Tautomers

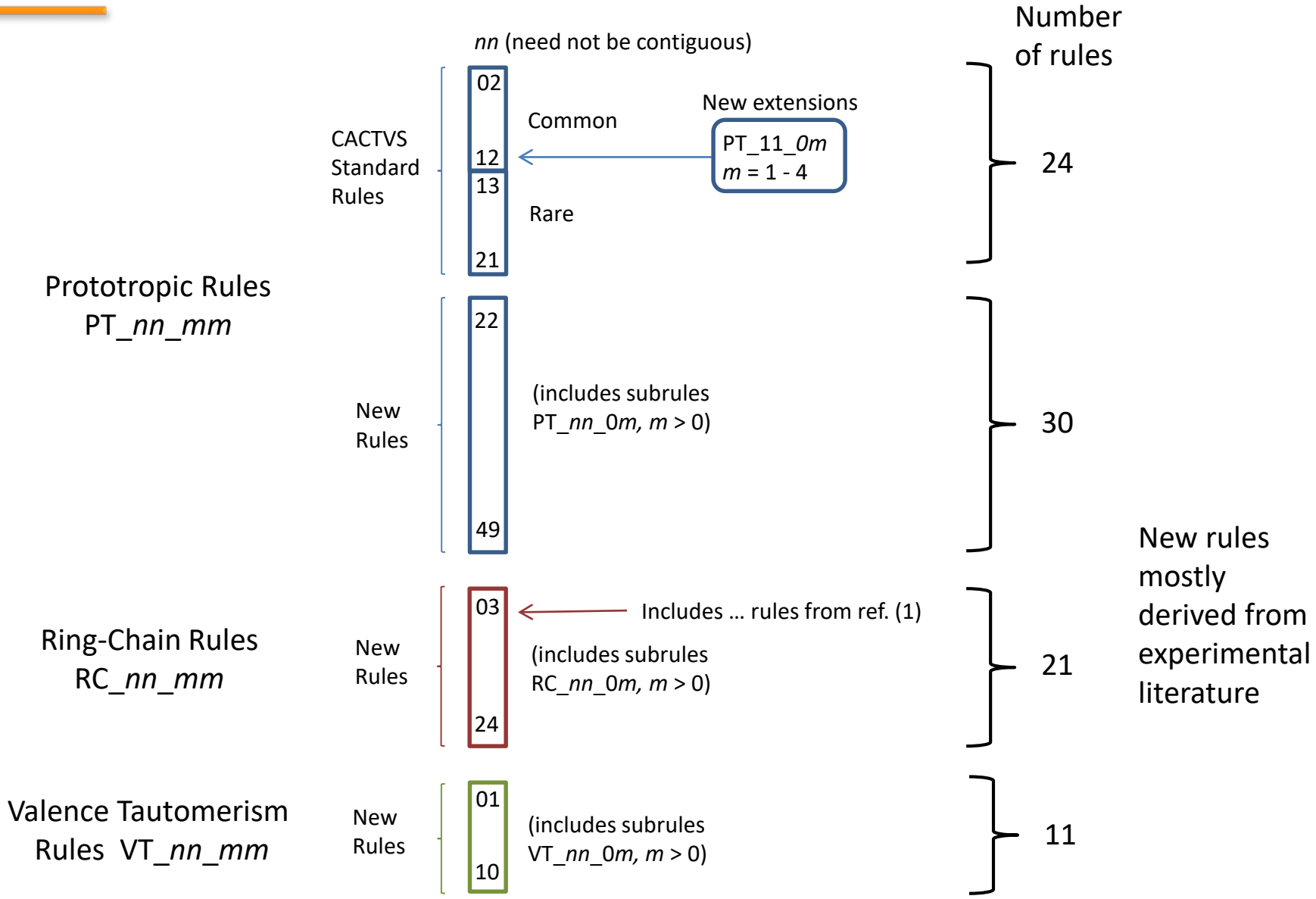
Devendra K. Dhaked and Marc C. Nicklaus

IUPAC Project #2012-023-2-800 “Redesign of Handling of Tautomerism for InChI V2”

CADD Group

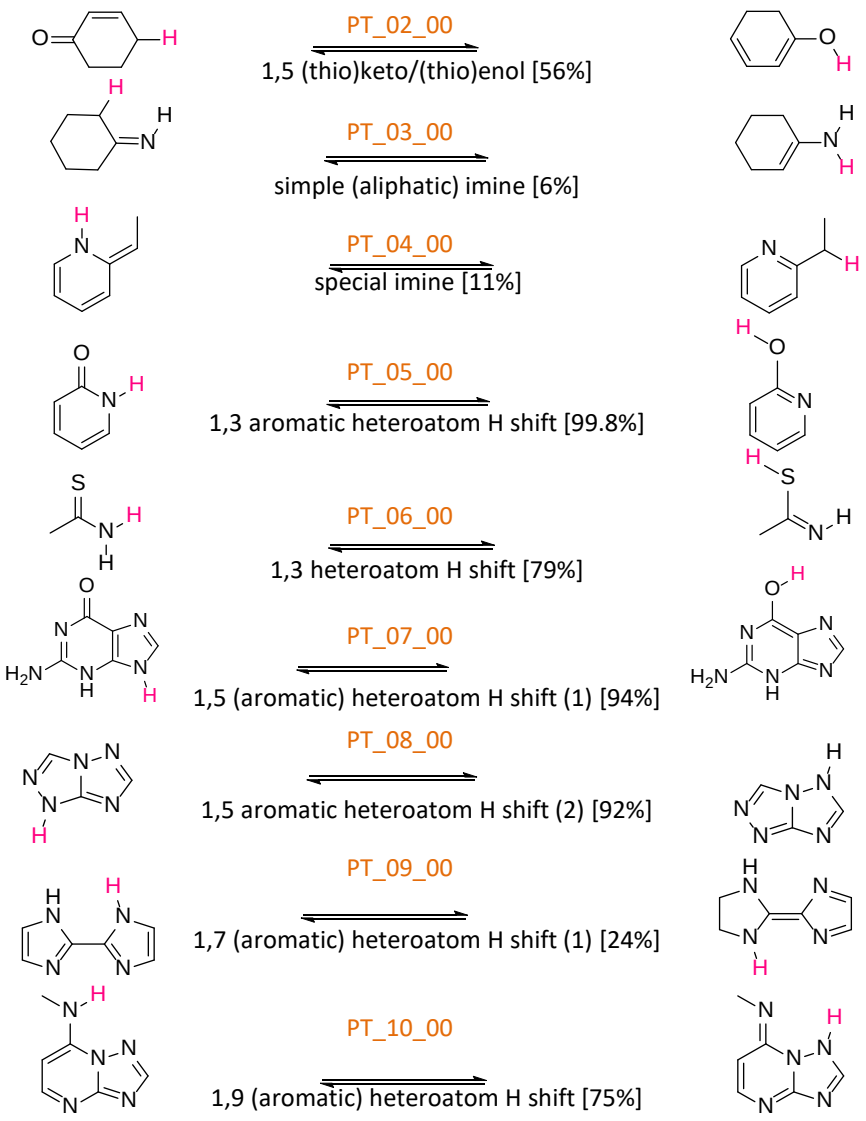
Chemical Biology Laboratory
Center for Cancer Research
National Cancer Institute
National Institutes of Health

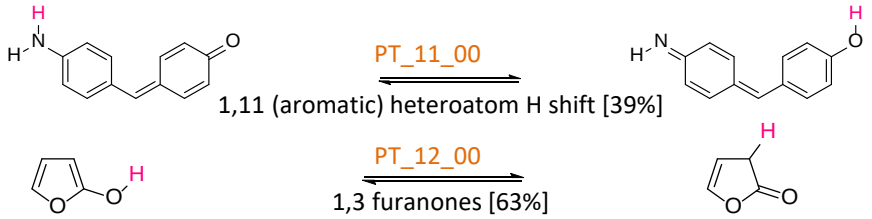
Rule Set



(1) Guasch *et al.*, J. Chem. Inf. Model. **2014**, 54, 2423−2432

CACTVS standard tautomeric transforms – Common Rules



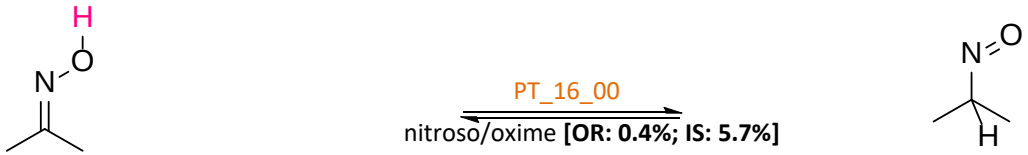


- Occurrence rates in PubChem (96M unique structures): between 65% (06) and 0.6% (11)
- **[x%]** = Matches of **NonStd InChIKey**s with rule-enumerated tautomers, sum of partial and complete matches ("**InChI Success**") (Options used: **KET 15T** DONOTADDH W0 RECMET NEWPS SPXYZ SAsXYZ Fb Fnud)
- Generally, complete match counts are higher than partial ones
- StdInChI: Success = 0% for 03, 04, 12; reduced for all others
- Long-range extensions (PT_11_0*n*), n = 13, 15, 17, 19: NonStd InChI Success pretty good, between 15% and 44%
- Rare Standard CACTVS rules (PT_12_00 – PT_21_00): rare in PubChem (occur. rates <0.5%), InChI Success = ~0% except 2 rules
- Note that example structures are just that: examples. Similar for the names. The SMIRKS are really defining the rule, such as:

  PT_02_00: 1.5 (thio)keto/(thio)enol
  [O,S,Se,Te;X1:1]=[Cz1H0:2][C:5]=[C:6][CX4z0,NX3:3][#1:4]>>
  [#1:4][O,S,Se,Te;X2:1][Cz1:2]=[C:5][C:6]=[Cz0,N:3]

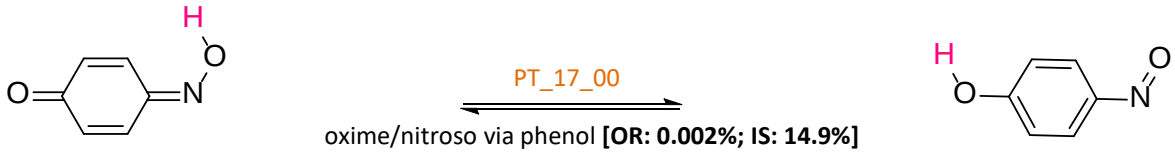
  Note that specific CACTVS transforms may be used that influence execution of specific rules.

CACTVS standard tautomeric transforms – Rare Rules with Some InChI Success



nitroso/oxime **[OR: 0.4%; IS: 5.7%]**

[#1:1][O;!R:2][N+0z1:3]=[CX3:4]>>[O;!R:2]=[N+0z1:3][CX4:4][#1:1]

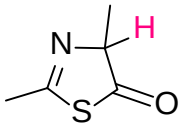


oxime/nitroso via phenol **[OR: 0.002%; IS: 14.9%]**

[#1:1][O:2][Nz1:3]=[C:4][C:5]=[C:6][C:7]=[O:8]>>[O:2]=[Nz1:3][c:4]=[c:5][c:6]=[c:7][O:8][#1:1]

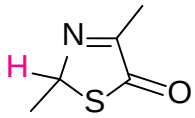
Occurrence Rate (OR) as measured in PubChem (96M)
InChI Success (IS) for Nonstandard InChI

New Prototropic Transforms with High Occurrence Rates

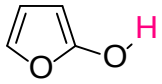


PT_22_00

imine/imine **[OR: 3.1%; IS: 0.05%]**

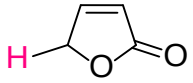


[#1:1][CX4:2][NX2:3]=[CX3:4]>>[CX3:2]=[NX2:3][CX4:4][#1:1]



PT_23_00

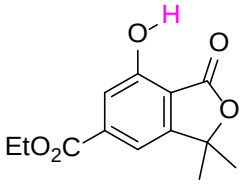
1,5 furanones **[OR: 1.3%; IS: 1.9%]**



[#1:1][O,S,NX3:2][cX3;z2;r5:3]=[c;r5:4][c;r5:5]=[c;z{1-2};r5;R{1-2}:6]>>
[O,S,NX2:2]=[CX3;z2;r5:3][C;r5:4]=[C;r5:5][Cz{1-2};r5;R{1-2}:6][#1:1]

Occurrence Rate (OR) as measured in PubChem (96M)
InChI Success (IS) for Nonstandard InChI

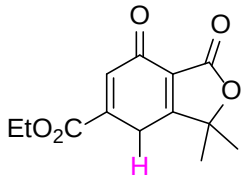
Rare Rules (Prototropic and Ring-Chain) with InChI Success > 1%

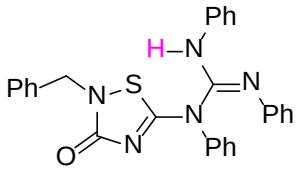


PT_48_00

Benzofuranone **[OC: 473; IS: 6.3%]**

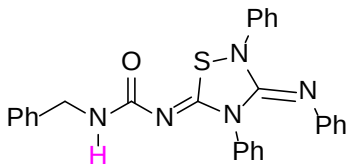
[#1:12][OX2:10][c:2]1=[c:3][c:4]=[c:5][c:6]2=[c:1]1[C:8](=[O:11])[O:7][C:9]2>>
[OX1:10]=[C:2]1[C:3]=[C:4][CX4:5]([#1:12])[C:6]2=[C:1]1[C:8](=[O:11])[O:7][C:9]2



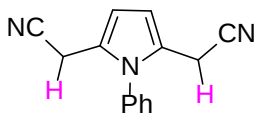


RC_14_00

thiadiazoline rearrangement **[OC: 108; IS: 1.9%]**

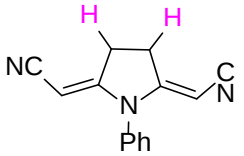


[#1:1][NX3:2][CX{2-3}:3][NX3:4][CX3;R1:5]1[SX2;R1:6][NX3;R1:7][CX3;R1:8](=[O:9])[NX2:10]=1>>
[NX3;R:2]1[CX{2-3};R:3][NX3;R:4][CX3;R:5](=[NX2:10][CX3:8](=[O:9])[NX3:7][#1:1])[SX2;R1:6]1

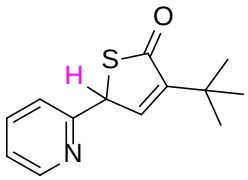


PT_44_00

2-subsituted-pyrrole **[OC: 9,599; IS: 1.7%]**



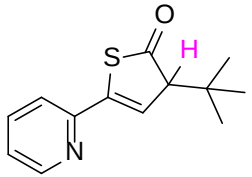
[#1:7][CX4;$([C][C]#[N]),$([C][C](=[O])[O]):6][c:5]1=[cR1:4][c:3]=[c:2][nX3:1]1>>
[#1:7][CX4R1:4]1[CX3:3]=[CX3:2][NX3:1][CX3:5]1=[CX3;$([C][C]#[N]),$([C][C](=[O])[O]):6]

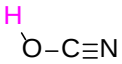


PT_42_00

△3-/△4-pyrro(thio/seleno)lin-2-one **[OC: 437,342; IS: 1.4%]**

[#1:1][CX4:4]1[NX3,O,S,Se:5][CX3:6](=[O:7])[CX3:2]=[CX3;a0:3]1>>
[#1:1][CX4:2]1[CX3;a0:3]=[CX3:4][NX3,O,S,Se:5][CX3:6]1=[O:7]

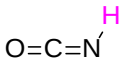




PT_18_00

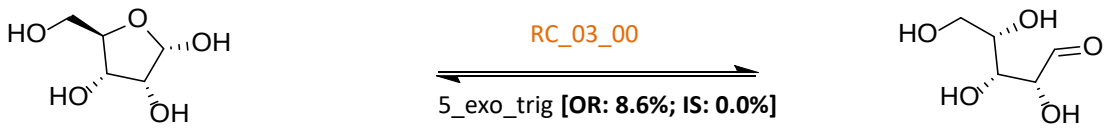
cyanic/iso-cyanic acids **[OC: 1,872; IS: 1.2%]**

[#1:1][O:2][C:3]#[N:4]>>[O:2]=[C:3]=[N:4][#1:1]

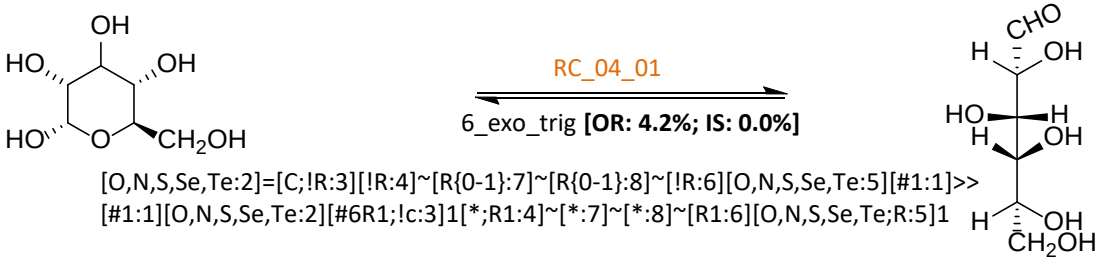


Occurrence Count (OC) as measured in PubChem, absolute numbers (96M)
InChI Success (IS) for Nonstandard InChI

And What About Sugars…?



[#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*:4]~[*:7]~[R1:6][O,N,S,Se,Te;R:5]1>>
[O,N,S,Se,Te:2]=[C;!R:3][R{0-1}:4]~[R{0-1}:7][!R:6][O,N,S,Se,Te:5][#1:1]

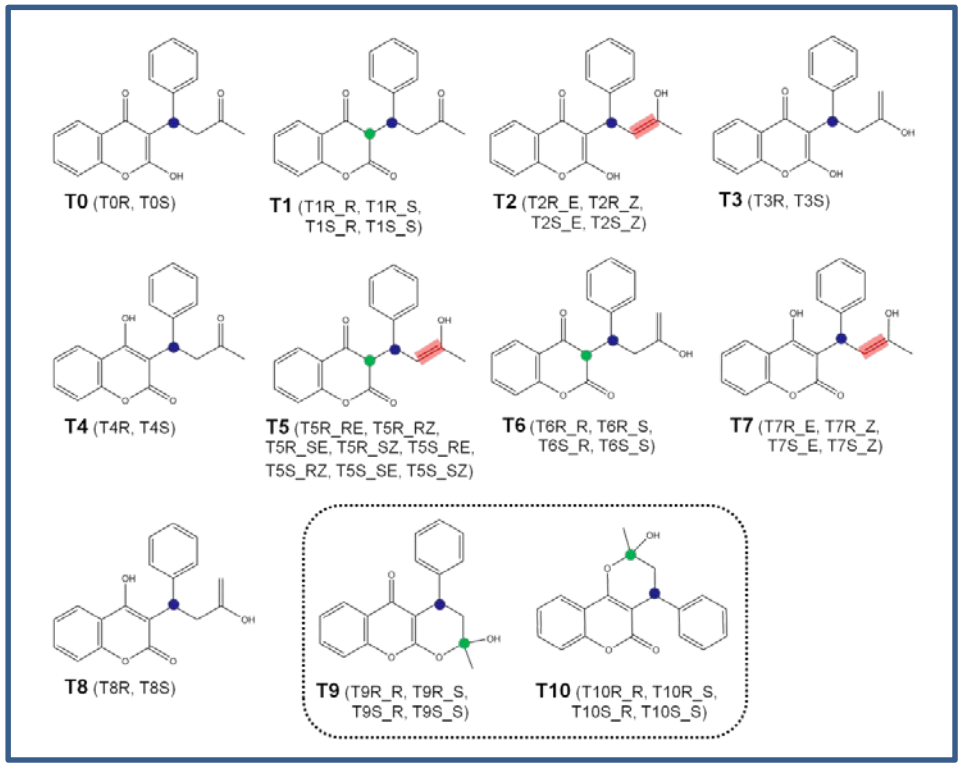


[O,N,S,Se,Te:2]=[C;!R:3][!R:4]~[R{0-1}:7]~[R{0-1}:8]~[!R:6][O,N,S,Se,Te:5][#1:1]>>
[#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*;R1:4]~[*:7]~[*:8]~[R1:6][O,N,S,Se,Te;R:5]1

Occurrence Rate (OR) as measured in PubChem (96M)
InChI Success (IS) for Nonstandard InChI

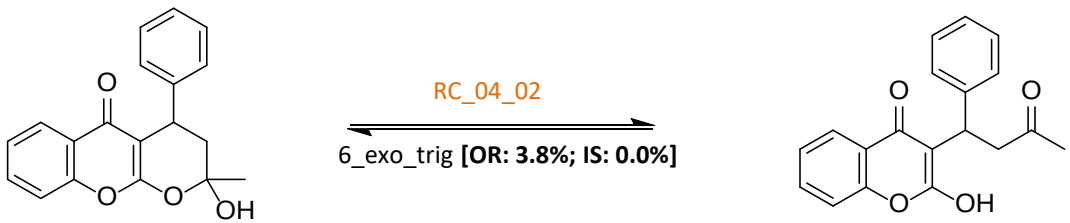
And Then There is Warfarin...

Igor Pletnev: " […] 40 tautomers of warfarin listed in [(1)], should also converge to the same [InChI] string."



(1) Guasch *et al.*, J. Org. Chem. **2015**, 80, 9900−9909

20 tautomers per enantiomer via non-persistent stereogenic elements



[O,N,S,Se,Te:2]=[C;!R:3][!R:4]~[!R:7]~[R{0-1}:8]~[R{0-1}:6][O,N,S,Se,Te;!R:5][#1:1]>>
[#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*;R1:4]~[*;R1:7]~[*:8]~[R:6][O,N,S,Se,Te;R1:5]1

Summary and Further Questions

- 20-30 transforms can be assigned higher priority
- Should we set 0.1% occurrence rate as threshold (>100,000 in PubChem)?
    - 25 rules
        - 4 of those: ring-chain
        - 1 of those: valence tautomerism
- Only 1 rule had 0 occurrence in PubChem

- Only 3 rules have (NonStandard) InChI Success rates > 90%
- Only 7 rules have (NonStandard) InChI Success rates > 50%
- 57 rules have (NonStandard) InChI Success rates = 0%

- New prototropic transforms: should be doable
- Ring-chain transforms: more difficult
- Valence tautomerism rules: may be nearly impossible with current InChI chemical structure model