Large scale storage stability analysis of molecules in the NCATS SMR *(formerly MLSMR)*

Christian Laggner,[1] Yuda Shayo,[1] Chayling Hendarto,[1] Charles R. Johnson,[1] Carson R. Loomis[2]

[1)] Evotec (US) Inc., 385 Oyster Point Blvd. Suite 1, South San Francisco, CA 94080, USA; [2)] National Center for Advancing Translational Sciences, 6701 Democracy Blvd., Bethesda, MD 20892, USA

The NCATS SMR at Evotec

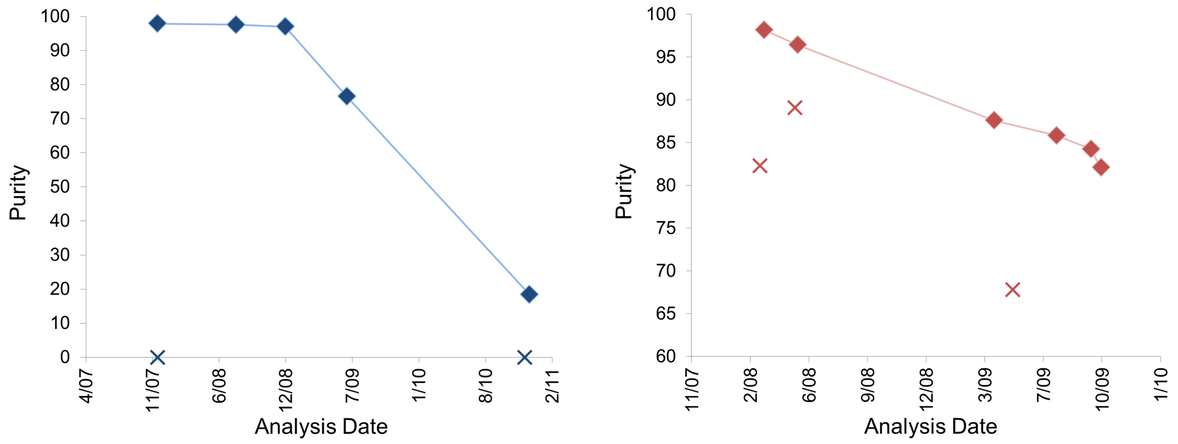
In 2004, NIH founded the Molecular Libraries Small Molecule Repository (MLSMR) to provide the Molecular Libraries Probe Production Centers Network (MLPCN) with a collection of diverse small molecules for use in high-throughput biochemical probe and drug discovery screening.[1] After completion of the initial 10 year Molecular Libraries Program (MLP), the collection transitioned to the National Center for Advancing Translational Sciences (NCATS), and was renamed to NCATS SMR (Small Molecule Resource).

Since its inception, the collection of almost 390,000 unique and diverse compounds has been managed in the same facility that is now part of Evotec.[2] One of defining criteria for the MLSMR has been its stringent QC requirements: new compounds are inspected for solubility in DMSO, and have to pass a purity cutoff of >90% by LCMS. The storage conditions are mild: nitrogen atmosphere, -20°C. Working solutions, 10mM in DMSO, undergo periodic thawing for distribution to screening centers and QC analysis is performed to verify integrity. Thus, we have accumulated over 3 million QC data points over the past 10 years.

We previously reported analysis of samples that failed QC after storage.[3] Here, we applied a systematic cheminformatic approach in order to identify scaffolds that are common among such compounds.

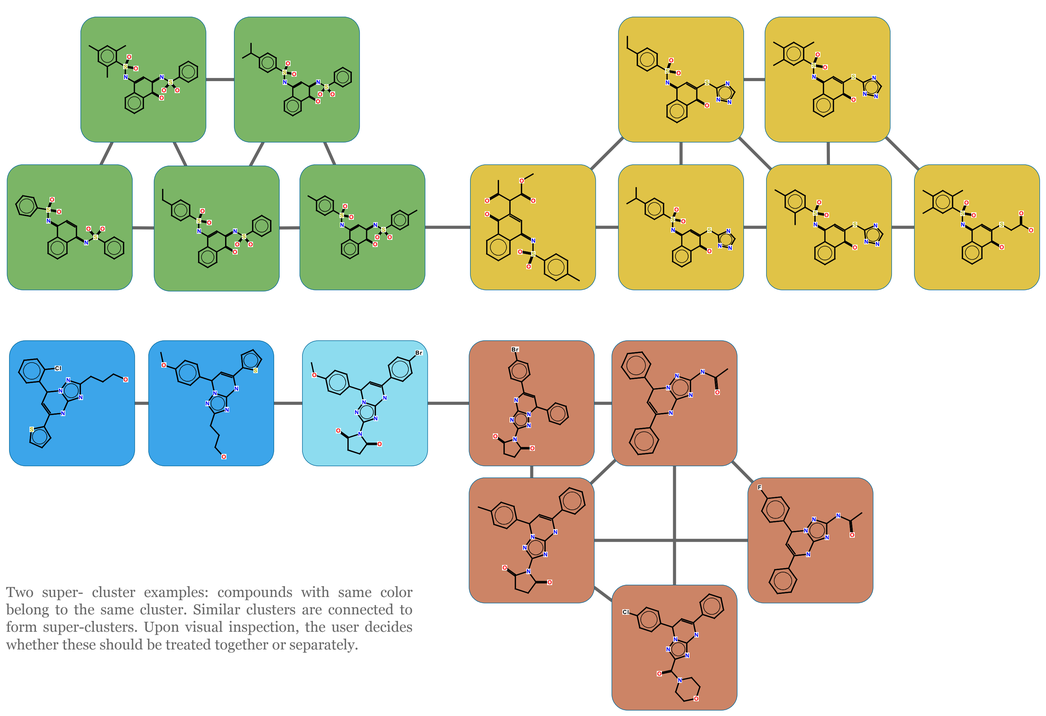
A cheminformatic workflow to identify clusters of unstable compounds

- Filter LCMS purity data from database for compounds that were initially >90% pure, yet later fell below 85% purity
- Conservative approach to reduce false negatives:
  - For every sample, take the highest purity value among multiple measurements on a given day
  - Higher and newer purity values invalidate any previous lower values:



Purity by LCMS over several years for two separate samples. X marks values that were ignored.

- A two-level clustering approach helps overcome the differences between computer- and human-perceived similarity:
  - Calculate pairwise chemical similarity (Dice coefficient, ChemAxon ECFP6 fingerprints), with a cutoff of 0.5
  - Use Markov Cluster Algorithm (MCL)[4] to define clusters of chemically similar compounds (tight similarity)
  - Define “super-clusters” using the strongest pairwise similarity between different clusters (broader similarity)

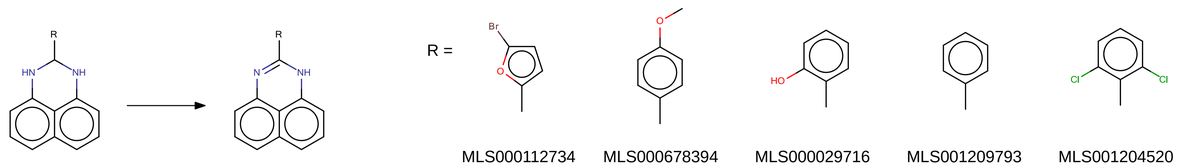


Two super- cluster examples: compounds with same color belong to the same cluster. Similar clusters are connected to form super-clusters. Upon visual inspection, the user decides whether these should be treated together or separately.

Identification of potential reaction products within chemical clusters

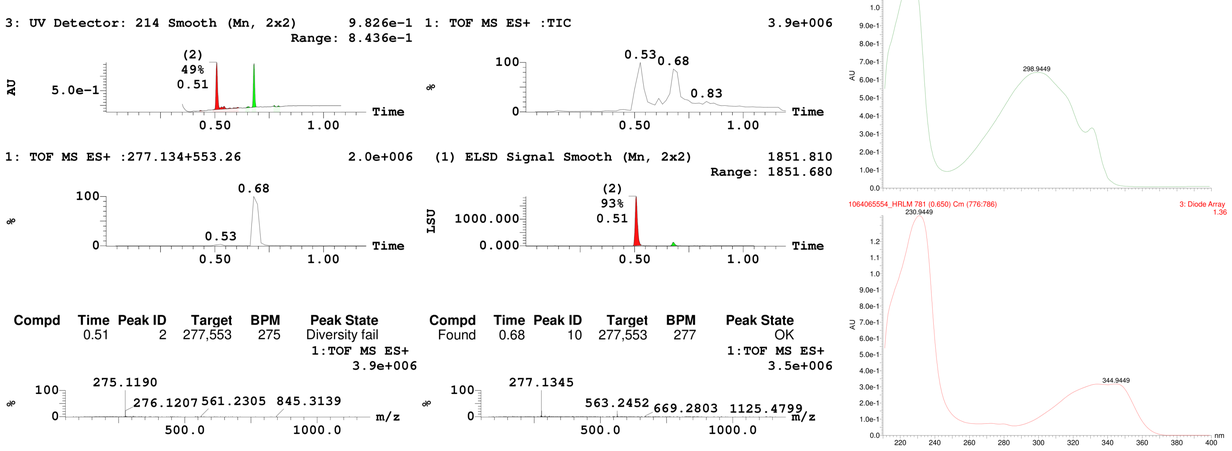
- Clusters and super-clusters are visually inspected to identify compounds that share a common chemical functionality
- Potential reaction products that fit the low resolution MS peaks of the products are identified
- A common reaction mechanism is proposed for the compounds within a given cluster
- High-resolution mass spectrometry (HRMS) is used on a list of selected compounds to verify the proposed molecular formula

Example: Oxidation of 2,3-dihydro-1*H*-perimidines



MLS000112734 MLS000678394 MLS000029716 MLS001209793 MLS001204520

We identified a cluster of 2,3-dihydro-1*H*-perimidines that undergo oxidation (-2H), presumably to the 1*H*-perimidines. Of 28 cluster members, 4 were selected for HRMS. Additionally, the phenyl analog MLS001209793 which had only been analyzed once before was included in the analysis and was found to have undergone the same transformation.



UPLC traces and high resolution mass spectra for MLS000678394:
Oxidized product (red) RT 0.51 min, 93% ELSD, 49% UV214 nm, M+1: calc. 275.1184, found 275.1190.
Original compound (green) RT 0.68 min, 7% ELS, 38% UV214 nm, M+1 : calc. 277.1341, found 277.1345.

UV spectrum of oxidized product (top panel) from starting material MLS000678394 (bottom panel).

Software and Instrumentation

Software: KNIME 2.10/2.11 with ChemAxon’s JChem extension; Cytoscape 3.2 with chemViz2 app; MCL clustering software[4] from *http://micans.org/mcl/*

High Resolution UPLC-MS: Phenomenex Kinetex C18 column, 2.6 um, 100 Å, 2.1x50 mm; 0.1% formic acid + 1.5 mM ammonium formate either in water (mobile phase A) or in acetonitrile:water 9:1 (mobile phase B). Samples were dissolved in DMSO to 1 mM concentration and 1 uL was injected. Gradient elution 99.8% A to 2% A over 1 min for total of 1.3 minute run; Positive or negative electrospray (ESI) MS detection using Waters Xevo G2 QTof mass spectrometer. Where necessary, lock mass was applied as external standard (leucine enkephalin). Detection by UV 214 nm, 254 nm and/or DAD in parallel with ELS methods; data was processed using Waters MassLynx software.

Background: Compound storage and QC at Evotec

Evotec processes all SMR samples in a standard, high-throughput Quality Control (QC) regimen; one that checks for weight after drying, solubility, identity and purity. Solubility in DMSO is checked and only samples completely soluble at 10 mM are accepted into the SMR. LCMS is used to determine sample identity and purity. SMR accepts samples where the molecular ion is identified and Area Under the Curve (AUC) is at least 90% by Evaporative Light Scattering (ELS) or UV (214,254 nm) detection. In addition to QC of incoming samples, ongoing analysis of approximately 10% of the collection is conducted annually to monitor stability over time.

The SMR stores samples that pass QC tests in automated storage equipment for fast retrieval. SMR uses a three-tier storage system: 1) Long Term Store, 2) Working Store and 3) Short Term Store. The MLSMR stores the bulk sample dry in glass vials as the Long Term Store. Working Store samples of each compound are created in 10 mM DMSO solution and are in turn used to create distribution-ready microtiter plates, which are held in the Short Term Store until shipment (typically for less than one week). Compounds are stored at -20 °C in all three tiers of the storage system and all DMSO solutions are stored and handled under nitrogen atmosphere to prevent water absorption and ensure optimal stability and purity.

https://mlsmr.evotec.com/MLSMR_HomePage/

Corporate Headquarters: Evotec AG, Manfred Eigen Campus, Essener Bogen 7, 22419 Hamburg, Germany **Email:** info@evotec.com **Web:** www.evotec.com

Part 2 – Large scale storage stability analysis of molecules in the NCATS SMR *(formerly MLSMR)*

Christian Laggner,[1] Yuda Shayo,[1] Chayling Hendarto,[1] Charles R. Johnson,[1] Carson R. Loomis[2]

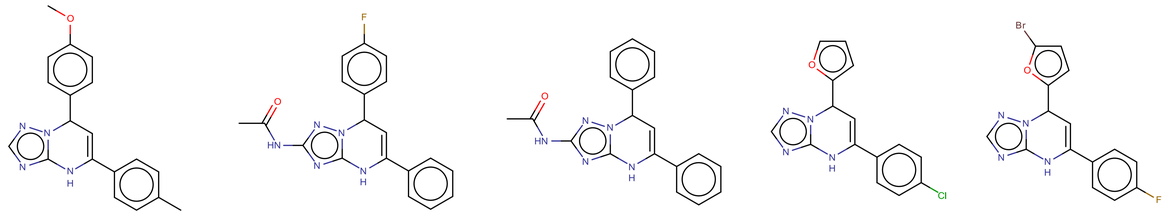
[1)] Evotec (US) Inc., 385 Oyster Point Blvd. Suite 1, South San Francisco, CA 94080, USA; [2)] National Center for Advancing Translational Sciences, 6701 Democracy Blvd., Bethesda, MD 20892, USA

Commonly encountered reactions during long-term storage

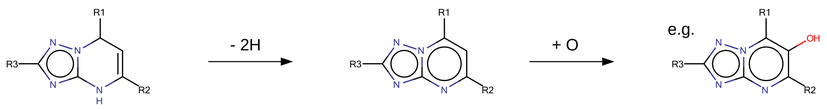
- Oxidation, often to form aromatic systems or conjugated double bonds
- Hydrolysis (e.g. of reactive esters and condensation products)
- Other elimination reactions (e.g. -HCN)
- Isomerization of double bond geometry

The product structures below are based entirely on HRMS data. None of the products were purified and fully characterized, so the structures shown are suggestive but not definitive.

Oxidation: 1,7-dihydro-[1,2,4]triazolo[1,5-*a*]pyrimidines

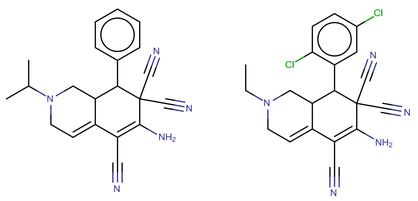


MLS002582392 MLS002582570 MLS000112698 MLS000535379 MLS000680652

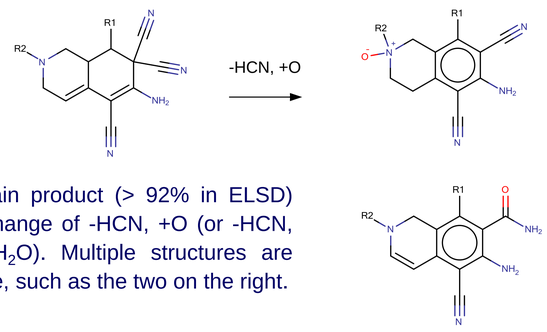


Both the first (-2H) and second (-2H, +O) oxidation products are detected for all 5 starting materials. The site of oxygen attachment is yet unknown.

Oxidation and Elimination: 6-amino-2,3,8,8a-tetrahydro-5,7,7(1*H*)-isoquinolinetricarbonitriles



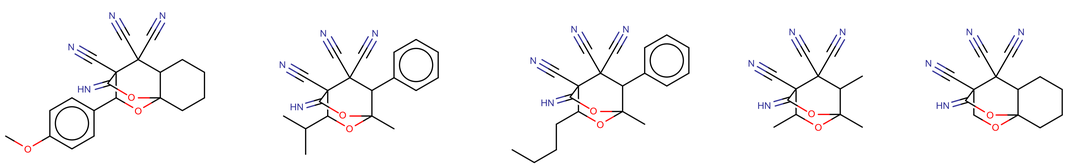
MLS002581216 MLS002538741



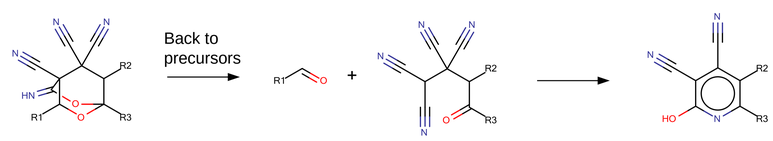
The main product (> 92% in ELSD) fits a change of -HCN, +O (or -HCN, -2H, $+H_2O$). Multiple structures are possible, such as the two on the right.

Hydrolysis, followed by internal condensation and elimination: 2,6-Dioxabicyclo[2.2.2]octane-4,8,8-tricarbonitriles

Reaction from tetracyanoethylated ketones, a precursor of these compounds,[5] to the proposed 3,4-dicyano-2(1*H*)-pyridone product has been reported to occur in boiling EtOH[6] or with pyruvic acid in aqueous acetone at RT.[7] The first three compounds are converted to a product that matches the suggested mechanism below (52-90% by ELSD). MLS001029868 has multiple products overlapping in LCMS, including one with the expected mass. MLS001049393, which has the smallest aldehyde precursor, was found mostly unconverted (86% by ELSD).



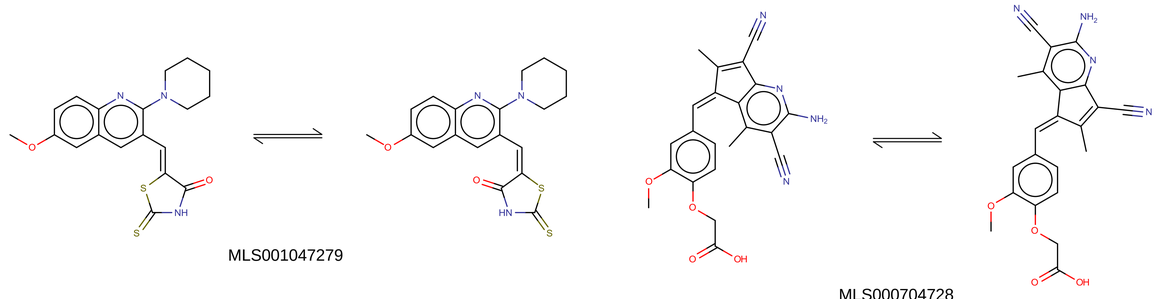
MLS000554241 MLS000589309 MLS000591416 MLS001029868 MLS001049393



(not detected) Matches observed HRMS

Isomerization

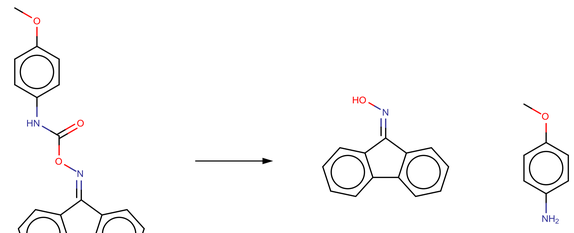
We observed several cases of C=C double-bond containing compounds where the HR molecular weight of the new compound matches that of the original one, and where conversion is never complete, but over time appears to reach an equilibrium. This gives us reason to believe that these compounds isomerize around the double bond.



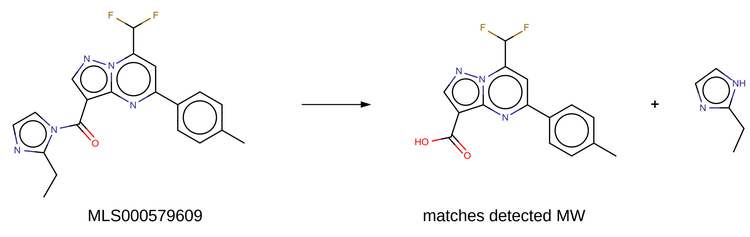
MLS001047279 MLS000704728

Two hydrolysis examples

Certain oxime esters, *N*-acylated di-, tri-, and tetrazoles, and imines (not shown) undergo hydrolysis:



MLS000578655 matches detected MW



MLS000579609 matches detected MW

Overlap with PAINS[8] and known reactive groups

Some of the compounds we analyzed have been previously described as pan assay interference compounds (PAINS)[8] or in other collections of undesirable compounds. Many of the known problematic compounds still lack a proper explanation as to why exactly they produce a false assay readout. We hope that our work contributes to the understanding of some of these mechanisms. Compounds that react before or during the assay do not necessarily need to show assay promiscuity, yet they can still lead to wrong SAR results. On the other hand, compounds that in some way interact unfavorably with the assay components (e.g. by reaction with thiol groups)[9] may be stable over many years in long-term storage.

We expect some of the reactions that occur with trace amounts of water and oxygen at low temperature to be accelerated in aqueous solution at assay conditions. For these compounds, fresh samples that pass our QC might thus still give wrong assay results. Aside from not having the expected structure when binding to the protein, they might also interfere with the assay readout (e.g. oxidation can change fluorescence/UV absorbance).

Conclusions

Since the inception of the SMR over 10 years ago, various groups of compounds that, for various reasons, lead to false assay results have been reported. Their usefulness for screening campaigns might be limited over all, or at least require specific storage and assay conditions. In an effort to add to the ever-growing knowledge about potentially problematic screening compounds, we developed an approach to identify clusters of compounds that possess limited storage stability, even when stored under state of the art storage conditions. It should be pointed out that this problem currently only affects about 1.1% of the overall SMR collection, and that overall our storage conditions appear to work well for the vast majority of diverse compound samples that we are managing.

Internally, the results from this study will help us understand the strengths and weaknesses in our compound storage and analysis process, and improve the overall quality of the SMR. The reaction mechanisms we discover can be helpful in two ways: they may explain the mechanisms of known problematic compounds or present novel approaches to certain chemical reactions in a mild and catalyst-free manner, leading to Green Chemistry applications. At the end of this ongoing project, we are planning to release to the scientific community a collection of searchable SMARTS patterns for the problematic scaffolds.

This work was funded by The National Institutes of Health, contract number HHSN271201200003I.

For questions and feedback please email ***christian.laggner@evotec.com***

References

1) *http://mli.nih.gov/mli/secondary-menu/mlscn/ml-small-molecule-repository/*
2) *http://mlsmr.evotec.com*
3) Johnson, C.R. et al. Reactions during storage of the NIH Molecular Libraries Small Molecule Repository. 245th ACS Meeting, 2013, poster MEDI-127.
4) van Dongen, S. Graph Clustering by Flow Simulation, PhD thesis, University of Utrecht, May 2000. (*http://www.library.uu.nl/digiarchief/dip/diss/1895620/inhoud.htm*)
5) Nasakin, O.E. et al. Reaction of tetracyanoethylated ketones with aldehydes. Synthesis of 3-imino-2,6-dioxabicyclo[2.2.2]octanes. *Chem. Heterocycl. Comp*. **1984**, 1206-1210.
6) Dickinson C.L. Cyanocarbon Chemistry. XV. A new synthesis of 3,4-dicyano-2 (1H)-pyridones. *J. Am. Chem. Soc.* **1960**, 4367–4369.
7) Nasakin, O.E. et al. Tetracyanoalkanones in the synthesis of heterocycles. Synthesis of 3,4-dicyano-2(1H)-pyridones by the interaction of tetracyanoalkanones with pyruvic acid. *Chem. Heterocycl. Comp.* **1985**, 1019-1022.
8) Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, 2719-2740.
9) Dahlin, J.L. et al. PAINS in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *J. Med. Chem.* **2015**, 2091-2113.