgreSQL database

SAVI – Publicly Downloadable Dataset

www.nature.com/scientificdata

SCIENTIFIC DATA

 Check for updates

OPEN

DATA DESCRIPTOR

SAVI, *in silico* generation of billions of easily synthesizable compounds through expert-system type rules

Hitesh Patel ¹, Wolf-Dietrich Ihlenfeldt ², Philip N. Judson ³, Yurii S. Moroz ⁴, Yuri Pevzner ^{1,5}, Megan L. Peach ⁶, Victorien Delannée ¹, Nadya I. Tarasova ⁷ & Marc C. Nicklaus ¹ 

Nature Sci. Data | (2020) 7:384 | <https://doi.org/10.1038/s41597-020-00727-4>

See also <https://doi.org/10.26434/chemrxiv.12185559.v1>

Data set: https://cactus.nci.nih.gov/download/savi_download/
<https://doi.org/10.35115/37n9-5738>

SAVI Downloads

- SAVI-2020 became available in April 2020
- Available in SDF and as SMILES table
- Each format consists of 55 tar files with 200 gzipped files each
- SDFs in total: 4.4TB; SMILES tables: 1.1TB
- Download accesses through October 2020: 59,456
- Counting 55 download clicks as one full SAVI download:
~5 full downloads/day, ~1,000 downloads in total

SAVI-2020 Syntheses Success Rates

- About 150 SAVI-2020 products synthesized in NIH drug development projects so far
- Anti-cancer and antiviral projects
- Some in-house, some by Enamine
- All from the Plus subset
- Success rate: 97%

Multistep Reactions

➤ First foray into two-step reactions:

Products by transform 2875 (“azide-alkyne Click chemistry”): 1.2M.
Checked susceptibility of those with all 53 transforms and 152k BBs for reaction with Enamine BBs:

➔ Possible reactant pairs: >50 billion

Re-using all SAVI-2020 products as new building blocks yields >1 trillion predicted actually accepted products

We may need “SAVI à la carte” to selectively expand into parts of these two-step spaces

Plans

- GUI for fast searches, on public server
- Expansion of transforms
- Broadening of building block set, especially to better feed “starved” transforms
- Easier writing of new transforms in CHMTRN/PATRAN
- More modern way of applying transforms for product generation

SLICE

SLICE (Smarts and Logic In ChEmistry) – XML format (Victorien Delannée)

```
<transforms>
  <reactions>
    <reaction subtype="1">
      <reactants>
        <reactant id="left">
          <smarts>[C,S:2][N:1]=[N;+1:3]=[N;-1:4]</smarts>
          <logic>
            if bond between atom 1 and 2 is in ring then kill
            if molecule has an amine1 or an azide or an acetylen or a nitrile anywhere then kill
            if atom 2 is a carbon {
              if atom alpha to atom 2 has more than one hetero then kill
              if atom 2 is not aromatic and if atom 2 has not a double bond
            }
            if atom 2 is a sulfur {
              #check sulfur
              if molecule has not any carbon alpha to atom 2 offpath or if molecule has not two oxygen on atom 2 then kill
              #check sochk
              foreach oxygen atom alpha to atom 2 defined as oxygenAtom {
                if bond between atom 2 and oxygenAtom is double then kill
              }
            }
          </logic>
        </reactant>
        <reactant id="right">
          <smarts>[C:5]#[C:6]</smarts>
          <logic>
            ...
          </logic>
        </reactant>
      </reactants>
      <products>
        <product id="1">
          <smarts>[n:1]1[[c:6]=[c:5][n:4]=[n:3]1][C,S:2]</smarts>
          <logic>
            ...
          </logic>
        </product>
      </products>
      <conditions properties=""/>
      <actualConditions properties=""/>
    </reaction>
  </reactions>
</transform>
</transforms>
```

Logic

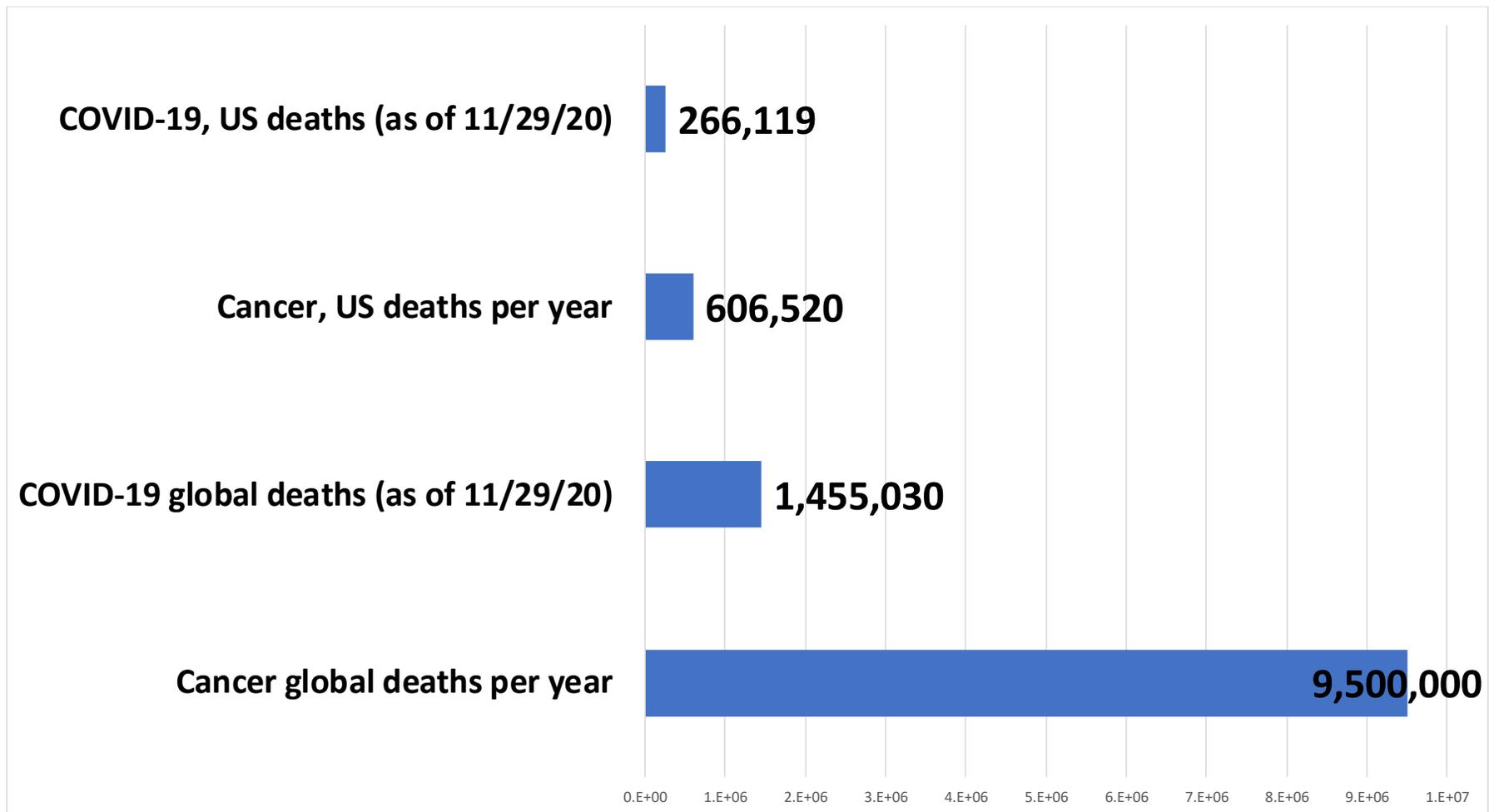
<https://doi.org/10.1021/scimeetings.0c00191>

B. Mining SAVI for potential drugs

Major purposes of large chemical databases:

- **Expanding our ability to address unmet therapeutic needs.**
- **Making drug discovery more effective**

Cancer remains one of the leading causes of death



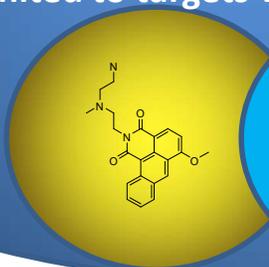
The Problem of Intractable Targets

80-90% of human proteins cannot be targeted by established modalities

Universe of potential targets

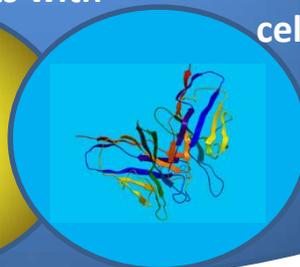
Small molecules

Are believed to be limited to targets with deep clefts, 10-15%



Biologics

Limited to targets outside cells, ≈10%



Is it REALLY impossible to target cleft-less proteins with small molecules?

Or

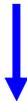
Have we been searching in a wrong galaxy of the chemical universe?

Use of SAVI for identification of ligands for cancer and inflammation-related molecular targets

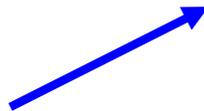
Virtual screens
using ICM-Pro (Molsoft) on
NIH supercomputer cluster
Biowulf



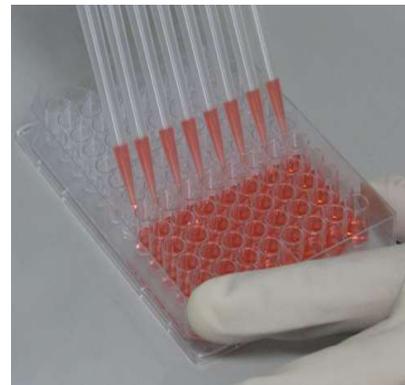
2 rounds of docking
followed by manual
verification



Chemical
synthesis of hits



Biophysical, biochemical and cell-
based assays



Screening targets

SAVI was tested in screens for challenging, non-druggable targets:

16 targets,

38 x-ray structures

Total compounds ordered: 70

Successful synthesis: 68

Typical current workflow:

- 1. Verification of x-ray structure suitability for virtual screens.**
- 2. Docking screen of SAVI diversity set of 2,955,416 compounds**
- 3. Synthesis and testing of hits**
- 4. Fragment-based and 2D-similarity searching of entire database for analogs of identified hits**
- 5. Docking of analogs**
- 6. Structure optimization using traditional medicinal chemistry**

Even for the most challenging targets, we were able to identify compounds with binding affinity in at least micromolar range.

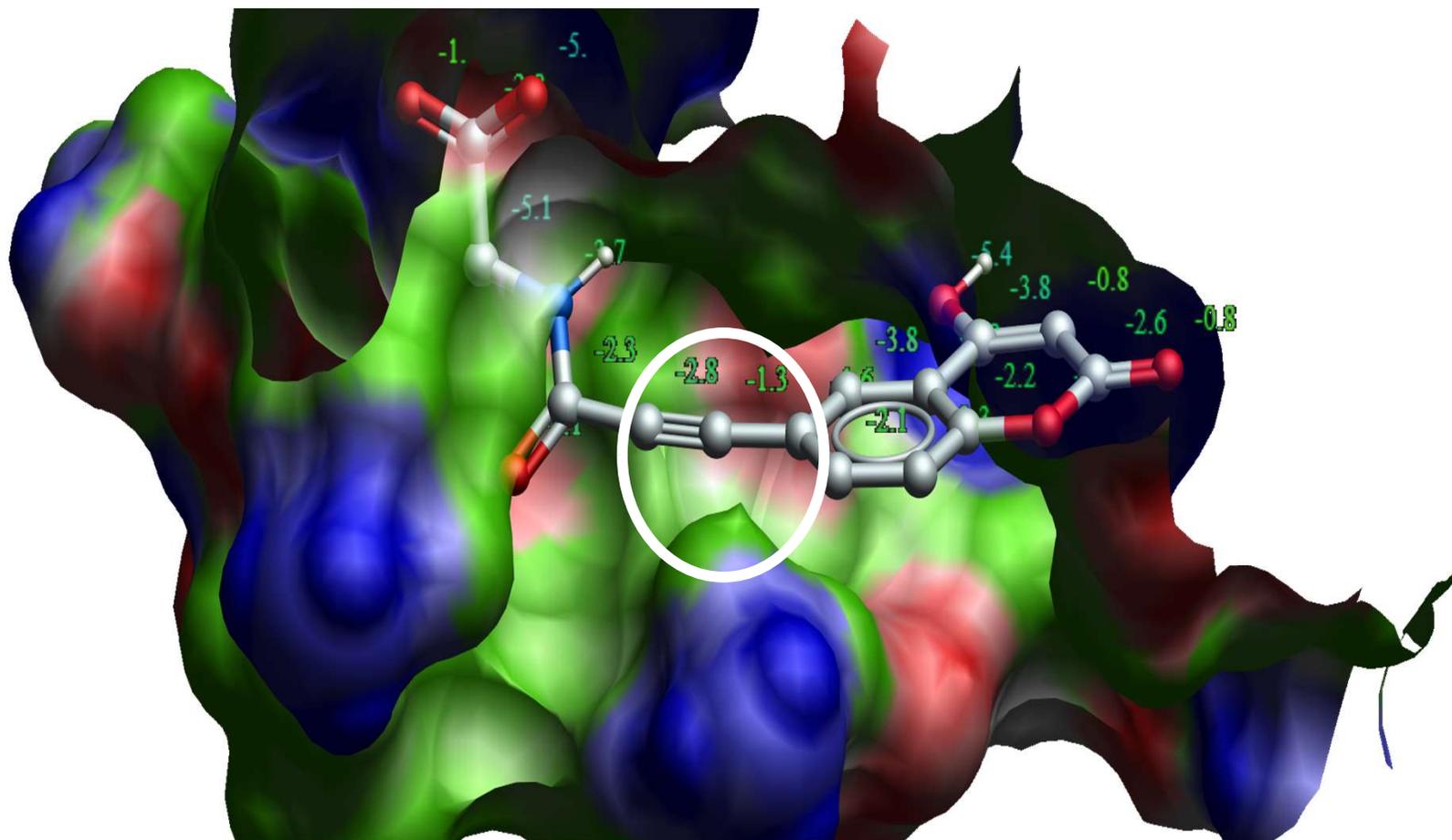
SAVI allowed for identification of inhibitors with nanomolar affinity for two targets widely considered non-druggable.

Although SAVI and REAL are made from the same blocks, they appear to perform differently in screens

Target	SAVI diversity set of 2,955,416 compounds		Enamine diversity set of 2,725,753 compounds	
	Number of hits	Best Score	Number of hits	Best Score
A	4346	-55.8	1480	-46.5
B	2008	-46.8	541	-43.2
C	2255	-50.1	6325	-50.4
D	326	-41.8	726	-43.0

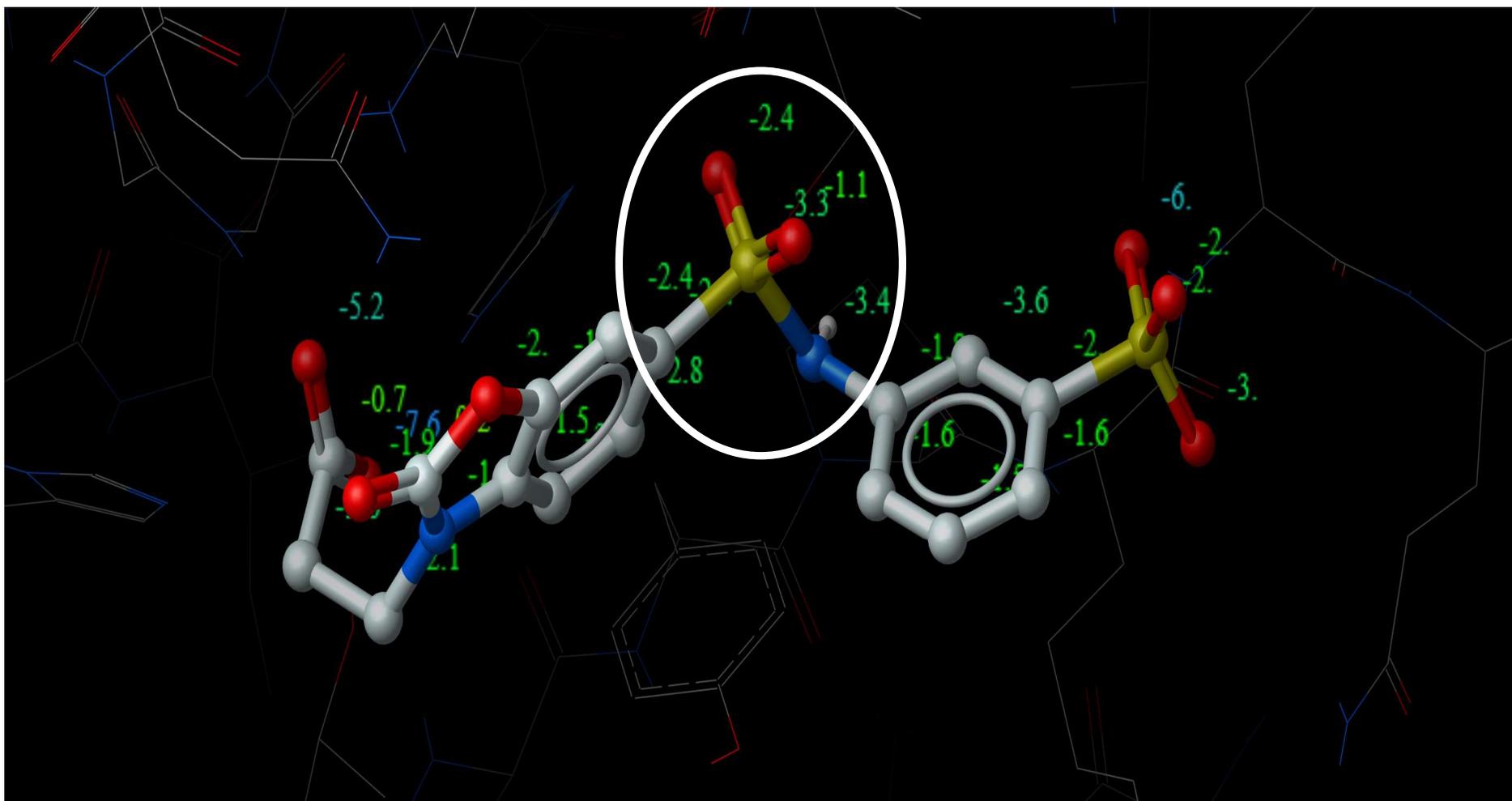
The transform defines the relative positioning of the fragments in the resulting compound

Example: Sonogashira coupling (transform 2267) creates rigid scaffold with unique shape:



Part of a molecule generated by the transform can also contribute directly to the binding

Sulfonamide group in this example contributes significantly to binding energy:



In some cases, however, we are unable to fill binding pockets completely and effectively

Most likely reason: limited diversity of the libraries.

Ways of improving library diversity:

- 1. More building blocks**
- 2. More transforms**
- 3. More synthesis steps**

Addition of new chemistries can increase size and diversity of the databases

LETTER

<https://doi.org/10.1038/s41586-019-1589-1>

Modular click chemistry libraries for functional screens using a diazotizing reagent

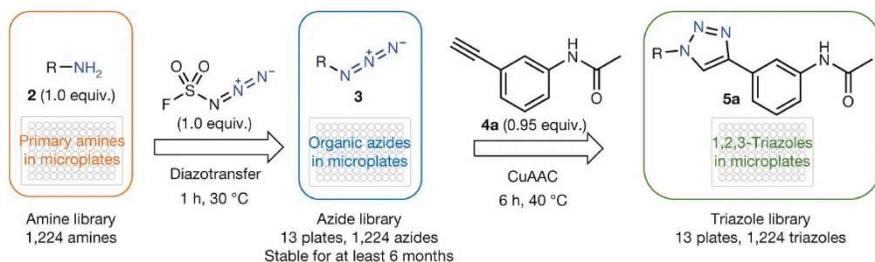
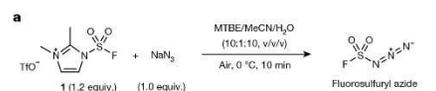
Genyi Meng^{1,2}, Taijie Guo^{1,2}, Tiancheng Ma^{1,2}, Jiong Zhang¹, Yucheng Shen¹, Karl Barry Sharpless^{1*} & Jiajia Dong^{1*}

Click chemistry is a concept in which modular synthesis is used to rapidly find new molecules with desirable properties¹. Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) triazole annulation and sulfur(vI) fluoride exchange (SuFEx) catalysis are widely regarded as click reactions²⁻⁴, providing rapid access to their products in yields approaching 100% while being largely orthogonal to other reactions. However, in the case of CuAAC reactions, the availability of azide reagents is limited owing to their potential toxicity and the risk of explosion involved in their preparation. Here we report another reaction to add to the click reaction family: the formation of azides from primary amines, one of the most abundant functional groups⁵. The reaction uses just one equivalent of a simple diazotizing species, fluorosulfonyl azide⁶⁻¹¹ (FSO₂N₃), and enables the preparation of over 1,200 azides on 96-well plates in a safe and practical manner. This reliable transformation is a powerful tool for the CuAAC triazole annulation, the most widely used click reaction at present. This method greatly expands the number of accessible azides and 1,2,3-triazoles and, given the ubiquity of the CuAAC reaction, it should find application in organic synthesis, medicinal chemistry, chemical biology and materials science.

Non-aromatic organic azides are typically prepared by substitution

This solution of FSO₂N₃ in methyl *tert*-butyl ether (MTBE)/MeCN (Fig. 1a) is used directly for the diazotransfer reaction. A solvent system of dimethyl sulfoxide (DMSO)/MTBE/H₂O or dimethylformamide (DMF)/MTBE/H₂O was found to be optimal for this reaction, and DMF was preferred for product purification. Unlike the previously reported diazotransfer reactions with TN₃, the related processes based on the parent reagent FSO₂N₃ do not benefit from the addition of a metal catalyst²². The quantitative conversion of amlodipine (2a; Fig. 1b) or pazufloxacin mesylate (2b; Fig. 1c) was complete at room temperature after 5 min (see Supplementary Information 1, section 8 for details of the reactions).

The high reactivity of FSO₂N₃ was further demonstrated by the diazotransfer of trifluoromethanesulfonamide (TFNH₂) to TFN₃ under mild basic biphasic reaction conditions (Supplementary Information 1).



nature
COMMUNICATIONS

ARTICLE

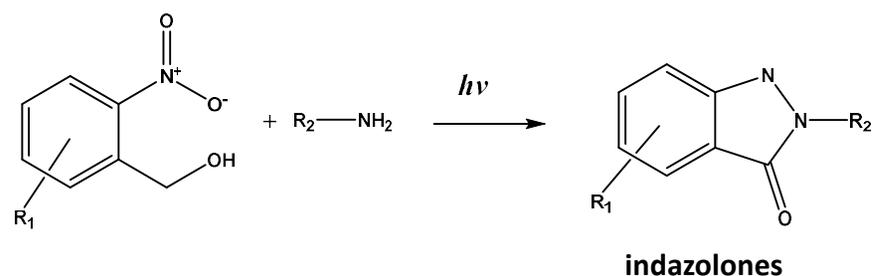
<https://doi.org/10.1038/s41467-020-19274-y>

OPEN

Check for updates

Light-induced primary amines and *o*-nitrobenzyl alcohols cyclization as a versatile photoclick reaction for modular conjugation

An-Di Guo^{1,2,4}, Dan Wei^{1,2,4}, Hui-Jun Nie^{1,4}, Hao Hu³, Chengyuan Peng³, Shao-Tong Li^{1,2}, Ke-Nian Yan^{1,2}, Bin-Shan Zhou¹, Lei Feng^{1,2}, Chao Fang¹, Minjia Tan³, Ruimin Huang³ & Xiao-Hua Chen^{1,2}



Acknowledgements

CADD Group

- **Hitesh Patel**
- Devendra Dhaked
- **Victorien Delannée**
- **Yuri Pevzner** (now: AbbVie)

Xemistry, GmbH

- **Wolf-Dietrich Ihlenfeldt**

Lhasa Limited, UK

- **Philip Judson**
- Martin Ott

Leidos (Frederick Natl Lab Canc Res)

- Raul Cachau
- Megan Peach
- Jeff Saxe

SBDD Group

- **Pedro Andrade Bonilla**
- **Karen Stefanisko**
- **Lyuba Khavruskii**

NCATS

- Alexey Zakharov

Enamine/Chemspace

- **Yurii Moroz**

Elsevier

- Scott Hutton (then: ChemNavigator)
- Matt Clark

InfoChem

- Hans Kraut
- Josef Eiblmaier

MilliporeSigma

- Bret Daniel
- Chad Hurwitz
- Subir Ghorai
- Daniel Boesch

LHASA LLC, U.S.

- Alan Long
- Alex Sukharevsky

Novartis

- Peter Ertl

Oncolinx

- Sourav Sinha