

# Tautomer Database: A Comprehensive Resource for Tautomerism Analyses

Devendra K. Dhaked, Laura Guasch, **Marc Nicklaus**

Submitted date: 22/11/2019 • Posted date: 29/11/2019

Licence: CC BY-NC 4.0

Citation information: Dhaked, Devendra K.; Guasch, Laura; Nicklaus, Marc (2019): Tautomer Database: A Comprehensive Resource for Tautomerism Analyses. ChemRxiv. Preprint.

<https://doi.org/10.26434/chemrxiv.10790369.v1>

We report a database of tautomeric structures that contains 2,819 tautomeric tuples extracted from 171 publications. Each tautomeric entry has been annotated with experimental conditions reported in the respective publication, plus bibliographic details, structural identifiers (e.g. NCI CADD identifiers FICTS, FICuS, uuuuu, and Standard InChI), and chemical information (e.g. SMILES, molecular weight). The majority of tautomeric tuples found were pairs, the remaining 10% were triples, quadruples, or quintuples, amounting to a total number of structures of 5,977. The types of tautomerism were mainly prototropic tautomerism (79%), followed by ring-chain (13%) and valence tautomerism (8%). The experimental conditions reported in the publications included about 50 pure solvents and 9 solvent-mixtures with 26 unique spectroscopic or non-spectroscopic methods.  $^1\text{H}$  and  $^{13}\text{C}$  NMR were the most frequently used methods. A total of 77 different tautomeric transform rules (SMIRKS) are covered by at least one example tuple in the database. This database is available as a spreadsheet for free download from <https://cactus.nci.nih.gov/download/tautomer/>.

## File list (2)

Tautomer\_Database\_MCN\_V8\_2019-11-22.pdf (1.74 MiB)

[view on ChemRxiv](#) • [download file](#)

Supporting\_information\_MCN\_V7\_2019-11-01.docx (1.07 MiB)

[view on ChemRxiv](#) • [download file](#)

# Tautomer Database: A Comprehensive Resource for Tautomerism Analyses

Devendra K. Dhaked, Laura Guasch, Marc C. Nicklaus\*

Computer-Aided Drug Design Group, Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, NIH, Frederick, MD 21702, USA

## ABSTRACT

We report a database of tautomeric structures that contains 2,819 tautomeric tuples extracted from 171 publications. Each tautomeric entry has been annotated with experimental conditions reported in the respective publication, plus bibliographic details, structural identifiers (e.g. NCI CADD identifiers FICTS, FICuS, uuuuu, and Standard InChI), and chemical information (e.g. SMILES, molecular weight). The majority of tautomeric tuples found were pairs, the remaining 10% were triples, quadruples, or quintuples, amounting to a total number of structures of 5,977. The types of tautomerism were mainly prototropic tautomerism (79%), followed by ring-chain (13%) and valence tautomerism (8%). The experimental conditions reported in the publications included about 50 pure solvents and 9 solvent-mixtures with 26 unique spectroscopic or non-spectroscopic methods.  $^1\text{H}$  and  $^{13}\text{C}$  NMR were the most frequently used methods. A total of 77 different tautomeric transform rules (SMIRKS) are covered by at least one example tuple in the database. This database is available as a spreadsheet for free download from <https://cactus.nci.nih.gov/download/tautomer/>.

## INTRODUCTION

Tautomerism is a phenomenon in which a set of molecules can interconvert by movement of a hydrogen or group of atoms and/or molecular rearrangement. The movement of hydrogen atoms along with the migration of pi-bonds is called prototropic tautomerism. The intermolecular arrangements leading to isomerization due to ring opening or cyclization are known as ring-chain tautomerism. (We have recently compiled 11 sets of rules for ring-chain tautomerism in SMIRKS notation.<sup>1</sup>) Another type of isomerization, which involves rapid reorganization of single and double bonds without migration of any atom or group is termed valence tautomerism.

At its core, tautomeric interconversion is a chemical reaction although one that is seen as happening “by itself” given the right conditions, not forced by a chemist’s intervention. The boundary between the two is, however, to a certain degree a matter of definition. In this vein, the interconversion of tautomers is affected by solvent, pH, light, temperature, pressure, etc. Tautomers may be in rapid equilibrium or may take days to years to interconvert under experimental, storage, or transport conditions. It is well documented that a tautomer that is dominant in solution may not necessarily be favored in the solid state.<sup>2,3</sup>

There has been significant interest in the pharmaceutical industry in the tautomeric polymorphism of drug-related molecules. Tautomerism is one of the causes of polymorphism of drug molecules. These polymorphs may have important, differing, properties such as solubility, bioavailability, physical/chemical stability etc. and may affect the patent life of the molecule.<sup>4,5</sup> Well-known examples of drug molecules that show tautomeric polymorphism are omeprazole, ranitidine, sulfasalazine. Polymorphs of these drugs have been isolated as different tautomeric forms or sometime mixtures of their tautomers.

In prototropic tautomers, hydrogen transfer may take place by either intra- or inter-molecular processes. Intramolecular transfer may involve an isolated molecule in the gas phase or in dilute solution in aprotic solvents while intermolecular hydrogen transfer may be observed in the self-associated aggregates of tautomeric forms, in highly concentrated solution in aprotic solvent, and in the solid state. Protic solvents may facilitate transfer of hydrogen by means of their own proton(s) after forming a complex with the solute’s tautomer.<sup>6,7</sup> In most cases, migration of a single hydrogen is involved, however molecules with more polar groups may transfer more than one proton by an either concerted or stepwise mechanism.<sup>8,9</sup>

Tautomers usually have different physicochemical properties such as logP, hydrophobicity, pKa, solubility, electrostatic potential, similarity index, etc. with concomitantly computation of such properties and molecular descriptors yielding different results from one tautomer to another.<sup>10</sup> This leads to the question: how to select the tautomer(s) of a molecule that allow one to most accurately predict its properties? One of the issues in this context has been the lack of a publicly accessible database providing a significant number of quantitative ratios or qualitative data of tautomeric forms in different solvents. The difference in the hydrogen bond donors and acceptor

patterns between different tautomers of a screening or lead compound may affect interaction of these tautomers with its target protein and so may have profound influence on the drug discovery process. Therefore, consideration of tautomers has been of high interest to the drug design community for decades.<sup>11–14</sup> A study suggested that using all possible (predicted) tautomers including unstable and stable ones yielded poorer docking results when compared to including only the stable one(s) in water.<sup>12</sup> The oftentimes small structural changes between members of a congeneric series nevertheless can lead to significant changes in the relative abundance of different tautomers.

It has also been quite difficult to identify the tautomer that binds to a target protein as it may be any one from the pool of low- to high-energy tautomers found in a different environment. Martin<sup>10</sup> discussed several examples where a minor or less stable tautomer was recognized as binding to the macromolecule. X-ray crystallography, unless solving macromolecular structures at ultra-high resolution (say, below 0.8 Å), typically suffers from insufficient precision in the position of heavy atoms and negligible electron density of hydrogen atoms. This lack of experimental evidence of hydrogen positions leads to placement of hydrogens based on chemical assumptions if not default settings of software. Neutron crystallography is more reliable in the determination of the tautomeric state of molecule, unfortunately there are very few structures in the Protein Data Bank (PDB)<sup>15</sup> or Cambridge Structural Database (CSD)<sup>16</sup> solved by neutron diffraction. Computational approaches, many of which need suitable training sets of sufficient size of experimentally determined tautomeric prevalences, may be an, if not in many cases the only, alternative. Borbulevych *et al.*<sup>17</sup> suggested refinement of X-ray structures based on a quantum-mechanical force field could aid in the determination of the protein-bound tautomer.

Different studies have given widely varying estimates of the tautomeric potential of molecules in small-molecule databases. Trepalin *et al.*<sup>14</sup> indicated the presence of 0.78% of tautomeric molecules in seven commercial database of bioscreening compounds using 23 pairs of tautomeric fragments. Cruz-Cabeza and Groom<sup>15</sup> suggested that 10% of the molecules in the CSD have tautomeric potential. Milletti *et al.*<sup>20</sup> estimated that an average of 29% molecules in four databases are tautomeric among about 0.7 million records. Chemoinformatics analysis by our group on 103.5 million records aggregated from 150 small molecule databases suggested that more than 66% of the molecules are susceptible to tautomerism based on about 20 tautomerism transform rules applied.<sup>21</sup> Recently, we found more than 30,000 cases where two or more

products from the Aldrich Market Select (AMS) database of (then) ~ 6 million chemicals were just different tautomeric forms of the same compound – with the interesting corollary that these different tautomeric forms of a compound were offered at different unit prices by the same vendor.<sup>22</sup> In another study, we reported numbers in the same range for typical small synthetic organic molecules from the AMS and natural products, yielding an average of 76% and 56%, respectively, of structures susceptible to prototropic tautomerism. In contrast hereto, the occurrence rate of ring-chain tautomerism was found to be twice in natural products compared to AMS (16% vs. 8%).<sup>1</sup>

Organizations oftentimes develop their own protocols for compound registration, which may lead to incompatible ways of handling tautomerism during, e.g., database and company mergers or switch to other compound registration and management software. Warr<sup>23</sup> reviewed different approaches used by 27 software vendors and database producers for compound registration including how tautomers were handled.

While thus a significant body of work exists of experimental, theoretical (quantum chemical), and – to a limited extent – chemoinformatics (rule-based), studies of tautomerism, very few systematic collections of experimental results in this field have been undertaken so far. A set of 785 transformations belonging to 11 types of tautomeric reactions with equilibrium constants measured in different solvents and at different temperatures was recently used in an effort to build QSPR models of equilibrium constants of tautomeric molecules.<sup>24</sup> To the best of our knowledge, there is currently no database publicly available that provides details of a significant number of molecules and their experimentally investigated tautomers under specific experimental conditions along with a detailed chemoinformatics analysis. Here, we report on a tautomer database we have created from literature in an attempt to compile experimental, quantitative, tautomeric preferences together with chemical and bibliographic information as well as an analysis along a set of more than 70 tautomeric transforms. We hope this resource, which we have made freely downloadable at <https://cactus.nci.nih.gov/download/tautomer/> will allow the scientific community to more easily explore the phenomenon of tautomerism by finding several thousand such molecules in one place.

## METHODS AND DATA

### Dataset of database

The current tautomer database consists of 2,819 entries, each comprising an  $n$ -tuple of tautomers ( $n = 2 \dots 5$ ) studied in a particular set of experimental conditions (pH, solvent, solvent mixture, temperature, experimental technique used). All these tuples together comprise a total of 5,977 records. The data were extracted from 171 publications, which included a number of reviews (see full list in the Supporting Information Table S1). The initial extraction from these literature sources was done by a contract mechanism (Parthys Reverse Informatics, <http://www.reverseinformatics.com>) whereas the significant work-up and curation of the initial data was performed by the authors.

For each entry for all  $n$ -tuples in the tautomer database the corresponding NCI/CADD Chemical Structure Identifiers<sup>25</sup> were calculated using the chemoinformatics toolkit CACTVS<sup>26</sup> (in which they have been implemented as standard molecular properties). The nature of these identifiers, which are based on the standard CACTVS molecular hashcodes, is based on turning off or on sensitivity to the following five chemical features: fragments, isotopes, charges, tautomers, and stereochemistry. In this database, we used the FICTS, FICuS and uuuuu identifiers out of possible 32 possible set of variants (see the original publication for explanation of the nomenclature).<sup>27</sup> The FICTS identifier, in which all five features are turned on, represents the original input structure as is. It is sensitive to fragments (such as counterions), isotopes, charges, and stereochemistry in the input structure as well as to the specific tautomer drawn. The FICuS identifier is tautomer invariant (but sensitive to all four other features), meaning that different tautomers have the same FICuS hashcode. In the case of the uuuuu identifier, all five features are turned off, implying that molecules differing only by fragment, isotope, charge, tautomer and stereochemistry have the same uuuuu (which can thus be regarded as a sort of parent structure identifier). Based on its design, the FICuS identifier is thus conceptually similar to the InChIKey, though the latter handles tautomerism less comprehensively than FICuS due to an only limited range of tautomerism transforms implemented in its current version (v. 1.05). Additionally, it is not currently possible to add entirely new types of tautomerism to the InChI[Key] calculation. It should be noted that this and other shortcomings of the current InChI in the handling of tautomerism has led to an IUPAC-sanctioned project of redesigning the handling of tautomerism

for an InChI V2<sup>28</sup>, for which the tautomer database described in this paper forms an experimental backdrop of sorts, and whose authors are involved in the IUPAC project. The standard InChI and InChIKey were also calculated with CACTVS and are included in our database.

In order to describe the tautomeric transformation(s) between the members of each of the tautomeric n-tuples in the database, we used a total of 77 rules. This set is closely related to, and essentially a major subset of, the 86 rules described in the context of redesigning of the handling of tautomerism for InChI V2 in the accompanying paper.<sup>29</sup> All these rules were encoded in SMIRKS line notation developed by Daylight Chemical Information Systems, Inc.<sup>30</sup> They were all processed in the CACTVS chemoinformatics toolkit, which comes with a default set of 20 prototropic rules covering a wide range of common and some rarer types of tautomerism, of which 12 have a representative in our database. To these, we added a subset of 8 SMIRKS from our recently published 11 types of ring $\rightleftharpoons$ chain rules (encoded in a total of 38 SMIRKS)<sup>1</sup>, plus 57 out of 61 heretofore unpublished rules, which are detailed in the accompanying publication.<sup>29</sup>

We use the following nomenclature (again aligned with the accompanying paper<sup>29</sup>) for the three types of rules discussed in this paper: 1. Prototropic tautomerism rules are called PT\_*nn*\_*mm*, where *nn* and *mm* are the number of the rule, and a possible variant, respectively. The names of most rules end with an 00 indicator, indicating that there is only one variant. 2. Ring-chain tautomerism rules are named RC\_*nn*\_*mm*, where *nn* and *mm* have the same meaning as described above. 3. Valence tautomerism rules are termed VT\_*nn*\_*mm* according to the same scheme.

To determine the single transform or sequence of transforms connecting the tuple members with each other, we applied the following procedure: In a first step, we enumerated all possible tautomers from each tautomeric tuple; in the second step, we generated a tautomer network among those enumerated tautomers. In such a network, we typically have several pathways that connect one tautomer to the other by different tautomeric transforms. As the final step, we searched for the shortest pathway, defined by the smallest number of transformation steps within the tautomeric pair. If two different paths had the same number of steps, we used a notation of the type of: {PT\_03\_00/PT\_06\_00} > PT\_09\_00. This means the pathway can either use PT\_03\_00 or PT\_06\_00 in the first step, followed by PT\_09\_00 in the second step.

## Database description

To recapitulate, our database of tautomers contains 2,819 entries, extracted from 171 publications, including reviews and research articles. It is provided to the user as a spreadsheet in Excel format. Each entry consists of three major segments: conditions, tautomer and publication. Each segment has several fields as listed in Table 1. For each additional (second, third,...) tautomer of a compound, fields in the second (and third etc.) instance of the tautomer-specific columns are populated with data, otherwise left empty. The following provides a brief explanation of some key columns in the spreadsheet. Others should be self-explanatory. A Legends worksheet is also available in the spreadsheet providing explanations for all columns.

- A. *Size*. Number of tautomers reported in the publication as being in equilibrium. In a few publications, only the main tautomer of the compound was described, in such cases we entered a second entry based on a possible (calculated) tautomer.
- B. *Solvent*. Solvent in which the tautomers were observed. This can be a mixture of solvents. If their concentration is indicated, then it is also mentioned in the solvent column.
- C. *Solvent\_proportion*. Fraction of solvents or their mixtures (percentage, ratio) used.
- D. *Solvent\_Mixture*. Indicates whether a single solvent or mixture was used. This column has a “yes” if the “*Solvent*” column indicates a solvent mixture, otherwise “no”.
- E. *Temperature*. Temperature (K) at which the tautomers were observed or the experiment carried out. In case of mass spectroscopy experiments, the temperature of the injector was used as the experimental temperature.
- F. *pH*. pH of the medium at which the tautomers were observed or experiment was conducted.
- G. *Experimental\_Method*. This describes the spectroscopic or physical methods that were used in the experimental determination of the tautomers. It may be a single method or a combination of several methods, by which the tautomers were established in the experiments. If the experimental details were not available in the review then those were extracted from the original references.

To give an example, data extraction from ref. 9 in the spreadsheet had yielded five entries for five different solvents used in the experiments (acetonitrile, CCl<sub>4</sub>, chloroform, dichloromethane, and THF), each with (the same) two tautomers reported. The following columns therefore carry a suffix (“1” or “\_1” for the first tautomer, “2” or “\_2” for the second). Explanations are only given for columns pertaining to the first tautomer since columns for subsequent tautomer(s) are entirely equivalent.



**Table 1.** Summary of data fields used in the tautomer database<sup>a</sup>

| Conditions          | Tautomer_1               | Publication_1     |
|---------------------|--------------------------|-------------------|
| Ref                 | Entry_ID1                | Filename_1        |
| Size                | Type_1                   | Publication_Doi_1 |
| Solvent             | Transf_1_2               | Publication_ID_1  |
| Solvent_Proportion  | ID_Hash_1                | Authors_1         |
| Solvent_Mixture     | FICTS_1                  | Affiliation_1     |
| Temperature         | HASHISY_1                | Title_1           |
| pH                  | FICuS_1                  | Section_1         |
| Experimental_Method | Uuuuu_1                  | Page_Number(s)_1  |
| Solvent_Mixture     | Std_InChIKey_1           | Notes_1           |
|                     | Std_InChI_1              | Cmpd_Number_1     |
|                     | SMILES_1                 |                   |
|                     | Mol_Formula_1            |                   |
|                     | Mol_Weight_1             |                   |
|                     | IUPAC_Name_1             |                   |
|                     | Quantitative_ratio_1     |                   |
|                     | Qualitative_Prevalence_1 |                   |
|                     | Prevalence_Category_1    |                   |

<sup>a</sup>Entries with the value “nul” in any column indicate that it was not possible to extract sufficiently specific information from the publication.

- H. *Entry\_ID1*. Unique ID composed from the publication reference (journal name, year, volume, page numbers) along with the tautomer ID in that publication (if given) and the nature of the tautomerism (e.g., “Keto $\rightleftharpoons$ enol”).
- I. *Type\_1*. The chemical nature of the tautomer, e.g. keto, hydroxy, imine, enamine, etc. An entry with “nul” in this column indicates that it was difficult to assign any specific name from the molecule’s common name or based on similar structures in the database.
- J. *Transf\_1\_2*. The rule(s) (prototropic [PT], ring $\rightleftharpoons$ chain [RC], or valence tautomerism [VT]) which transform(s) tautomer\_1 into tautomer\_2 (single or multiple steps). A forward slash “/” is used to indicate alternative rules for any step. Curly braces “{}” are used to group together alternative rules if these appear in multi-step transforms. The greater than sign “>” is used to separate steps in multi-step transformations. An entry with “no\_transform” in this column indicates that these pairs are not covered by our rules because these examples are related to zwitterionic and complex protonated structures, hence we did not develop any rules for them.
- K. *ID\_Hash\_1*. A hashed unique ID generated for each tautomer by the original contractor (Reverse Informatics). Some entries added later by us do not have an ID\_Hash.

- L. *FICTS\_1*. Tautomer-sensitive NCI/CADD structural identifier of *tautomer\_1*.
- M. *HASHISY\_1*. Tautomer-sensitive CACTVS structural identifier of *tautomer\_1*.
- N. *FICuS\_1*. Tautomer-insensitive NCI/CADD identifier, which therefore is the same for all tautomers of the same molecule.
- O. *uuuuu\_1*. NCI/CADD identifier for the parent compound of *tautomer\_1*.
- P. *Quantitative\_ratio\_1*. Quantitative ratio of *tautomer\_1* compared to other tautomers. This can be a single number, a range, or an upper or lower bound between 0 to 1. Decimal numbers are reported up to the third decimal digit.
- Q. *Qualitative\_Prevalence\_1*. Qualitative prevalence category of *tautomer\_1* reported in the publication. These keywords describe prevalence of one tautomer over other tautomer(s) and are mostly extracted from the papers, assigned based on the quantitative data or the spectra, or other informations in the text of paper. "nul" is used if no such keywords were available in the papers or reviews.
- R. *Prevalence\_Category\_1*. In order to make both quantitatively and qualitatively reported prevalences of tautomers comparable at least in a categorical way, we numerically categorized *tautomer\_1* into five classes: 0, 1, 2, 3, and 4 based on its quantitative ratio and/or qualitative prevalence as described below

Numeric classification of qualitative prevalence's keywords:

- 0:** Not observed
- 1:** Less favored, less stable, minor, observed
- 2:** Equally, favored, major, in equilibrium, preferred, similar spectra
- 3:** More favored, more stable, predominant, strongly favored
- 4:** Exclusively observed, only observed, only tautomer, identical tautomer

Numeric classification of quantitative amount of tautomers

- 0:** *Quantitive\_ratio* = 0.0 - 0.0099
- 1:** *Quantitive\_ratio* = 0.01 - 0.30
- 2:** *Quantitive\_ratio* = 0.31 - 0.69
- 3:** *Quantitive\_ratio* = 0.70 - 0.99
- 4:** *Quantitive\_ratio* = 1

If there were 3 or more tautomers reported, there would be corresponding columns in the spreadsheet with “3” or “\_3”, e.g. *Transf\_1\_3*, etc.

## DATABASE ANALYSIS

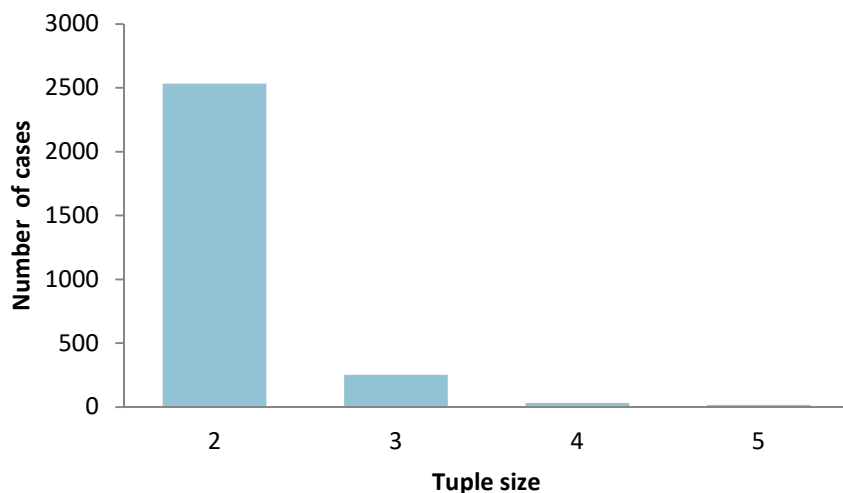
### Provenance and relationship of tuples

We did not identify any direct tuples’ duplicates in terms of both chemical structure and experimental conditions. Purely chemical duplicates were found for 479 tautomeric tuples in the database but they differ in conditions such as temperature, solvent, pH, or spectroscopy method.

All tuples came from the same original source, i.e. typically a specific publication. We did not look for, or include, sets of tautomers of the same molecule for which different tautomers in a large database such as PubChem or large vendor catalogs come from different original sources (source set providers, original sample vendors etc.). Such cases – which we have termed “tautomic conflicts”<sup>22</sup> – are numerous within, and even more numerous across, large small-molecule databases but are not the topic of this paper. They will be more thoroughly discussed in a future publication.

### Size distribution of the tuples

The database contains tautomeric tuples ranging in size from 2 to 5. The majority of tautomeric equilibria (2,530 cases) in the database are for 2 tautomers although we also found equilibria with 3 (250), 4 (27), or 5 (11) tautomers, respectively (Figure 1). The tautomeric equilibria for more than 2 tautomers comprise around 11% of the cases.

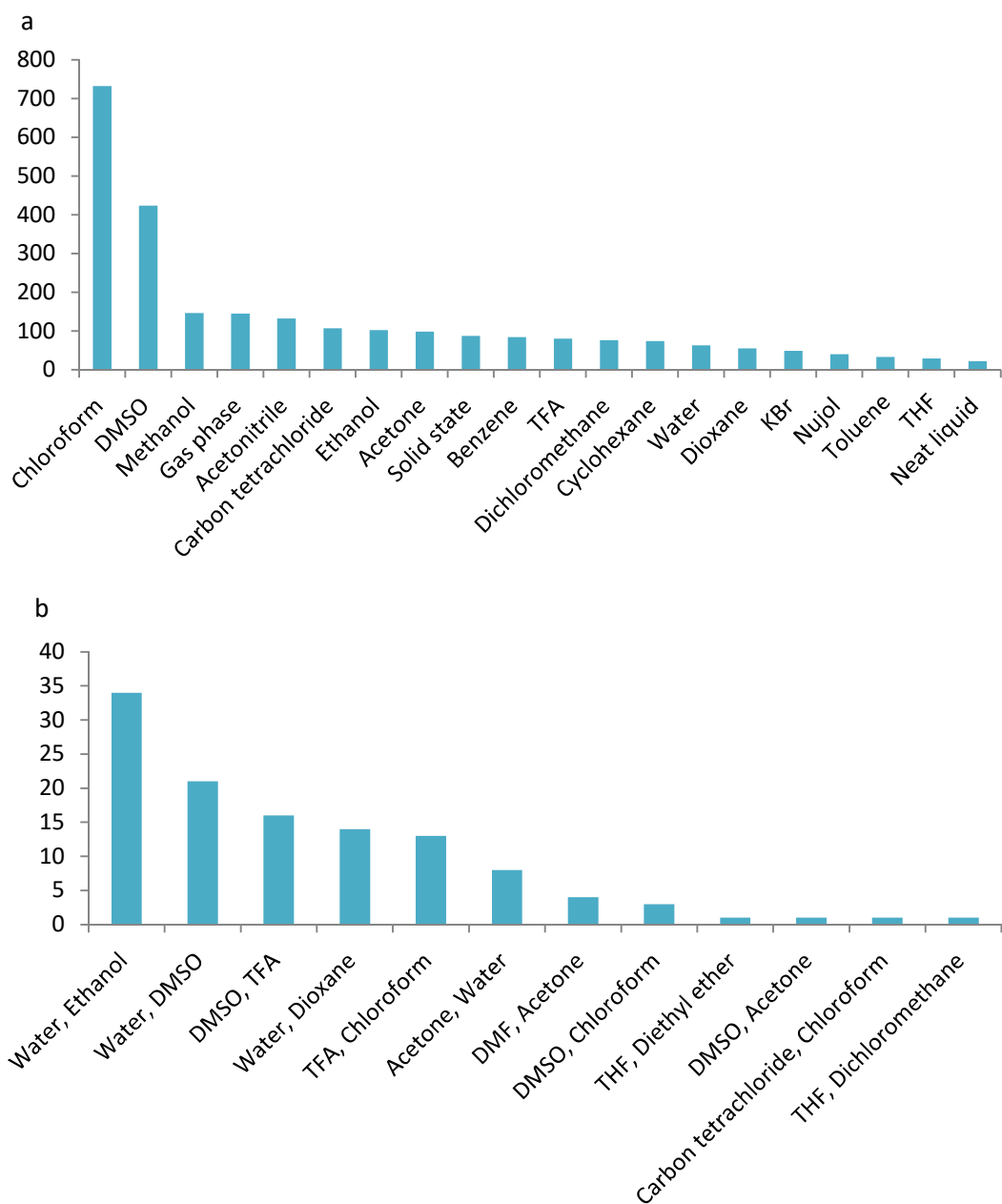


**Figure 1.** The distribution of the sizes of tautomer tuples in the database.

## Solvent

The database contains tautomeric equilibrium studies performed not only in solvents as the medium but also in the solid state, neat liquid, gas matrix, and vapor phase. Still, the majority of experiments were conducted in some kind of solvent or solvent mixture. About 50 different types of solvents were reported for the determination of tautomeric contents of the molecules under study. The frequency distribution of the top 20 solvents is given in Figure 2a. Chloroform (normal or deuterated) was the most frequently used solvent, with dimethyl sulfoxide (DMSO) in second place both together accounting for around 45% of all reported solvents. If we add methanol (~6%) and acetonitrile (~5%), the top 4 solvents cover about 57% of the cases. Looking at the tail end of the distribution shows that polar solvents such as water were used infrequently (~2%), at least as a pure solvent.

Turning to the non solvent-based experiments, we note that 6.8% and 3.8% of the cases were determined in gas phase and solid state, respectively. In ~1% of the cases, tautomerism was determined in the neat liquid form of the molecules under investigation. The database has 12 solvent mixtures, in which 12 different types of solvents were used. Around 5% molecules of the database were studied in such mixtures. The majority of these cases were water+ethanol (34), water+DMSO (21) and DMSO+TFA (16), respectively. Their quantitative distribution is given in Figure 2b. In these mixtures, water was one of the common co-solvents mixed with ethanol, DMSO, dioxane, or acetone, respectively, in ~66% of the cases of mixtures.

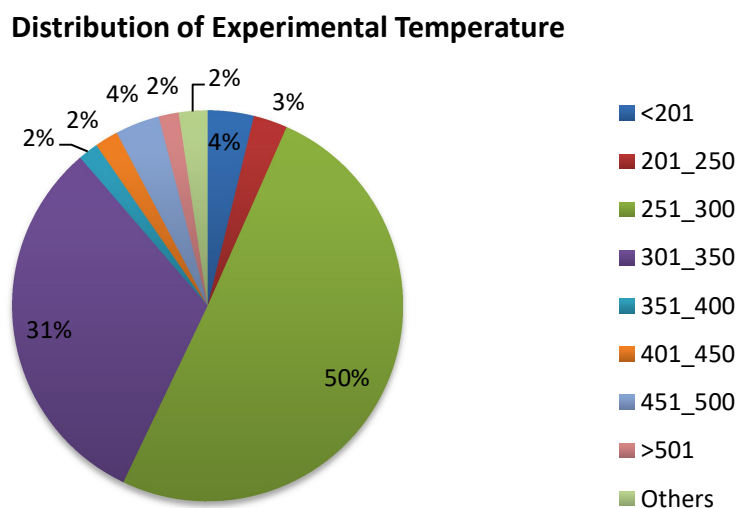


**Figure 2.** The frequency distribution of most commonly used (a) solvents and (b) solvent mixtures.

### Temperature distribution

Experimental temperature information is available for 1389 entries in the form of either exact value, range, room (RT), or ambient temperature. About 82% of the studies represented in the

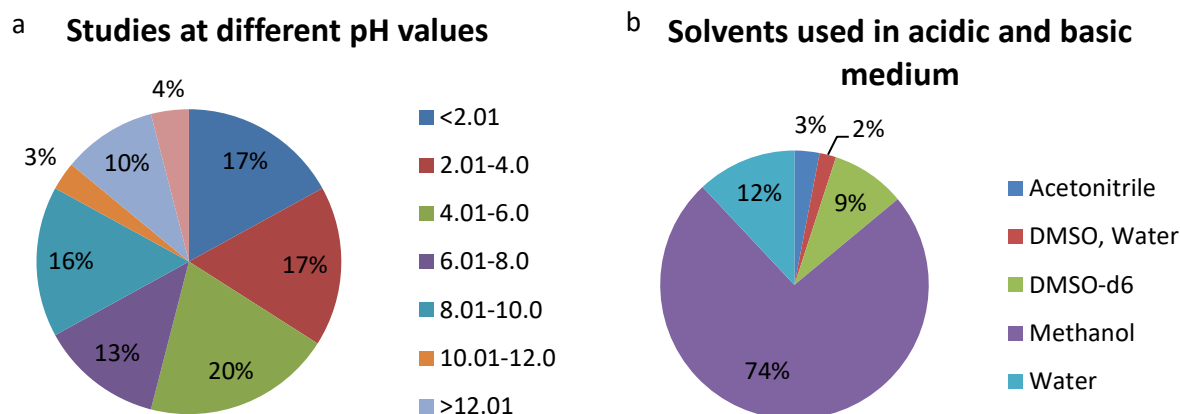
database were carried out at a temperature range of 250-350K. The majority of those (50%) were carried out in the range of 251-300K. There were only 53 entries below 201K, and 23 entries at higher temperature (e.g. 523K).



**Figure 3.** Temperature range distribution of experimental studies. (The range of 251-300 K includes studies that simply reported “room temperature.”)

### pH distribution

The database has experimental pH details for 100 entries. Of those, 63% of the records were reported to have used acidic medium. The distribution of entries by pH range is shown in Figure 4a. Except for the pH value range 10.01-12.0, most other ranges have at least 10% of the studies each. Most of the acidic and basic condition studies fell in the range of 4.01-6.0 and 8.01-10.0, respectively. We did not find any tautomeric tuples containing four or more tautomers having all been studied within one particular pH range. In most of the cases (91 entries), a set of 2 tautomers was observed at the given pH range and in a few cases (9 entries) a set of 3 tautomers was reported in pH based studies. In these studies, medium polar to polar solvents (98%) or their mixtures (2%) were used. The most commonly used solvent for such studies was methanol (Figure 4b). These studies used the following spectroscopy methods:  $^1\text{H}$  NMR, Flash photolysis, Raman, UV and UV/VIS. Of these, UV/VIS spectroscopy was used in 79% of the cases with methanol, acetonitrile and DMSO-water.

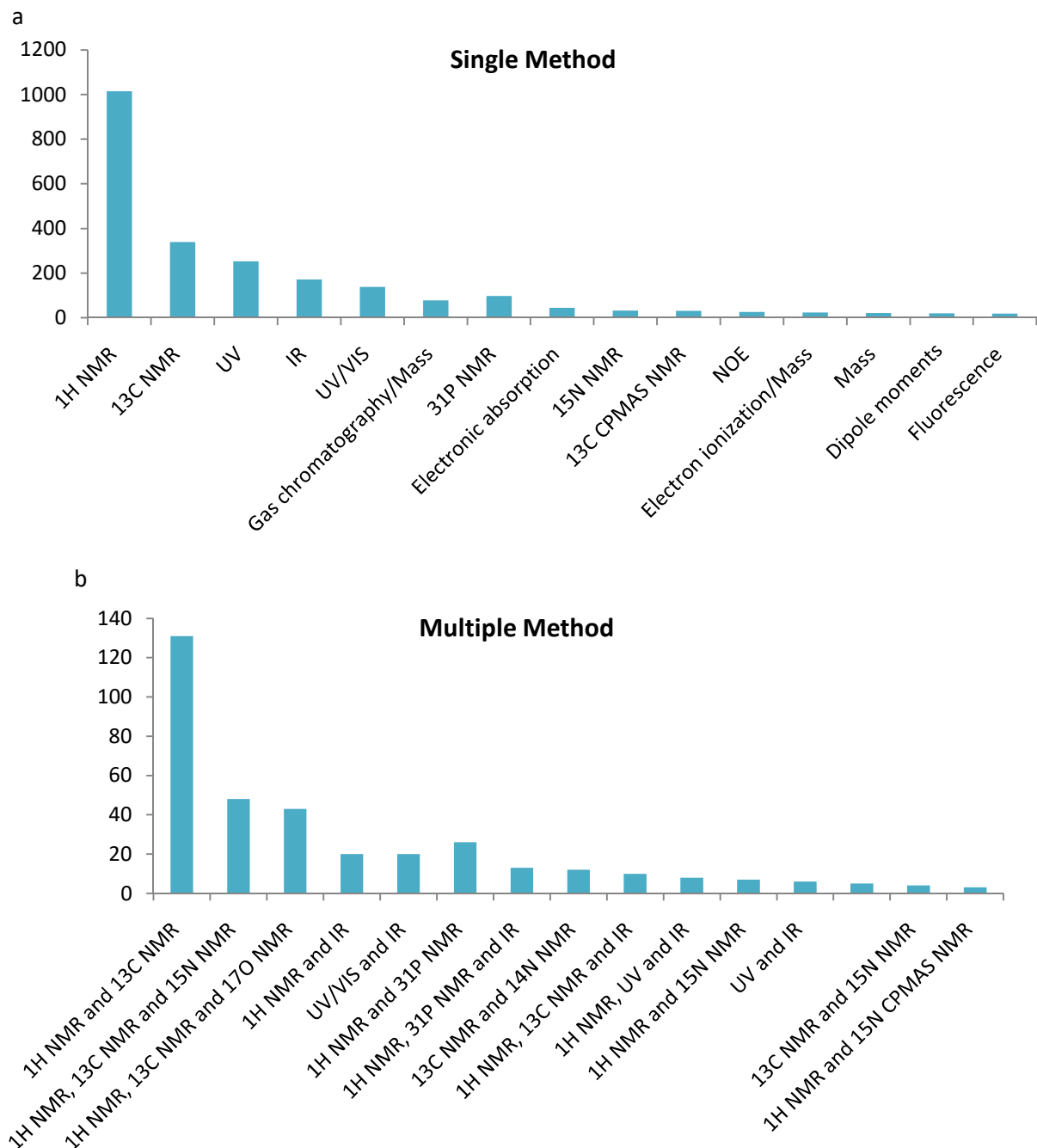


**Figure 4.** (a) Distribution of studies at different pH ranges. (b) Distribution of different solvents used in pH based studies.

## Experimental methods

In the most of the studies (85%), a single spectroscopy or physical method was used while in remainder of the studies 2 to 3 methods were used, often by way of an additional method used as support of the primary method. In the multiple method studies, spectroscopic methods from among  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{14}\text{N}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ , and/or  $^{31}\text{P}$  NMR spectroscopy were the most common ones and these NMR methods accounts for  $\sim 75\%$  cases in multiple method. Out of the total 29 unique methods,  $^1\text{H}$  NMR (1014),  $^{13}\text{C}$  NMR (340), UV (253), IR (172) and UV/VIS (139) are the top 5 spectroscopy methods which were used as the major experimental technique (Figure 5a). In the multiple method studies,  $^1\text{H}$  NMR and  $^{13}\text{C}$  NMR were frequently used together (131). In addition,  $^1\text{H}$  NMR was commonly used together with other methods such as  $^{31}\text{P}$ ,  $^{15}\text{N}$  and  $^{17}\text{N}$  NMR, and IR. Figure 5a/5b clearly shows that spectroscopy methods have been substantially more often used in the study of tautomers than non-spectroscopy physical methods such as dipole moments, X-ray diffraction and electron diffraction. The solid state NMR methods  $^{13}\text{C}$  and  $^{15}\text{N}$  CPMAS were also used together with other NMR, IR and X-ray methods. It is interesting to note that some of the spectroscopic methods have used different types of solvents, for example,  $^1\text{H}$  NMR, UV, IR,  $^{31}\text{P}$  NMR and  $^{13}\text{C}$  NMR methods were performed in 41, 22, 19, 20 and 14 different solvents and solvent mixtures, respectively. Chloroform and DMSO were the

most important solvents in  $^1\text{H}$  NMR (279 and 227 cases, respectively) and  $^{13}\text{C}$  NMR (182 and 84 cases, respectively) (see the Solvents with Spectroscopy Methods in the Supporting Information, Table S2. In IR chloroform (59) and nujol (38), in UV/VIS methanol (89) and acetonitrile (8), and in UV ethanol (76) and water (31) were used extensively.



**Figure 5.** Frequency distribution of (a) single experimental methods and (b) multiple experimental methods (only top 15 methods shown).



### Analysis by tautomeric transform rules

As already mentioned, we used as the starting point for the tautomeric rule compilation (a) 20 standard prototropic rules (comprised by the 20 default CACTVS rules PT\_02\_00 - PT\_21\_00); (b) 11 ring $\rightleftharpoons$ chain (RC\_01\_00 – RC\_11\_00) rules that have been published by our group recently.<sup>1</sup> We point out that most of these 11 rules have more than one variant, yielding in fact a total of 38 SMIRKS strings. In addition, we have compiled<sup>29</sup> 61 new tautomeric rules derived from various literature sources. These new rules consist of 34 prototropic rules (PT\_22\_00 - PT\_49\_00) plus two variants with *nn* > 00 and variants of PT\_11\_*nn* for long range hydrogen migration, where *nn* ranges from 01-04), 16 ring $\rightleftharpoons$ chain rules (RC\_03\_03, RC\_03\_04, RC\_04\_04, and RC\_12\_00 – RC\_24\_00), and 11 valence rules (VT\_01\_00, VT\_01\_01 – VT\_10\_00). (See footnotes of Table 2 for rule naming and numbering nomenclature.)

Table 2 shows the frequency of the applicability of all these rules to the entries in our database showing both the cases where the transformation between the experimental tautomers only required the application of a single rule as well as of cases that needed additional, or allowed alternative, rules in the single/multistep transformation between observed tautomers.

It needs to be pointed out that there are cases for all “New” rules in our database simply because we added to the database all the cases from the literature source(s) that actually gave rise to each of the new rules as described in the accompanying paper.<sup>29</sup> In contrast hereto, the original (contractor-based) extraction from literature had not been rule-aware. Therefore some of the standard rules, de facto being quite rare even though labeled as “Standard,” are not represented in the current version of the database (PT\_13\_00 – PT\_15\_00, PT\_17\_00 – PT\_21\_00).

**Table 2.** Frequency distribution of prototropic, ring $\rightleftharpoons$ chain, and valence tautomerism rules for single-rule transformations and those with combined/alternative rules<sup>a,b</sup>

| Type              | Standard Rules |                           |             |                              |
|-------------------|----------------|---------------------------|-------------|------------------------------|
|                   | Rule number    | Rule name                 | Single rule | Combined or alternative rule |
| Prototropic Rules | PT_02_00       | 1,5 (thio)keto/(thio)enol | 0           | 230                          |
|                   | PT_03_00       | simple (aliphatic) imine  | 0           | 323                          |

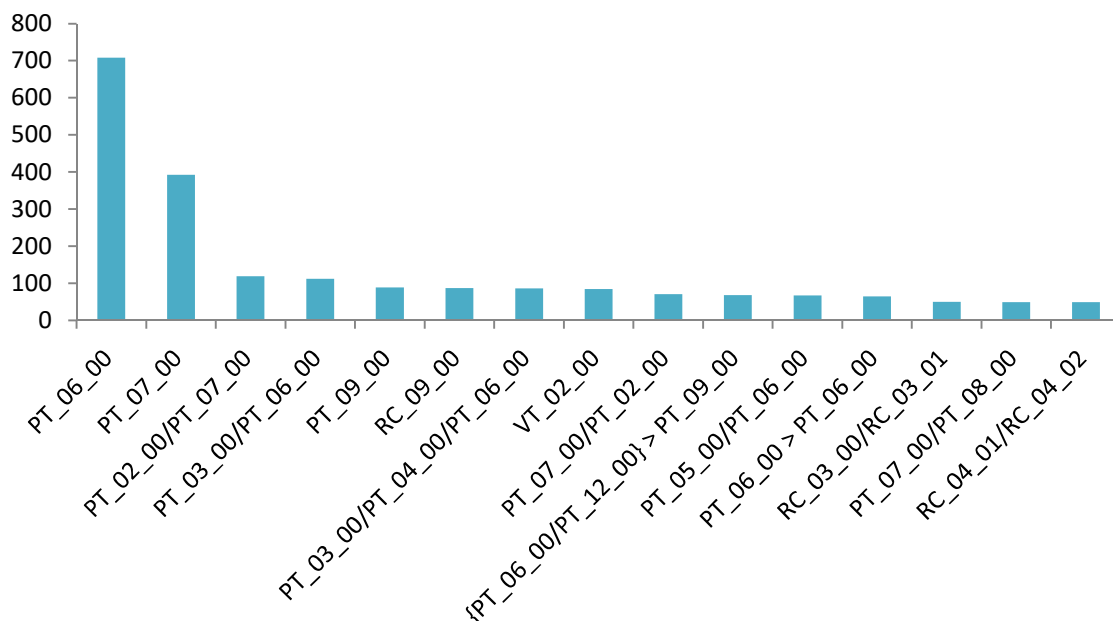
|                   |                       |  |     |     |
|-------------------|-----------------------|--|-----|-----|
|                   | PT_04_00              | special imine                                      | 0   | 127 |
|                   | PT_05_00              | 1,3 aromatic heteroatom H shift                    | 0   | 184 |
|                   | PT_06_00              | 1,3 heteroatom H shift                             | 708 | 891 |
|                   | PT_07_00              | 1,5 (aromatic) heteroatom H shift (1)              | 391 | 463 |
|                   | PT_08_00              | 1,5 (aromatic) heteroatom H shift (2)              | 0   | 88  |
|                   | PT_09_00              | 1,7 (aromatic) heteroatom H shift                  | 89  | 256 |
|                   | PT_10_00              | 1,9 (aromatic) heteroatom H shift                  | 0   | 72  |
|                   | PT_11_00 <sup>b</sup> | 1,11 (aromatic) heteroatom H shift                 | 0   | 33  |
|                   | PT_12_00              | 1,3 furanones                                      | 0   | 84  |
|                   | PT_16_00              | nitroso/oxime                                      | 0   | 14  |
| Ring-Chain Rules  | RC_03_00              | 5_exo_trig   | 0   | 50  |
|                   | RC_03_01              | 5_exo_trig   | 0   | 50  |
|                   | RC_03_02              | 5_exo_trig   | 19  | 0   |
|                   | RC_04_01              | 6_exo_trig   | 0   | 49  |
|                   | RC_04_02              | 6_exo_trig   | 0   | 49  |
|                   | RC_09_00              | 5_endo_trig  | 67  | 0   |
|                   | RC_10_00              | 6_endo_trig  | 10  | 15  |
|                   | RC_10_01              | 6_endo_trig  | 29  | 15  |
| <b>New Rules</b>  |                       |  |     |     |
| Prototropic Rules | PT_22_00              | imine/imine  | 3   | 0   |
|                   | PT_23_00              | 1,5 furanones                                      | 12  | 0   |
|                   | PT_24_00              | 1,4 N-oxide/N-hydroxide                            | 8   | 0   |
|                   | PT_25_00              | 1,6 N-oxide/N-hydroxide (1)                        | 4   | 0   |
|                   | PT_26_00              | 1,6 N-oxide/N-hydroxide (2)                        | 5   | 0   |
|                   | PT_27_00              | acene  | 13  | 0   |
|                   | PT_27_01              | thiophene analogue of acene                        | 15  |     |
|                   | PT_28_00              | nitro/aci-nitro via aromatic ring (1): 1,7 H-shift | 2   | 0   |
|                   | PT_29_00              | nitro/aci-nitro via aromatic ring (1): 1,5 H-shift | 3   | 0   |
|                   | PT_29_01              | o-tolualdehyde                                     | 2   | 0   |
|                   | PT_30_00              | nitramide/N-nitronic acid                          | 1   | 0   |
|                   | PT_31_00              | sulfone-based aliphatic compounds                  | 1   | 0   |
|                   | PT_32_00              | nitrile/ketenimine: 1,3 H-shift                    | 8   | 0   |
|                   | PT_33_00              | nitrile/ketenimine: 1,5 H-shift                    | 8   | 0   |
|                   | PT_34_00              | triad phosphorus-carbon                            | 5   | 0   |
|                   | PT_35_00              | sulfenyl/sulfinyl: 1,2 H-shift                     | 2   | 0   |
|                   | PT_36_00              | oxime/nitrone: 1,2 H-shift                         | 5   | 0   |
|                   | PT_37_00              | sulfenyl/S-oxide: 1,4 H-shift                      | 1   | 0   |
|                   | PT_38_00              | silane-hemiaminal/silanoic amide                   | 2   | 0   |

|                  |          |   |    |   |
|------------------|----------|---|----|---|
|                  | PT_39_00 | nitrono/azoxy or Behrend rearrangement                              | 19 | 0 |
|                  | PT_40_00 | tetrad phosphorus-carbon  | 1  | 0 |
|                  | PT_41_00 | pyridine 1-oxide/1-hydroxypyridine                                  | 2  | 0 |
|                  | PT_42_00 | $\Delta^3$ -/ $\Delta^4$ -pyrro(thio/seleno)lin-2-one               | 27 | 0 |
|                  | PT_43_00 | isobenzofuran/phthalan  | 4  | 0 |
|                  | PT_44_00 | 2-substituted-pyrrole   | 6  | 0 |
|                  | PT_45_00 | isopropylidenecycloalkane/<br>isopropylcycloalkene                  | 17 | 0 |
|                  | PT_46_00 | 4-picoline  | 1  | 0 |
|                  | PT_47_00 | isoindole/isoindolenine   | 24 | 0 |
|                  | PT_48_00 | benzofuranone   | 4  | 0 |
|                  | PT_49_00 | N-hydroxyindole   | 6  | 0 |
| Ring-Chain Rules | RC_03_03 | boronic acid/oxaborole  | 19 | 0 |
|                  | RC_03_04 | 5_exo_trig  | 15 | 0 |
|                  | RC_04_04 | 6_exo_trig  | 25 | 0 |
|                  | RC_12_00 | 5_endo_tet or minophosphorane<br>/benzoxazaphospholine              | 39 | 0 |
|                  | RC_13_00 | 6_endo_dig  | 1  | 0 |
|                  | RC_14_00 | thiadiazoline rearrangement   | 9  | 0 |
|                  | RC_15_00 | 5_exo_trig: 1,4 H-shift   | 3  | 0 |
|                  | RC_16_00 | boryl/borate  | 2  | 0 |
|                  | RC_17_00 | boryl/borate: ion-complex   | 2  | 0 |
|                  | RC_18_00 | 5_exo_tet or hydroxyphosphorane                                     | 4  | 0 |
|                  | RC_19_00 | nitroolifin/1,2-oxazine N-oxide                                     | 6  | 0 |
|                  | RC_20_00 | 5_endo_trig: 1,4 H-shift or aminoethyl<br>nitrono/imidazolidin-1-ol | 6  | 0 |
|                  | RC_21_00 | cyclobutane/enamine   | 3  | 0 |
|                  | RC_22_00 | 5_endo_trig: 1,5 H-shift  | 12 | 0 |
|                  | RC_23_00 | 6_endo_trig: 1,4 H-shift  | 1  | 0 |
|                  | RC_24_00 | $\lambda^5$ -/ $\lambda^3$ -phosphane                               | 2  | 0 |
| Valence Rules    | VT_01_00 | monothio-o-benzoquinone/benzoxathiete                               | 2  | 0 |
|                  | VT_01_01 | $\alpha$ -dithione and 1,2-dithiete                                 | 12 | 0 |
|                  | VT_02_00 | tetrazole/azide   | 84 | 0 |
|                  | VT_03_00 | isothiocyanate/triazinethione                                       | 8  | 0 |
|                  | VT_04_00 | tetrazine/azodiazo  | 21 | 0 |
|                  | VT_05_00 | 1,2,3-triazole/diazoamidine   | 8  | 0 |
|                  | VT_06_00 | norcaradiene/cycloheptatriene or<br>benzene-oxide/oxepin            | 18 | 0 |
|                  | VT_07_00 | phospha-münchnones  | 11 | 0 |

|  |          |                                  |    |   |
|--|----------|----------------------------------|----|---|
|  | VT_08_00 | 1,2,3,4-tetrazinium/azodiazonium | 15 | 0 |
|  | VT_09_00 | diazaphosphazole/phosphinoimine  | 25 | 0 |
|  | VT_10_00 | phosphine/phosphonium salt       | 25 | 0 |

<sup>a</sup>Different classes of tautomerism are defined by prefixing each rule with PT, RC, or VT for prototropic tautomerism, ring⇌chain tautomerism and valence tautomerism, respectively. The second placeholder in rule name between underscore indicates rule number in that category (i.e. “02” in PT\_02\_00) and the last number in name indicates a variant of that rule (i.e. “01” in VT\_01\_01, “03” in RC\_03\_03). Rule ending with “\_00” occurs only in one variant for that rule. This naming scheme allows us to add more variants in that rule if it is required in the future. <sup>b</sup>We also have four variants of PT\_11\_*nn* for long range hydrogen migration, where *nn* ranges from 01-04. <sup>b</sup>SMIRKS of these tautomeric rules are given in Table S3 of Supporting Information.

The most commonly encountered prototropic, ring⇌chain and valence rules are shown in Figure 6. The tautomeric transformation of one tautomer to another can take place in one or more steps, where each step corresponds to one rule, however in some cases these tautomeric transformation can be achieved by other rules. The majority of transformations from our database occur in a single step (60%) while the others involve the use of additional rules to complete the transformation. About 35% of transformations are achieved by the application of PT\_06\_00 and PT\_07\_00 in a single step. There are some rules (PT\_02\_00 - PT\_05\_00 , PT\_08\_00, PT\_10\_00 - PT\_16\_00 ) which do not appear in any single-step transformation, but appear several times in multistep transformations or as alternative rules to others. For example, 353 cases need additional one-step (for a total of two steps) and 27 other cases require two or more steps (for a total of three or more steps) to complete the observed tautomeric transformations.



**Figure 6.** Frequency distribution of commonly observed rules.

Generally, a hydrogen atom migrates in a tautomeric transformation from its initial position in the molecule to an odd numbered (relative) position (such as 3, 5, 7, 9, or 11), designated as “1,3H shift,” “1,5H shift” etc. In contrast hereto, migration to an even position (such as 2, 4, or 6) is rare. Analysis of the database showed that in most cases hydrogen migrates via 1,3H shift (1,120), followed by 1,5H (707) and 1,7H (91) shifts, respectively, in the single step transformations. One notes that this distance travelled by the hydrogen is well correlated with the respective H shifts frequency of observation. We also have observed cases where the 1,3H shift can be achieved alternatively via long distance migration using 1,5H (30) or 1,7H shift (118), respectively. Likewise, 1,5H shift based transformations can be in competition with 1,7H and 1,9H shifts in a single step equilibria. For 2-step transformations, we observed the order by frequency of occurrence shown in Table 3. We note that as in single-step transformations, shorter distances hydrogen migrations are more prevalent than longer ones for the 2-step transformations too.

**Table 3.** Type of hydrogen shifts observed in the one step and two steps hydrogen migrations<sup>a</sup>

| One step hydrogen migrations <sup>a</sup> | Count | Two steps hydrogen migrations <sup>b</sup> | Count |
|---|-------|--|-------|
|---|-------|--|-------|

|         |       |            |     |
|---------|-------|------------|-----|
| 1,2     | 7     | 1,3 > 1,3  | 135 |
| 1,3     | 1,120 | 1,3 > 1,7  | 91  |
| 1,4     | 12    | 1,5 > 1,5  | 42  |
| 1,5     | 707   | 1,5 > 1,11 | 27  |
| 1,6     | 15    | 1,3 > 1,5  | 5   |
| 1,7     | 91    | 1,5 > 1,7  | 3   |
| 1,3/1,5 | 30    | 1,5 > 1,3  | 2   |
| 1,3/1,7 | 118   | 1,7 > 1,7  | 1   |
| 1,5/1,9 | 15    | Others     | 48  |
| 1,5/1,7 | 2     | -          | -   |
| Others  | 609   | -          | -   |

<sup>a</sup>“/” indicates alternative H shifts possible for the same transformation. <sup>b</sup>“>” denotes a first H shift is followed by a second one to achieve the transformation.

The database contains significantly fewer cases (388) of ring⇌chain tautomerism than of prototropic tautomerism, generally belonging to cyclization to 4, 5 and 6 membered ring systems. This ring cyclization can occur either via endocyclic or exocyclic process where the double bond becomes part of the ring or of the side chain, respectively. Examples for both types of ring cyclization occur in our database. In 180 cases of endocyclic ring⇌chain transformations, ring closure happens at digonal (sp), trigonal (sp<sup>2</sup>), or tetrahedral (sp<sup>3</sup>) centers. In our recent NMR experimental study, we found 5 or 6 membered endocyclization at a trigonal center (represented by rules RC\_09\_00 and RC\_10\_00) to be the most “reliable” ring⇌chain rules in the sense that they predicted truly occurring tautomeric interconversion.<sup>22</sup> As per our new rule RC\_13\_00, 6-membered ring closure can also occur at a digonal center (sp) through an endocyclization process. Moreover, according to new rules RC\_12\_00, RC\_18\_00 and RC\_24\_00, 5-membered ring cyclization can also occur at a tetrahedral center. These three rules RC\_12\_00, RC\_18\_00, and RC\_24\_00 do not follow the concept of ring closing/opening according to Baldwin’s rules. In contrast to other rules, RC\_24\_00 involves tautomerization between trivalent (chain) and pentavalent (ring) tautomers. It is worth pointing out that during ring closure the hydrogen atom usually travels by 1,3 H shift as noted in the case of all of old, and some of the new, ring⇌chain rules. However, there are some rules that involve 1,2 H shift (i.e. RC\_24\_00), 1,4 H shift (i.e. RC\_15\_00, RC\_20\_00 and RC\_23\_00) and 1,5 H shift (RC\_22\_00) during ring closure. We thus note that experimental evidence can introduce the need

for new rules (at the chemoinformatics level) that are not covered by previous work aimed at creating comprehensive rule sets.

In 193 cases of exocyclic ring $\rightleftharpoons$ chain transformations, the ring closing/opening takes place at trigonal or tetrahedral centers. 5- or 6-membered cyclization occurs only at trigonal centers while 5-membered cyclization via RC\_18\_00 takes place only at tetrahedral centers. There are some instances of ring $\rightleftharpoons$ chain tautomerism in the thiadiazoline (RC\_14\_00), boryl/borate (RC\_16\_00 and RC\_17\_00), and  $\lambda^5/\lambda^3$ -phosphane (RC\_24\_00) systems that do not involve any unsaturated electrophilic center (or endo- or exo-cyclic bonds) during interconversions but rather involve saturated sulfur, boron, and phosphorus centers, respectively.

The ring-chain rules do not occur in combination with any prototropic or ring $\rightleftharpoons$ chain rule, i.e. in all cases transformation proceeds in a single step. Therefore, ring $\rightleftharpoons$ chain rules appear to be more selective and specific than prototropic rules. Generally, ring $\rightleftharpoons$ chain tautomerism showed a high prevalence for the chain form over of the ring form. There are 20 cases of ring $\rightleftharpoons$ chain tautomerism where three tautomers are in equilibrium with each other in solution, the two ring tautomers existing as *cis* and *trans* isomers, respectively. This provides interesting systems in which cyclic stereoisomers are in equilibrium with each other via the chain form; however, the chain form dominates over the two ring forms.

There are 228 cases of valence tautomerism in the database. They all involve ring opening/closing in 4, 5 or 6 membered ring systems without migration of any hydrogen atom. The ring-opened tautomers of four rules (VT\_02\_00, VT\_04\_00, VT\_05\_00 and VT\_08\_00) have a charge-separated moiety in their structures and this charge disappears in the ring-closed tautomers. In contrast hereto, a charge-separated moiety is present in the ring-closed tautomer of both VT\_07\_00 and VT\_10\_00. The tautomeric equilibrium via VT\_06\_00 involves ring-contraction (6 membered) and -expansion (7 membered) in the tautomers. Out of the 11 rules, VT\_09\_00 is the only one rule that involves valency change during tautomerization:, between trivalent phosphinoimine and pentavalent diazaphosphazole tautomers. Among these 11 valence tautomerism rules, our database contains significant counts only for the tetrazole $\rightleftharpoons$ azide tautomerism (VT\_02\_00). The dataset allows us to conclude that for the tetrazole $\rightleftharpoons$ azide tautomerism, the tetrazole tautomer is more favored in polar aprotic solvent while the azide tautomer is dominant in nonpolar solvent. The available dataset for benzoxathiete derivatives is

quite small, thus does not allow one to deduce any specific preference for one or the other tautomer, both being described in the available literature references.

## Type of tautomerism

Many of the transforms listed in Table 2 align quite closely with chemotypes the way the organic chemist would usually perceive them. However, others among these transforms, as they are expressed as general SMIRKS patterns<sup>{Formatting Citation}</sup>, cover a broader range of compound types. For example, transform PT\_06\_00 (1,3 heteroatom H shift), recognizes C, O, N, S, P, Se, and Te in its SMIRKS pattern, thus covering quite diverse types of compounds and tautomerism based on those. Conversely, the interconversion between the hydrazine and the azo species of a compound can be effected at the transform level by 1,3H, 1,5H and 1,7H shifts, which are encoded in different transforms. We therefore felt it to be useful to analyze the distribution of the records in our database along these more chemical types of tautomerism.

The database in this sense comprises of more than 50 types of tautomeric pairs. The most commonly encountered tautomer types are listed in Table 4 in the order of their occurrence (see molecular examples in the Supporting Information Table S4). We see that azo $\rightleftharpoons$ hydrazone tautomerism is one of the most studied types of prototropic tautomerism, followed by keto $\rightleftharpoons$ enol and oxo-enamine $\rightleftharpoons$ oxo-imine forms, respectively. The number of records for azo $\rightleftharpoons$ hydrazone is more than twice compared to the simple keto $\rightleftharpoons$ enol form. Table 5 shows commonly identified sets of three tautomers with their occurrences. There are 51 cases in which a set of geometrical tautomers of enethiol (*cis* and *trans* or *Z* and *E*) is found in equilibrium with the thioketo form. On the other hand, in 42 cases of enol-imine $\rightleftharpoons$ oxo-enamine $\rightleftharpoons$ oxo-imine tautomerism, three chemically distinct tautomers are in equilibrium. Moreover, a set of 2 similar thioimidol, enamine, enol, and keto tautomers are found to exist in equilibria with thioamide, imine, keto, and enol form, respectively. Table 6 shows the distribution of some of the common tautomers across the 5 different prevalence categories described above (0-4). To remind the reader, category 0 indicates that the tautomer was not observed or was not in the detectable range; 4 indicates that the tautomer was exclusively observed; and categories 1-3 form the class that comprises various degrees of predominance between the two extremes. There are cases where the numbers of exclusively observed tautomers are higher than the numbers in the predominant class. For example, enethiol, oxo-imine, imine (counter tautomer of enamine) and phenol-imine



are exclusively identified in 16, 9, 19, and 16 cases, respectively. In the case of keto-enethiol $\rightleftharpoons$ thioketo-enol type tautomerism, all the tautomers fall in 3 categories (1-3). Keto-enethiol, oxo-enamine, pyridol, isoindole, phosphine and thioketo are some of the tautomers for which there is no single example in prevalence category 4 in the database. This does not exclude the possibility of their observation in other experiments, however, our database does not have examples from them. As per the data in Table 6, azo, keto, oxo-enamine, keto-enol, and thioketo-enol are examples of tautomers usually preferred over their corresponding other tautomer(s). It is interesting to note that oxo-enamine can be in equilibrium with oxo-imine, enol-imine, and phenol-imine tautomers. It is usually preferred over both oxo-imine and enol-imine, however phenol-imine always dominates the oxo-enamine. Imine can be in equilibrium with enamine and amine tautomers, being typically preferred over enamine while it is less preferred over amine.

**Table 4.** Types of tautomeric equilibria in tautomeric pairs with their occurrences. (Examples for each of these types are given in Table S4 of Supporting Information.)

| Type of tautomerism   | Count | Type of tautomerism   | Count |
|---|-------|---|-------|
| Azo $\rightleftharpoons$ Hydrazone                                      | 333   | NH $\rightleftharpoons$ OH                                      | 10    |
| Ring $\rightleftharpoons$ Chain <sup>a</sup>                            | 318   | 1,2-Dihydro $\rightleftharpoons$ 1,4-Dihydro                    | 9     |
| Enol $\rightleftharpoons$ Keto  | 138   | Carbamoylimino $\rightleftharpoons$ Guanidine <sup>a</sup>      | 9     |
| Oxo-enamine $\rightleftharpoons$ Oxo-imine                              | 113   | Thiol $\rightleftharpoons$ Thione                               | 9     |
| Diketo $\rightleftharpoons$ Keto-enol                                   | 108   | Nitroso-enamine $\rightleftharpoons$ Oxim-imine                 | 7     |
| Enol-imine $\rightleftharpoons$ Oxo-enamine                             | 104   | 1,2-Dihydro $\rightleftharpoons$ 2,5-Dihydro                    | 6     |
| Amine $\rightleftharpoons$ Imine  | 83    | Cycloheptatriene $\rightleftharpoons$ Norcaradiene <sup>b</sup> | 6     |
| Keto-enethiol $\rightleftharpoons$ Thioketo-enol                        | 82    | Pyrrole $\rightleftharpoons$ Pyrrolidine                        | 6     |
| Azide $\rightleftharpoons$ Tetrazole <sup>b</sup>                       | 82    | 1,4-Dihydro $\rightleftharpoons$ 4,6-Dihydro                    | 5     |
| nul $\rightleftharpoons$ nul <sup>c</sup>                               | 78    | Nitrone $\rightleftharpoons$ Oxime                              | 5     |
| Ring $\rightleftharpoons$ Chain <sup>b</sup> (Valence)                  | 77    | Triazole $\rightleftharpoons$ Triazole                          | 4     |
| Enamine $\rightleftharpoons$ Imine                                      | 72    | 2H $\rightleftharpoons$ 4H                                      | 4     |
| Oxo-enamine $\rightleftharpoons$ Phenol-imine                           | 65    | Enol-enamine $\rightleftharpoons$ Oxo-enamine                   | 4     |
| Pyridol $\rightleftharpoons$ Pyridone                                   | 58    | Amino-thieno $\rightleftharpoons$ Imine-thieno                  | 4     |
| NH $\rightleftharpoons$ NH  | 57    | Isobenzofuran $\rightleftharpoons$ Phthalan                     | 4     |
| Phenol-quinone $\rightleftharpoons$ Phenol-quinone                      | 51    | CH $\rightleftharpoons$ OH                                      | 3     |
| Enol-imine $\rightleftharpoons$ Oxo-imine                               | 40    | 1,4-Dihydro $\rightleftharpoons$ 4,5-Dihydro                    | 3     |
| Benzoxazaphospholine $\rightleftharpoons$ Iminophosphorane <sup>a</sup> | 39    | N(1)H $\rightleftharpoons$ N(3)H                                | 3     |
| CH $\rightleftharpoons$ NH  | 35    | Amine $\rightleftharpoons$ Zwitterion                           | 3     |

|   |    |   |   |
|---|----|---|---|
| Keto-enol $\rightleftharpoons$ Keto-enol                                | 33 | Selenol $\rightleftharpoons$ Selone   | 3 |
| NH-imidazole $\rightleftharpoons$ NH-imidazole                          | 31 | Imine $\rightleftharpoons$ Imine  | 3 |
| Lactam $\rightleftharpoons$ Lactim                                      | 31 | Nitro $\rightleftharpoons$ aci-Nitro  | 3 |
| Amine-imine $\rightleftharpoons$ Amine-imine                            | 27 | 5,6-Dihydro $\rightleftharpoons$ 5,6-Dihydro                                    | 2 |
| Cyclohexadienone $\rightleftharpoons$ Phenol                            | 27 | 5,6-Dihydro-2H $\rightleftharpoons$ 5,6-Dihydro-4H                              | 2 |
| 3H-2-one $\rightleftharpoons$ 5H-2-one                                  | 27 | C1-H $\rightleftharpoons$ C3-H  | 2 |
| Enethiol $\rightleftharpoons$ Thioketo                                  | 26 | Thiol $\rightleftharpoons$ Zwitterion   | 2 |
| N-hydroxide $\rightleftharpoons$ N-oxide                                | 25 | Sulfenyl $\rightleftharpoons$ Sulfinyl  | 2 |
| Diazaphosphazole $\rightleftharpoons$ Phosphinoimine <sup>b</sup>       | 25 | Sila-hemiaminal $\rightleftharpoons$ Silanoic-amide                             | 2 |
| Phosphine $\rightleftharpoons$ Phosphonium salt <sup>b</sup>            | 25 | $\lambda^3$ -Phosphane $\rightleftharpoons$ $\lambda^5$ -Phosphane <sup>a</sup> | 2 |
| Isoindole $\rightleftharpoons$ Isoindolenine                            | 24 | 1H $\rightleftharpoons$ 2H  | 1 |
| 1,4-Dihydro $\rightleftharpoons$ 1,6-Dihydro                            | 19 | 2H $\rightleftharpoons$ 2H  | 1 |
| Nitrone $\rightleftharpoons$ Nitrone                                    | 19 | 1,6-Dihydro $\rightleftharpoons$ 3,6-Dihydro                                    | 1 |
| Isopropylcycloalkene $\rightleftharpoons$ Isopropyliden<br>ecycloalkane | 17 | 4H $\rightleftharpoons$ 6H  | 1 |
| Thioamide $\rightleftharpoons$ Thioimidol                               | 16 | Nitroso-imine $\rightleftharpoons$ Oxim-imine                                   | 1 |
| Keteneimine $\rightleftharpoons$ Nitrile                                | 16 | C3-H $\rightleftharpoons$ N(5)H   | 1 |
| Tropolone $\rightleftharpoons$ Tropolone                                | 12 | Oxo-thione $\rightleftharpoons$ nul <sup>c</sup>                                | 1 |
| 2H $\rightleftharpoons$ 6H  | 12 | Pyridol $\rightleftharpoons$ Zwitterion   | 1 |
| Amide $\rightleftharpoons$ Imidol                                       | 12 | 1H $\rightleftharpoons$ 3H  | 1 |
| Amino $\rightleftharpoons$ Imino  | 12 | N-nitronic acid $\rightleftharpoons$ Nitramide                                  | 1 |
| Arene-imine $\rightleftharpoons$ Azepine <sup>b</sup>                   | 12 | Enol $\rightleftharpoons$ Ylide   | 1 |
| Anaquinoid $\rightleftharpoons$ Paraquinonimine                         | 10 | S-oxide $\rightleftharpoons$ Sulfenyl   | 1 |

<sup>a</sup> Ring-chain tautomerism type (Total count for ring-chain tautomerism of two tautomers including benzoxazaphospholine $\rightleftharpoons$ iminophosphorane, Carbamoylimino $\rightleftharpoons$ Guanidine, and  $\lambda^5$ -Phosphane $\rightleftharpoons$  $\lambda^3$ -Phosphane pairs is 368).

<sup>b</sup> Valence tautomerism type.

<sup>c</sup> "nul" indicates cases of tautomeric equilibria for which no name for one or the other or both tautomer was given in the references and we were not able to assign any specific name.

**Table 5.** Types of tautomeric equilibria in three-tautomer sets with their occurrences

| Type of tautomerism  | Count | Type of tautomerism   | Count |
|--|-------|---|-------|
| Enethiol $\rightleftharpoons$ Enethiol $\rightleftharpoons$ Thioketo                     | 51    | Enamine $\rightleftharpoons$ Imine $\rightleftharpoons$ nul                             | 6     |
| Enol-imine $\rightleftharpoons$ Oxo-enamine  | 42    |   |       |
| $\rightleftharpoons$ Oxo-imine   |       | Enamine $\rightleftharpoons$ Enamine $\rightleftharpoons$ Imine                         | 6     |
| Phenol-quinone $\rightleftharpoons$ Phenol-quinone $\rightleftharpoons$ Phenol-quinone   | 25    | 1,2-Dihydro $\rightleftharpoons$ 1,4-Dihydro $\rightleftharpoons$ 1,5-Dihydro           | 5     |
| CH $\rightleftharpoons$ NH $\rightleftharpoons$ OH                                       | 21    | 1,7-Dihydro-7-oxo $\rightleftharpoons$ 4,7-Dihydro-7-oxo $\rightleftharpoons$ 7-Hydroxy | 5     |
| Chain $\rightleftharpoons$ Ring $\rightleftharpoons$ Ring                                | 20    | Nitroso-enamine $\rightleftharpoons$ Nitroso-imine $\rightleftharpoons$ Oxim-imine      | 4     |
| 5-Hydroxytriazine $\rightleftharpoons$ Orthoquinonoid $\rightleftharpoons$ Paraquinonoid | 11    | Enol $\rightleftharpoons$ Keto $\rightleftharpoons$ Zwitterion                          | 4     |
| Thioamide $\rightleftharpoons$ Thioimidol $\rightleftharpoons$ Thioimidol                | 11    | Triazole $\rightleftharpoons$ Triazole $\rightleftharpoons$ Triazole                    | 4     |
| nul $\rightleftharpoons$ nul $\rightleftharpoons$ nul <sup>a</sup>                       | 9     | Azo $\rightleftharpoons$ Hydrazone $\rightleftharpoons$ Zwitterion                      | 3     |
| Enol $\rightleftharpoons$ Enol $\rightleftharpoons$ Keto                                 | 6     | Diketo $\rightleftharpoons$ Keto-enol $\rightleftharpoons$ Keto-enol                    | 3     |
| Enol $\rightleftharpoons$ Keto $\rightleftharpoons$ Keto                                 | 6     | Others  | 8     |

<sup>a</sup> See Table 4.

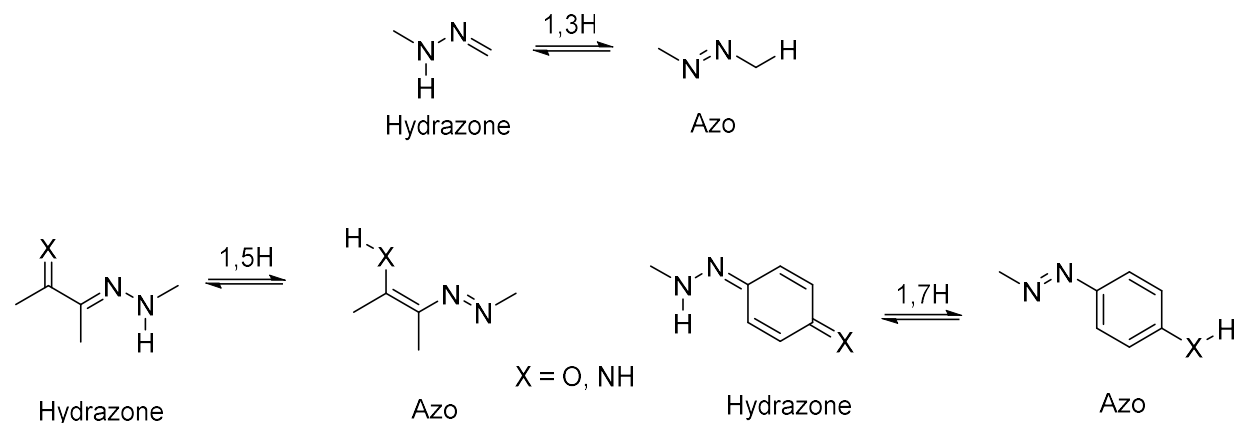
Closely following the generic ring $\rightleftharpoons$ chain case count of 388 (Table 4 and 5), the database contains 333 records for the hydrazone $\rightleftharpoons$ azo type of tautomerism. Additionally, there are 3 cases where a zwitterionic species competes with both of the above-mentioned tautomers. Most of the studies were carried out in methanol (82) or DMSO (68), mainly using <sup>1</sup>H NMR (109) and UV/VIS (88) spectroscopic methods. This tautomerism can involve 1,3H, 1,5H, or 1,7H shift. In the simplest case, the hydrogen is hopping between carbon and nitrogen atoms via 1,3H shift as shown in Figure 8. If the X position in that figure is O, then the azo form is termed hydroxy-azo or enol. Its hydrazone counter tautomer is found with several different names in the literature such as quinone-hydrazone, keto-hydrazone, oxo-hydrazone, or keto. If X is NH, then this pair is described in the literature by the terms azo and quinoimine-hydrazone tautomers, respectively. Of these, 145 cases included a single-step migration described by transforms PT\_07\_00 (99), PT\_09\_00 (79), or PT\_06\_00 (17) while others involved two or more steps of migration. The numbers in Table 6 show that the azo form was more favored in 122 cases whereas the

hydrazone form was found to be more dominant in 72 cases as defined by prevalence category 3 and 4 collectively.

**Table 6.** Distribution of some of the common tautomeric pairs in different prevalence categories

| Tautomer_1                        | Prevalence_Category |     |     |    |    | Tautomer_2                    | Prevalence_Category |     |     |    |    |
|-----------------------------------|---------------------|-----|-----|----|----|-------------------------------|---------------------|-----|-----|----|----|
|                                   | 0                   | 1   | 2   | 3  | 4  |                               | 0                   | 1   | 2   | 3  | 4  |
| Azo                               | 39                  | 73  | 99  | 79 | 43 | Hydrazone                     | 54                  | 127 | 80  | 38 | 34 |
| Enol                              | 63                  | 35  | 20  | 14 | 5  | Keto                          | 20                  | 34  | 10  | 43 | 30 |
| Oxo-enamine                       | 9                   | 2   | 37  | 44 | 21 | Oxo-imine                     | 25                  | 70  | 7   | 2  | 9  |
| Diketo                            | 27                  | 41  | 9   | 25 | 6  | Keto-enol                     | 6                   | 37  | 22  | 39 | 4  |
| Enol-imine                        | 6                   | 53  | 0   | 29 | 16 | Oxo-enamine                   | 16                  | 29  | 2   | 51 | 6  |
| Keto-enethiol                     | 0                   | 59  | 19  | 3  | 0  | Thioketo-enol                 | 0                   | 4   | 19  | 58 | 0  |
| Amine                             | 10                  | 32  | 5   | 28 | 8  | Imine                         | 18                  | 33  | 5   | 24 | 3  |
| Enamine                           | 19                  | 13  | 26  | 9  | 5  | Imine                         | 5                   | 14  | 26  | 8  | 19 |
| Oxo-enamine                       | 16                  | 15  | 34  | 0  | 0  | Phenol-imine                  | 0                   | 0   | 34  | 15 | 16 |
| Pyridol                           | 3                   | 38  | 8   | 8  | 0  | Pyridone                      | 2                   | 15  | 12  | 27 | 2  |
| Enethiol                          | 0                   | 6   | 3   | 1  | 16 | Thioketo                      | 18                  | 1   | 4   | 3  | 0  |
| Enol-imine                        | 6                   | 53  | 0   | 29 | 16 | Oxo-enamine                   | 16                  | 29  | 2   | 51 | 6  |
| Lactam                            | 10                  | 1   | 4   | 15 | 1  | Lactim                        | 4                   | 11  | 5   | 9  | 1  |
| 5H-2-one                          | 2                   | 4   | 2   | 13 | 6  | 3H-2-one                      | 6                   | 13  | 2   | 4  | 2  |
| Cyclohexadienone                  | 9                   | 1   | 5   | 4  | 8  | Phenol                        | 9                   | 3   | 5   | 0  | 10 |
| Isoindole                         | 1                   | 8   | 5   | 10 | 0  | Isoindolenine                 | 0                   | 10  | 5   | 8  | 1  |
| N-hydroxide                       | 1                   | 3   | 5   | 9  | 1  | N-oxide                       | 1                   | 10  | 5   | 2  | 1  |
| Keteneimine                       | 5                   | 11  | 0   | 0  | 0  | Nitrile                       | 0                   | 6   | 0   | 9  | 1  |
| Ring <sup>a</sup>                 | 28                  | 125 | 114 | 64 | 7  | Chain <sup>a</sup>            | 11                  | 65  | 171 | 65 | 26 |
| Benzoxazaphospholine <sup>a</sup> | 0                   | 11  | 17  | 10 | 1  | Iminophosphorane <sup>a</sup> | 1                   | 10  | 17  | 11 | 0  |
| Diazaphosphazole <sup>b</sup>     | 2                   | 4   | 1   | 17 | 1  | Phosphinoimine <sup>b</sup>   | 1                   | 17  | 1   | 4  | 2  |
| Phosphine <sup>b</sup>            | 3                   | 9   | 5   | 8  | 0  | phosphonium salt <sup>b</sup> | 4                   | 4   | 5   | 12 | 0  |
| Ring <sup>b</sup>                 | 6                   | 22  | 8   | 26 | 15 | Chain <sup>b</sup>            | 18                  | 28  | 8   | 17 | 6  |
| Tetrazole <sup>b</sup>            | 5                   | 29  | 8   | 35 | 5  | Azide <sup>b</sup>            | 4                   | 70  | 1   | 3  | 4  |

<sup>a,b</sup> See Table 4. Category values: 0 = not observed or not in the detectable range; 4 = exclusively observed.



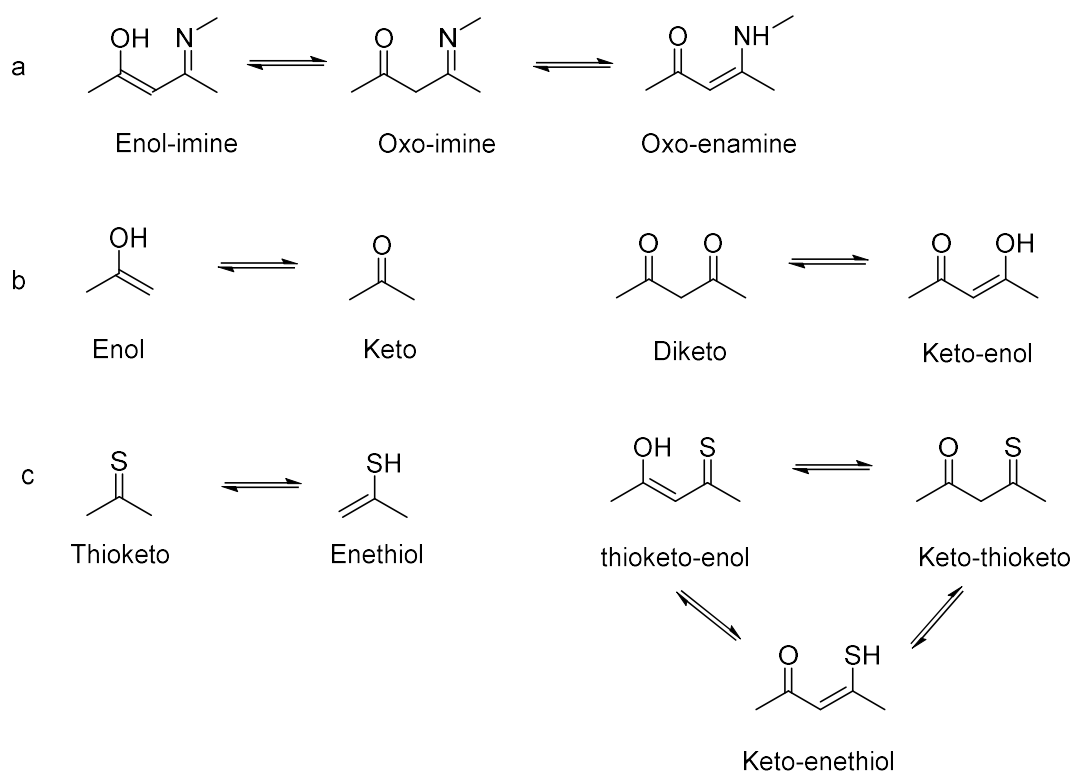
**Figure 8.** Schematic representation of some equilibria of the azo-hydrazone type tautomerism.

The database contains 257 cases where any two tautomers from the following three tautomers were in equilibrium with each other: enol-imine, oxo-enamine and oxo-imine (Figure 9a). For example, oxo-enamine $\rightleftharpoons$ oxo-imine, enol-imine $\rightleftharpoons$ oxo-enamine, and enol-imine $\rightleftharpoons$ oxo-imine type equilibria are present in the database in 113, 104, and 40 cases, respectively. These cases were mainly observed using  $^1\text{H}$  NMR, UV/VIS, and mass spectroscopic methods in chloroform, ethanol, acetonitrile, and gas phase. They can involve 1,3H, 1,5H, or 1,9H shifts in a single-step or multi-step transformation with rules PT\_02\_00, PT\_03\_00, PT\_04\_00, PT\_05\_00, PT\_06\_00, PT\_07\_00, PT\_09\_00, or PT\_10\_00. Based on prevalence category, we observe that oxo-enamine is found to be predominant or exclusively present in 65 cases, whereas oxo-imine is observed as predominant or exclusively present in 11 cases. The equilibrium between enol-imine and oxo-enamine was quite competitive, either state being observed predominantly or exclusively in 45 and 57 cases, respectively. The enol-imine is generally preferred in molecules in which the tautomeric skeleton is part of aromatic (benzene) or heteroaromatic (pyridine) ring systems. However, both tautomers are found with almost equal preference in molecules containing a hydroxyl group on the naphthalene ring, with only a few exceptions. Finally, in the third pair, enol-imine and oxo-imine were observed in prevalence category 3 and 4 in 12 and 19 examples, respectively.

Simple keto $\rightleftharpoons$ enol is the second most commonly observed type of tautomerism (137 pairs) after azo $\rightleftharpoons$ hydrazone tautomerism. They were mostly studied in chloroform (53) and gas phase (25), and the rest in about 10 different solvents. We find equilibria for 133 and 3 cases based on PT\_06\_00 and PT\_07\_00, respectively. Among these pairs, the keto form was exclusively

observed, or more stable, in 73 cases, whereas the enol form was found to be prevalent in 19 cases.  $\beta$ -dicarbonyl compounds can exist in the form of diketo and keto-enol type tautomers (Figure 9b). The acidity of such compounds is higher compared to monocarbonyl compounds hence are more prone to tautomerism. The database contains 108 pairs of diketo $\rightleftharpoons$ keto-enol type tautomerism, which were mainly reported in chloroform, gas phase, and DMSO using  $^1\text{H}$  NMR,  $^{13}\text{C}$  NMR and gas chromatography/mass spectroscopy methods. The keto-enol form was observed predominantly or exclusively in 43 cases, while the diketo form was favored in 31 cases. This higher preference for the keto-enol tautomer may be due to possibility of hydrogen bond stabilization of the keto-enol form and the presence of resonance-stabilized conjugated  $\pi$ -systems. In addition, in these molecules, the position of the tautomeric equilibrium is also affected by the presence of electron withdrawing and donating substituents at the one  $\beta$ - and the two  $\alpha$ -carbons. The database contains 33 pairs of keto-enol $\rightleftharpoons$ keto-enol type tautomers that exist in equilibrium with other tautomer(s) (mainly observed in chloroform) via 1,5H shift as per rule PT\_07\_00. Such pairs could have been in equilibrium with a third diketo tautomer via 1,3H shift; however, we have only 2 cases where such two keto-enol tautomers were found in equilibrium with the diketo form.

The keto-enethiol $\rightleftharpoons$ thioketo-enol type of tautomerism appears in monothio analogues of  $\beta$ -dicarbonyl compounds (Figure 9c). Analysis of 82 cases shows that the thioketo-enol tautomer is predominant mainly in acetonitrile (14), cyclohexane (43) and gas phase (1), and its counter tautomer is favorable in acetonitrile (3) and in the gas phase (1). Out of 26 cases of the enethiol $\rightleftharpoons$ thioketo type tautomerism in simple thioketo molecules, the enethiol was predominant in 17 cases. It is interesting to note that in thio analogues of  $\beta$ -dicarbonyl compounds the thioketo-enol tautomer is favorable while in simple thioketo molecules the enethiol is favored.

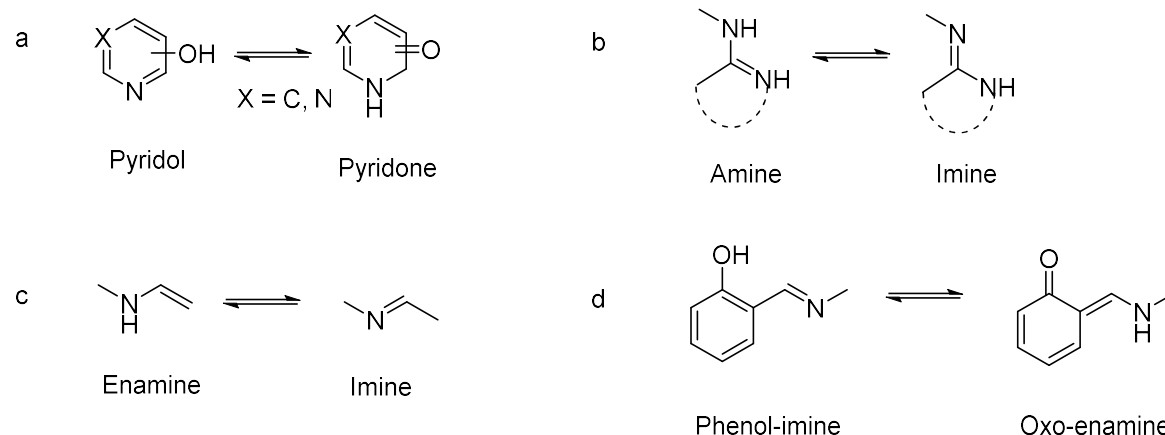


**Figure 9.** Schematic representation of (a) enol-imine $\rightleftharpoons$ oxo-imine $\rightleftharpoons$ oxo-enamine tautomerism (b) simple keto $\rightleftharpoons$ enol and diketo $\rightleftharpoons$ keto-enol tautomerism (c) simple thioketo $\rightleftharpoons$ enethiol and thioketo-enol $\rightleftharpoons$ keto-thioketo $\rightleftharpoons$ keto-enethiol tautomerism.

Pyridone $\rightleftharpoons$ pyridol type tautomerism (58 pairs in 2-pyridone, 4-pyridone and 4-pyrimidinone derivatives) (Figure 10a) was studied at least in 10 different media including solvents, vapor phase, solid phase, liquid phase, N<sub>2</sub>/Ar matrix, KBr, and others. In these molecules, the pyridone type tautomer was found to be predominant in 29 cases, mainly in cyclohexane, chloroform and acetonitrile solvents while its counter tautomer pyridol was found to be favorable only in 8 cases in polar and nonpolar solvents.

The amine $\rightleftharpoons$ imine type of tautomerism (83) involves hopping of hydrogen between nitrogen atoms (Figure 10b) where one of the nitrogen atoms usually is present in an aromatic or alicyclic ring and the other in a side chain of the molecule. These molecules were studied using mainly dipole moment measurement, IR spectra and <sup>15</sup>N NMR spectra in at least 7 media. The transformation between tautomers mostly occurs via 1,3H shift in a single step and in a few cases in two steps. Molecules with the aromatic ring involving such tautomerism usually show a strong predominance for the amine form. On the other hand, molecules with alicyclic rings were found

to exist predominantly in either the imine or the amine form. Overall, the database shows that the amine is favorable in 36 cases while the imine is predominant or exclusive in 27 cases irrespective of molecule types.



**Figure 10.** Schematic representation of (a) pyridol $\rightleftharpoons$ pyridone tautomerism (b) amine $\rightleftharpoons$ imine tautomerism (c) enamine $\rightleftharpoons$ imine and (d) phenol-imine $\rightleftharpoons$ oxa-enamine tautomerism.

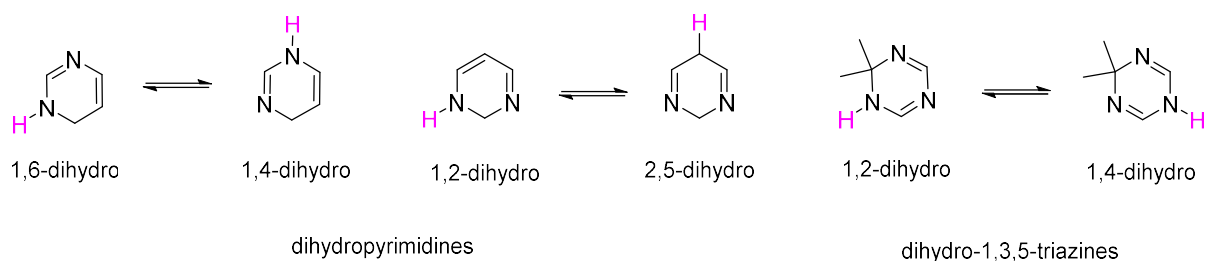
The database contains 72 tautomeric pairs of enamine $\rightleftharpoons$ imine type tautomerism (Figure 10c), which were mainly reported in TFA, DMSO and chloroform by  $^1\text{H}$  and/or  $^{13}\text{C}$  NMR spectroscopy. The imine form was predominant or exclusively observed in 27 cases (mainly in TFA and chloroform) while enamine form was favorable in 14 cases (mainly in chloroform and DMSO). Further, in a subset of 26 molecules out of the 72 pairs both tautomers were observed as major species.

A dataset of 65 pairs of oxo-enamine $\rightleftharpoons$ phenol-imine type tautomerism (Figure 10d) was mainly studied with  $^1\text{H}$  and/or  $^{13}\text{C}$  NMR spectroscopy or combination of those with  $^{17}\text{O}$  and  $^{15}\text{N}$  NMR spectroscopy in 5 different solvents. Chloroform (32) and acetone (11) were the main solvents for such tautomerism. In the literature, phenol-imines are often called by names such as hydroxy-imine or enol-imine, which have by other authors been given to counter tautomers of other types of tautomers such as oxo-imines. The phenol-imine tautomer was observed predominantly or exclusively in 31 cases, while its counterpart oxo-enamine was not found to be predominant in even a single case.

There are some dihydropyrimidines derivatives that show equilibrium between 1,4- and 1,6-dihydro forms, and 1,2- and 2,5-dihydro forms (Figure 11). The insertion of one more nitrogen



atom into dihydropyrimidine derivatives leads to dihydro-1,3,5-triazines that show equilibrium between 1,2- and 1,4-dihydro forms, where the 1,4-dihydro form predominates in solution. The keto group-containing triazine derivatives are supposed to exist in three tautomeric forms, namely: paraquinonoid, orthoquinonoid and 5-hydroxytriazine. Our data indicates predominance of the paraquinonoid form over orthoquinonoid and 5-hydroxytriazine.



**Figure 11.** Schematic representation of tautomers of diazine and triazine derivatives.

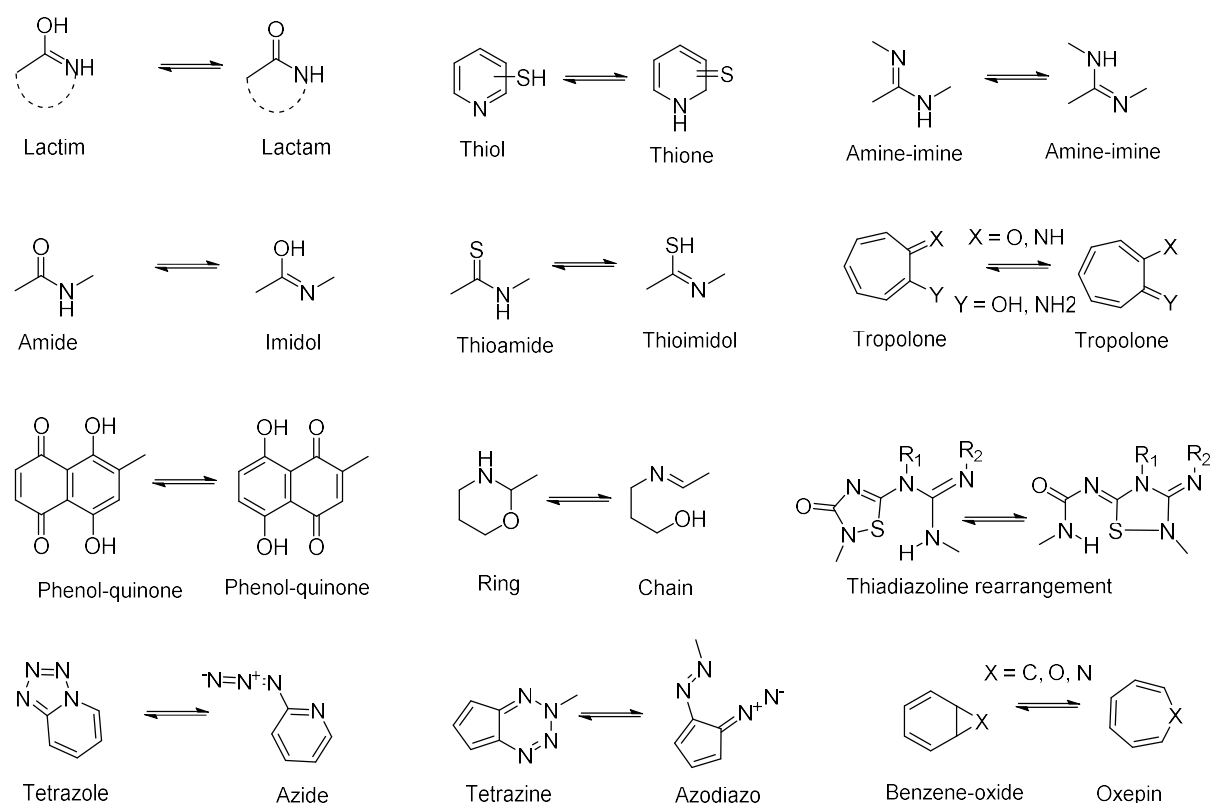
The migration of a hydrogen atom within a heterocyclic ring has been described by terms like CH, NH, and OH forms that are involved in the migration. Analysis of the 105 cases of isoxazol-5(2H)-one, oxadiazole-5-one, pyrazol-3-one and related derivatives in the database shows that these molecules can exist in NH, CH and OH forms and can transform from the CH form to NH and OH forms by 1,3H shift. The CH form was found to be more predominant in 27 cases. The database contains 57 cases where one NH form is in equilibrium with another NH form by 1,5H shift (annular tautomerism of pyrazoles).

In our database, 1,3H shift of the amide bond in nonaromatic heterocyclic ring is considered in the lactam $\rightleftharpoons$ lactim interconversion (Figure 12). This tautomerism was mainly investigated using electron ionization/mass spectra and UV spectra in gas phase and water. Out of 31 pairs, lactam tautomers were found to be predominant in 16 cases, while lactim tautomers were found predominantly or exclusively in 10 cases. The tautomers of the amide and thioamide bond in acyclic compounds are described by amide $\rightleftharpoons$ imidol and thioamide $\rightleftharpoons$ thioimidol tautomerism, respectively. In phenol-quinone $\rightleftharpoons$ phenol-quinone type tautomerism, the hydrogen migrates from one oxygen to another oxygen, so these tautomers cannot be differentiated for their preferences; still, we have some examples in the database where both identical tautomers were reported to be in an equilibrium. Similarly, in tropolone-based molecules, hydrogen migrates between nitrogen atoms or oxygen atoms via a long conjugation path generating similar tautomers, i.e. without a

change of chemotype. In amino-substituted tropolone molecules, the hydrogen shift occurs between oxygen and nitrogen atoms and in all cases, nitrogen in amino form and oxygen in keto form were exclusively observed in the examples in our database.

The database contains 388 examples of ring $\rightleftharpoons$ chain tautomerism covered by 24 rules, where 368 and 20 cases are from tuples of 2 and 3 tautomers, respectively. Of these, 85 pairs belong to 5\_endo\_trig (RC\_09\_00, RC\_20\_00, and RC\_22\_00), followed by 6\_exo\_trig (80 pairs from RC\_04\_01/RC\_04\_02, RC\_04\_04, and RC\_19\_00) and 5\_endo\_tet (39 pairs for RC\_12\_00) types, respectively. Their ring $\rightleftharpoons$ chain equilibria were explored using about 30 media. Chloroform (113) was the most frequently used medium followed by DMSO, acetone, and carbon tetrachloride. There are 22 examples that were studied in their neat liquid phase, i.e. without using any solvent. In addition, water mixed with DMSO, dioxane, ethanol, or acetone was used as the solvent in 51 cases.

Finally, our database contains 228 examples of valence tautomerism from 11 rules. Out of those, 82 pairs belong to azide $\rightleftharpoons$ tetrazole tautomerism, followed by diazaphosphazole $\rightleftharpoons$ phosphinoimine (25) and phosphine $\rightleftharpoons$ phosphonium salt (25) types, respectively. The tetrazole form was observed predominantly or exclusively in 40 cases while azide form was favored in 7 cases. The equilibrium between them was examined in 11 solvents or their mixtures, where DMSO (35) and chloroform (15) were used in most of the cases.  $^1\text{H}$  NMR was extensively used while in other cases  $^{13}\text{C}$ ,  $^{17}\text{O}$ ,  $^{14}\text{N}$  and/or  $^{15}\text{N}$  NMR spectroscopy was used for the equilibrium studies. Except benzene-oxide $\rightleftharpoons$ oxepin tautomerism, all types of valence tautomerism include an equilibrium between ring and chain type tautomers. Out of those 209 examples, ring valence tautomers were observed predominantly or exclusively in 111 cases while the chain form was found to be predominant in 44 cases.



**Figure 12.** Schematic representation of tautomeric equilibria of some other compounds including examples of ring-chain and valence tautomerism.

### Scoring for canonical tautomers

We analyzed two-tautomer tuples (2,530) to check how likely CACTVS determines the canonical form of a given experimentally determined tautomers via the scoring scheme<sup>21</sup> developed for the NCI/CADD calculable chemical structure identifiers.<sup>25</sup> The current scoring scheme was able to consistently predict the right tautomer in 26% of the cases while in 36% it consistently failed to predict the correct tautomer. The most important aspect is to determine always the same tautomer regardless of which tautomer was used as the starting input to algorithm. In contradiction to this, in the remaining 38% of the cases the current scoring scheme predicted two different canonical tautomers for a given tautomer tuple. We mention this brief experiment here to illustrate that this database may open up a significant tool to implement, or

improve currently implemented, tautomer scoring schemes to empirically predict a canonical tautomer.

## SUMMARY AND CONCLUSIONS

A significant variety of structures, chemotypes, analytical procedures, and experimental conditions including solvents has been compiled to form the Tautomer Database. It contains, at the time of this writing, a set of 2,819 molecules extracted from experimental literature as tautomeric pairs or higher-order tuples. We have associated them with prototropic, ring $\rightleftharpoons$ chain, or valence tautomeric transform rules, found applicable in 79%, 13%, and 8% of the cases, respectively. As expected,  $^1\text{H}$  and  $^{13}\text{C}$  NMR methods were most frequently used in experimental tautomer equilibrium studies. Out of more than 40 solvents reported in the publications, chloroform, DMSO and ethanol were the most commonly used solvents. In solvent mixtures, water with ethanol or DMSO was commonly used. The temperature of the experiment was reported for 1,389 cases, out of these 50% and 32% studies were carried out in the range of 251-300K and 301-350K, respectively. No more than 100 entries had pH details reported, with most of those entries coming from a pH range of 0-7 mainly in methanol and water solvents.

At the chemoinformatics level, most of the examples of prototropic tautomerism are covered directly or alternatively by 42 different prototropic transforms whereas the examples for ring $\rightleftharpoons$ chain and valence tautomerism are covered by 24 and 11 transforms, respectively. The latter examples are highly specific, i.e. the database does not contain examples that are amenable to transforms of prototropic, ring $\rightleftharpoons$ chain, and valence tautomerism combined.

We hope that this database of experimental data and its included analysis by chemoinformatics methods (by way of annotation with tautomeric transform rules) may provide a set of data useful for future work in the field of tautomerism. This would include tools such as software and chemical identifiers that could be used to avoid tautomeric duplication in chemical databases

and compound registration systems. We also hope it may help in developing approaches to predict the most “medicinally” relevant and “reasonable” tautomer forms. This dataset could be a useful training set for machine learning models based on quantum mechanics<sup>32,33</sup> to rapidly identify the lowest energy tautomer.

## **ACKNOWLEDGEMENTS**

We have to send copious thanks to Wolf-Dietrich Ihlenfeldt for his initial work with CACTVS and its treatment of tautomerism, as well as for his support in our generating and testing the new rules. We gratefully acknowledge Thomas Sander and Oya Wahl for providing us with a copy of their Tautomer Codex database, which helped in the generation of a handful of additional rules. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). This work was supported by the Intramural Research Program of the National Institutes of Health, Center for Cancer Research, National Cancer Institute. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

## **SUPPORTING INFORMATION**

There are 4 tables in the SI (contained in one Word document): List of the publications used in tautomer database generation (Table S1). Distribution of solvents or their mixtures, or general experimental environments, by spectroscopic methods (Table S2). SMIRKS of tautomeric rules (Table S3). Representative examples of chemical types of tautomerism (Table S4). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## **AUTHOR INFORMATION**

### **Corresponding Author**

\* Computer-Aided Drug Design Group, Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, NIH, Frederick, MD 21702, USA

### Present Addresses

Laura Guasch: Roche Pharmaceutical Research and Early Development, Roche Innovation Center, Basel CH-4070, Switzerland

### ORCID

Devendra K. Dhaked: 0000-0001-6349-7649

Laura Guasch: 0000-0003-4990-7765

Marc C. Nicklaus: 0000-0002-4775-7030

### Competing interests

The authors declare that they have no competing interests.

### REFERENCES

- (1) Guasch, L.; Sitzmann, M.; Nicklaus, M. C. Enumeration of Ring–Chain Tautomers Based on SMIRKS Rules. *J. Chem. Inf. Model.* **2014**, *54* (9), 2423–2432. <https://doi.org/10.1021/ci500363p>.
- (2) Murguly, E.; Norsten, T. B.; Branda, N. Tautomerism of 4-Hydroxyterpyridine in the Solid, Solution and Gas Phases: An X-Ray, FT-IR and NMR Study. *J. Chem. Soc. Perkin Trans. 2* **1999**, No. 12, 2789–2794. <https://doi.org/10.1039/a904323c>.
- (3) Lulinki, S.; Madura, I.; Serwatowski, J.; Szatyłowicz, H.; Zachara, J. A Tautomeric Equilibrium between Functionalized 2-Formylphenylboronic Acids and Corresponding 1,3-Dihydro-1,3-Dihydroxybenzo [c][2,1]Oxaboroles. **2006**. <https://doi.org/10.1039/b611195e>.
- (4) Llinàs, A.; Goodman, J. M. Polymorph Control: Past, Present and Future. *Drug Discov. Today* **2008**, *13* (5–6), 198–210. <https://doi.org/10.1016/j.drudis.2007.11.006>.
- (5) Censi, R.; Di Martino, P. Polymorph Impact on the Bioavailability and Stability of Poorly Soluble Drugs. *Molecules* **2015**, *20* (10), 18759–18776. <https://doi.org/10.3390/molecules201018759>.
- (6) McMorro, D.; Aartsma, T. J. Solvent-Mediated Proton Transfer. The Roles of Solvent Structure and Dynamics on the Excited-State Tautomerization of 7-Azaindole/Alcohol Complexes. *Chem. Phys. Lett.* **1986**, *125* (5–6), 581–585. [https://doi.org/10.1016/0009-2614\(86\)87103-2](https://doi.org/10.1016/0009-2614(86)87103-2).
- (7) Wu, Y.-S.; Huang, H.-C.; Shen, J.-Y.; Tseng, H.-W.; Ho, J.-W.; Chen, Y.-H.; Chou, P.-T. Water-Catalyzed Excited-State Proton-Transfer Reactions in 7-Azaindole and Its

- Analogues. *J. Phys. Chem. B* **2015**, *119* (6), 2302–2309.  
<https://doi.org/10.1021/jp506136v>.
- (8) Schlabach, M.; Limbach, H. H.; Bunnenberg, E.; Shu, A. Y. L.; Tolf, B. R.; Djerassi, C. NMR Study of Kinetic HH/HD/DH/DD Isotope Effects on the Tautomerism of Acetylporphyrin: Evidence for a Stepwise Double Proton Transfer. *J. Am. Chem. Soc.* **1993**, *115* (11), 4554–4565. <https://doi.org/10.1021/ja00064a018>.
  - (9) Aguilar-Parrilla, F.; Scherer, G.; Limbach, H. H.; Foces-Foces, M. de la C.; Hernandez Cano, F.; Smith, J. A. S.; Toiron, C.; Elguero, J. Observation of a Series of Degenerate Cyclic Double, Triple, and Quadruple Proton Transfers in Solid Pyrazoles. *J. Am. Chem. Soc.* **1992**, *114* (24), 9657–9659. <https://doi.org/10.1021/ja00050a055>.
  - (10) Martin, Y. C. Let's Not Forget Tautomers. *J. Comput. Aided Mol. Des.* **2009**, *23* (10), 693–704. <https://doi.org/10.1007/s10822-009-9303-2>.
  - (11) Masand, V. H.; Mahajan, D. T.; Gramatica, P.; Barlow, J. Tautomerism and Multiple Modelling Enhance the Efficacy of QSAR: Antimalarial Activity of Phosphoramidate and Phosphorothioamidate Analogues of Amiprophos Methyl. *Med. Chem. Res.* **2014**, *23* (11), 4825–4835. <https://doi.org/10.1007/s00044-014-1043-8>.
  - (12) Milletti, F.; Vulpetti, A. Tautomer Preference in PDB Complexes and Its Impact on Structure-Based Drug Discovery. *J. Chem. Inf. Model.* **2010**, *50* (6), 1062–1074. <https://doi.org/10.1021/ci900501c>.
  - (13) Kalliokoski, T.; Salo, H. S.; Lahtela-Kakkonen, M.; Poso, A. The Effect of Ligand-Based Tautomer and Protomer Prediction on Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49* (12), 2742–2748. <https://doi.org/10.1021/ci900364w>.
  - (14) Oellien, F.; Cramer, J.; Beyer, C.; Ihlenfeldt, W.-D.; Selzer, P. M. The Impact of Tautomer Forms on Pharmacophore-Based Virtual Screening †. *J. Chem. Inf. Model.* **2006**, *46* (6), 2342–2354. <https://doi.org/10.1021/ci060109b>.
  - (15) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
  - (16) The Cambridge Crystallographic Data Centre (CCDC). <https://www.ccdc.cam.ac.uk/>.
  - (17) Borbulevych, O.; Martin, R. I.; Tickle, I. J.; Westerhoff, L. M. XModeScore: A Novel Method for Accurate Protonation/Tautomer-State Determination Using Quantum-Mechanically Driven Macromolecular X-Ray Crystallographic Refinement. *Acta Crystallogr. Sect. Struct. Biol.* **2016**, *72* (4), 586–598. <https://doi.org/10.1107/S2059798316002837>.
  - (18) Trepalin, S. V.; Skorenko, A. V.; Balakin, K. V.; Nasonov, A. F.; Lang, S. A.; Ivashchenko, A. A.; Savchuk, N. P. Advanced Exact Structure Searching in Large Databases of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 852–860. <https://doi.org/10.1021/ci025582d>.
  - (19) Cruz-Cabeza, A. J.; Groom, C. R. Identification, Classification and Relative Stability of Tautomers in the Cambridge Structural Database. *CrystEngComm* **2011**, *13* (1), 93–98. <https://doi.org/10.1039/C0CE00123F>.
  - (20) Milletti, F.; Storch, L.; Sforna, G.; Cross, S.; Cruciani, G. Tautomer Enumeration and Stability Prediction for Virtual Screening on Large Chemical Databases. *J. Chem. Inf. Model.* **2009**, *49* (1), 68–75. <https://doi.org/10.1021/ci800340j>.
  - (21) Sitzmann, M.; Ihlenfeldt, W.-D.; Nicklaus, M. C. Tautomerism in Large Databases. *J. Comput. Aided Mol. Des.* **2010**, *24* (6–7), 521–551. <https://doi.org/10.1007/s10822-010-9346-4>.

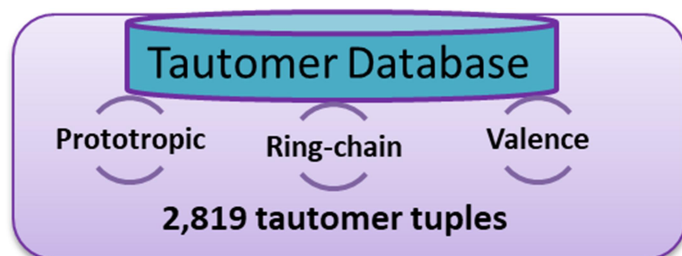


- (22) Guasch, L.; Yapamudiyansel, W.; Peach, M. L.; Kelley, J. A.; Barchi, J. J.; Nicklaus, M. C. Experimental and Chemoinformatics Study of Tautomerism in a Database of Commercially Available Screening Samples. *J. Chem. Inf. Model.* **2016**, *56* (11), 2149–2161. <https://doi.org/10.1021/acs.jcim.6b00338>.
- (23) Warr, W. A. Tautomerism in Chemical Information Management Systems. *J. Comput. Aided Mol. Des.* **2010**, *24* (6–7), 497–520. <https://doi.org/10.1007/s10822-010-9338-4>.
- (24) Gimadiev, T. R.; Madzhidov, T. I.; Nugmanov, R. I.; Baskin, I. I.; Antipin, I. S.; Varnek, A. Assessment of Tautomer Distribution Using the Condensed Reaction Graph Approach. *J. Comput. Aided Mol. Des.* **2018**, *32* (3), 401–414. <https://doi.org/10.1007/s10822-018-0101-6>.
- (25) Sitzmann, M.; Filippov, I. V.; Nicklaus, M. C. Internet Resources Integrating Many Small-Molecule Databases 1. *SAR QSAR Environ. Res.* **2008**, *19* (1–2), 1–9. <https://doi.org/10.1080/10629360701843540>.
- (26) Xemistry Chemoinformatics. <http://www.xemistry.com/>.
- (27) Sitzmann, M.; Filippov, I. V.; Nicklaus, M. C. Internet Resources Integrating Many Small-Molecule Databases 1. *SAR QSAR Environ. Res.* **2008**, *19* (1–2), 1–9. <https://doi.org/10.1080/10629360701843540>.
- (28) IUPAC projects [https://iupac.org/projects/project-details/?project\\_nr=2012-023-2-800](https://iupac.org/projects/project-details/?project_nr=2012-023-2-800).
- (29) Dhaked, D. K.; Guasch, L.; Nicklaus, M. C. Tautomer Database: A Comprehensive Resource for Tautomerism Including InChI V2. *J. Chem. Inf. Model.* (*In press*).
- (30) Daylight Theory Manual <http://www.daylight.com/dayhtml/doc/theory/theory.smirks.html>.
- (31) Dhaked, D. K.; Ihlenfeldt, W.-D.; Patel, H.; Nicklaus, M. C. Toward a Comprehensive Treatment of Tautomerism in Chemoinformatics Including InChI V2. *J. Chem. Inf. Model.* (*In press*).
- (32) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203. <https://doi.org/10.1039/C6SC05720A>.
- (33) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A Data Set of 20 Million Calculated off-Equilibrium Conformations for Organic Molecules. *Sci. Data* **2017**, *4* (1). <https://doi.org/10.1038/sdata.2017.193>.

**For Table of Contents Use Only**

## **Tautomer Database: A Comprehensive Resource for Tautomerism Analyses**

Devendra K. Dhaked, Laura Guasch, Marc C. Nicklaus<sup>\*</sup>



Tautomer\_Database\_MCN\_V8\_2019-11-22.pdf (1.74 MiB)

[view on ChemRxiv](#) • [download file](#)

---

**Table S1:** Publications used in tautomer database generation (new rules' references are 2, 3, 10, 44, 91-92, and 98-171).

| <b>No.</b> | <b>Publication</b>                                   | <b>Records extracted</b> |
|------------|--|--------------------------|
| 1          | Prog. NMR. Spect., 2006, 49, 169-206                 | 113                      |
| 2          | Adv. Heterocycl. Chem., 2001, 81, 253-303            | 76                       |
| 3          | Heterocycles, 1995, 41(8), 1805-1832                 | 57                       |
| 4          | Heterocycles, 1995, 41(9), 2057-2093                 | 57                       |
| 5          | Heterocycles, 1986, 24(8), 2321-2355                 | 87                       |
| 6          | J. Org. Chem., 2001, 66(12), 4132-4140               | 59                       |
| 7          | J. Chem. Soc. Perkin Trans. 2, 1989, 1613-1616       | 42                       |
| 8          | J. Organomet. Chem., 1977, 129, 41-53                | 45                       |
| 9          | Heterocycles, 1991, 32(2), 329-369                   | 46                       |
| 10         | J. Am. Chem. Soc., 1989, 111(1), 258-268             | 3                        |
| 11         | J. Chem. Soc., Faraday Trans. 2, 1986, 82, 2379-2392 | 5                        |
| 12         | J. Am. Chem. Soc., 1981, 103 (6), 1350-1353          | 4                        |
| 13         | J. Chem. Soc., Perkin Trans. 2, 1985, 1255-1264      | 4                        |
| 14         | Dyes Pigm., 2000, 47, 45-51                          | 5                        |
| 15         | Pol. J. Chem., 1981, 55, 1441-1444                   | 4                        |
| 16         | Mol. Phy., 2004, 102 (23-24), 2607-2615              | 7                        |
| 17         | Helv. Chim. Acta, 2010, 93, 466-481                  | 14                       |
| 18         | Magn. Reson. Chem., 1985, 23(9), 790-792             | 4                        |
| 19         | Tetrahedron, 1964, 20, 2835-2840                     | 8                        |
| 20         | J. Chem. Soc., Perkin Trans. 2, 1985b, 1265-1274     | 16                       |
| 21         | Russ. J. Org. Chem., 2002, 38(6), 851-854            | 10                       |
| 22         | Rapid Commun. Mass Spectrom., 2006, 20, 2548-2552    | 10                       |
| 23         | Spectrosc. Lett., 1989, 22(2), 123-160               | 10                       |
| 24         | Bull. Acad. Polon. Sci., 1973, XXI(11), 859-871      | 11                       |
| 25         | J. Chem. Soc., Perkin Trans. 2, 1993, 1067-1071      | 12                       |
| 26         | Tetrahedron, 1965, 21, 1681-1692                     | 7                        |
| 27         | Appl. Spectrosc., 1982, 36(6), 682-684               | 17                       |
| 28         | Org. Biomol. Chem., 2009, 7, 4110-4119               | 12                       |
| 29         | Bull. Korean Chem. Soc. 2011, 32, 1873-1878          | 10                       |
| 30         | J. Mol. Struct., 2000, 532, 139-142                  | 13                       |
| 31         | Tetrahedron, 1981, 37(15), 2633-2640                 | 13                       |
| 32         | J. Am. Chem. Soc., 1976, 98(16), 4736-4745           | 15                       |
| 33         | Tetrahedron, 1982, 38(24), 3679-3686                 | 23                       |
| 34         | Gazz. Chim. Ital., 1986, 116, 579-582                | 13                       |
| 35         | Eur. J. Mass Spectrom., 2006, 12, 25-29              | 14                       |
| 36         | J. Chem. Soc., Perkin Trans. 2, 1980, 1768-1773      | 14                       |
| 37         | Org. Magn. Reson., 1980, 13(5), 330-334              | 17                       |
| 38         | Magn. Reson. Chem., 1998, 36, 343-350                | 6                        |

|    |  |    |
|----|--|----|
| 39 | J. Am. Chem. Soc., 1973, 95(25), 8364-8371         | 14 |
| 40 | Org. Mass Spectrom., 1987, 22, 493-500             | 14 |
| 41 | Org. Magn. Reson., 1979, 12(6), 379-382            | 11 |
| 42 | Magn. Reson. Chem., 1998, 36, 315-324              | 16 |
| 43 | J. Chem. Res., (S), 1988, 236-237                  | 12 |
| 44 | J. Mol. Struct., 1999, 510, 165-178                | 17 |
| 45 | Tetrahedron, 1972, 28, 455-462                     | 10 |
| 46 | J. Mol. Struct., 2000, 499, 121-126                | 14 |
| 47 | J. Chem. Soc. Perkin Trans. 2, 1995, 1901-1907     | 17 |
| 48 | Arch. Pharm. (Weinheim), 1992, 325, 541-543        | 20 |
| 49 | Arkivoc, 2003, xv, 134-142                         | 22 |
| 50 | Pol. J. Chem., 1981, 55, 1809-1824                 | 20 |
| 51 | Russ. Chem. Bull., 2003, 52(1), 198-207            | 23 |
| 52 | J. Heterocyc. Chem., 1994, 31, 233-238             | 27 |
| 53 | Dyes Pigm., 1990, 13, 107-116                      | 23 |
| 54 | J. Chem. Soc., Perkin Trans. 2, 1993, 1447-1454    | 23 |
| 55 | Org. Mass Spectrom., 1981, 16(8), 347-350          | 11 |
| 56 | Spectrochim. Acta, Part A, 1985, 41 (1/2), 109-114 | 34 |
| 57 | Tetrahedron, 1961, 12, 41-50                       | 22 |
| 58 | Eur. J. Org. Chem., 2011, 4367-4372                | 14 |
| 59 | Tetrahedron, 1962, 18, 777-790                     | 43 |
| 60 | J. Am. Chem. Soc., 1986, 108(4), 630-638           | 64 |
| 61 | Tetrahedron, 1978, 34, 1377-1380                   | 29 |
| 62 | Spectrochim. Acta, Part A, 1990, 46(5), 803-808    | 34 |
| 63 | J. Mol. Struct., 2006, 790, 80-88                  | 18 |
| 64 | Eur. J. Mass Spectrom., 2007, 13, 291-306          | 15 |
| 65 | Can. J. Chem., 1988, 66, 1141-1146                 | 32 |
| 66 | J. Chem. Soc., Perkin Trans. 2, 1985, 1423-1427    | 18 |
| 67 | J. Heterocyc. Chem., 1997, 34, 305-309             | 12 |
| 68 | Adv. Heterocycl. Chem., 2000, 77, 1-50             | 52 |
| 69 | Russ. J. Gen. Chem., 2003, 73(10), 1595-1602       | 31 |
| 70 | Tetrahedron, 1972, 28(24), 5923-5947               | 36 |
| 71 | Org. Magn. Reson., 1984, 22(1), 16-23              | 36 |
| 72 | Color. Technol., 2016, 132, 387-398                | 16 |
| 73 | ChemPhysChem., 2015, 16, 649 -657                  | 3  |
| 74 | Photochem. Photobiol. Sci., 2010,9, 901-907        | 2  |
| 75 | Dyes Pigm., 2012, 94, 443-451                      | 5  |
| 76 | Dalton Trans., 2012, 41, 11107-11115               | 17 |
| 77 | Dyes Pigm., 2017, 137, 101-110                     | 68 |
| 78 | RSC Adv., 2015, 5, 102495-102507                   | 30 |
| 79 | Spectrochim. Acta, Part A, 2014, 132, 815-820      | 7  |
| 80 | J. Chem. Soc., Perkin Trans. 2, 2000, 1259-1266    | 12 |
| 81 | J. Chem. Soc., Perkin Trans. 2, 1994, 2461-2470    | 29 |

|     |   |    |
|-----|---|----|
| 82  | Eur. J. Org. Chem. 2009, 2009, 1377-1383                          | 21 |
| 83  | J. Mol. Struct., 2017, 1134, 546-551                              | 8  |
| 84  | Open J. Phys. Chem., 2013, 3, 138-149                             | 22 |
| 85  | Can. J. Chem., 1991, 69, 696-701                                  | 12 |
| 86  | Org. Magn. Reson. 1999, 37(4), 259-268                            | 43 |
| 87  | Tetrahedron, 1988, 44(11), 3319-3326                              | 22 |
| 88  | J. Chem. Soc., Perkin Trans. 1, 1991, (11), 2831-2840             | 8  |
| 89  | J. Mol. Struct. 2013, 1032, 138-146                               | 19 |
| 90  | Tetrahedron 2015, 71, 2349-2356                                   | 15 |
| 91  | Beilstein J. Org. Chem. 2017, 13, 2396-2407                       | 10 |
| 92  | Tetrahedron, 2016, 72(16), 1919-1927                              | 37 |
| 93  | Can. J. Chem. 1971, 49 (23), 3799-3806                            | 77 |
| 94  | Dyes Pigm., 1989, 10 (1), 33-45                                   | 56 |
| 95  | Ring-Chain Tautomerism; Springer US: Boston, MA, 1985; pp 169-209 | 22 |
| 96  | Adv. Heterocycl. Chem. 1996, 66, 1-71                             | 62 |
| 97  | J. Org. Chem. 1971, 36 (10), 1352-1355                            | 19 |
| 98  | Eur. J. Org. Chem. 2002, 2002 (3), 557-563                        | 3  |
| 99  | Tetrahedron Lett. 2009, 50 (1), 12-14                             | 7  |
| 100 | Tetrahedron Lett. 1998, 39 (16), 2433-2436                        | 5  |
| 101 | Magn. Reson. Chem. 1985, 23 (3), 181-184                          | 6  |
| 102 | J. Phys. Chem. A 2004, 108 (51), 11241-11248                      | 2  |
| 103 | Magn. Reson. Chem. 1987, 25 (8), 721-724                          | 3  |
| 104 | J. Org. Chem. 1982, 47 (22), 4323-4326                            | 1  |
| 105 | J. Am. Chem. Soc. 1999, 121 (43), 9977-9985                       | 4  |
| 106 | Tetrahedron Lett. 2016, 57 (24), 2624-2627                        | 5  |
| 107 | Chem. Ber. 1977, 110 (12), 3817-3828                              | 39 |
| 108 | Bull. Chem. Soc. Jpn. 1988, 61 (2), 501-504                       | 2  |
| 109 | Helv. Chim. Acta 2001, 84 (6), 1441-1458                          | 3  |
| 110 | J. Am. Chem. Soc. 1991, 113, 172-179                              | 1  |
| 111 | Acta Part A Mol. Biomol. Spectrosc. 2010, 77 (2), 367-373         | 8  |
| 112 | Tetrahedron 1973, 29 (24), 4103-4109                              | 6  |
| 113 | Int. J. Spectrosc. 2009, 2009, 1-18                               | 2  |
| 114 | Russ. Chem. Rev. 1983, 52 (11), 1012-1029                         | 6  |
| 115 | J. Am. Chem. Soc. 1974, 96 (11), 3502-3511                        | 12 |
| 116 | J. Org. Chem. 2000, 65 (22), 7284-7290                            | 19 |
| 117 | J. Org. Chem. 1985, 50 (14), 2593-2595                            | 2  |
| 118 | J. Mol. Struct. 1994, 323, 85-91                                  | 4  |
| 119 | Magn. Reson. Chem. 2011, 49 (5), 205-212                          | 1  |
| 120 | J. Phys. Chem. A 2016, 120 (21), 3767-3779                        | 1  |
| 121 | J. Org. Chem. 1975, 40 (17), 2504-2508                            | 10 |
| 122 | Chem. Pharm. Bull. 1985, 33 (3), 1297-1300                        | 9  |
| 123 | Tetrahedron Lett. 2002, 43 (15), 2741-2743                        | 3  |

|     |  |    |
|-----|--|----|
| 124 | Tetrahedron 2005, 61 (30), 7294-7303   | 21 |
| 125 | Chem. Heterocycl. Compd. 2006, 42 (2), 280-281   | 4  |
| 126 | Adv. Heterocycl. Chem. 1994, 61, 59-141  | 4  |
| 127 | Russ. Chem. Bull. 2006, 55 (7), 1239-1242  | 8  |
| 128 | Chem. Ber. 1994, 127 (8), 1479-1488  | 21 |
| 129 | New J. Chem. 2007, 31 (1), 144-154   | 13 |
| 130 | Fluor. Chem. 2016, 187, 1-8  | 6  |
| 131 | J. Org. Chem. 1970, 35 (10), 3444-3448   | 8  |
| 132 | J. Am. Chem. Soc., 1978, 100 (16), 5229-5230   | 4  |
| 133 | J. Am. Chem. Soc. 2010, 132 (20), 6912-6913  | 2  |
| 134 | Eur. J. Org. Chem. 2008, 2008 (23), 4025-4034  | 15 |
| 135 | Org. Biomol. Chem. 2015, 13 (45), 11026-11038  | 6  |
| 136 | Chem. Ber. 1989, 122 (9), 1765-1776  | 12 |
| 137 | J. Org. Chem. 2010, 75 (12), 4261-4273   | 11 |
| 138 | Croat. Chem. Acta 1986, 59(1), 165-170   | 4  |
| 139 | Tetrahedron Lett. 1978, 19 (26), 2339-2340   | 2  |
| 140 | Tetrahedron Lett. 2007, 48 (29), 5151-5155   | 6  |
| 141 | Tetrahedron Lett. 2011, 52 (23), 3029-3032   | 3  |
| 142 | Eur. J. Org. Chem. 2002, 2002 (22), 3821-3826  | 15 |
| 143 | Russ. J. Org. Chem. 2005, 41 (10), 1421-1429   | 12 |
| 144 | J. Org. Chem. 1979, 44 (11), 1819-1823   | 1  |
| 145 | J. Chem. Soc. 1960, 2937   | 2  |
| 146 | J. Org. Chem., 1982, 47, 705-709   | 6  |
| 147 | J. Org. Chem., 1971, 36(26), 3999-4004   | 6  |
| 148 | J. Org. Chem., 1971, 36(26), 4004-4012   | 3  |
| 149 | Adv. Phys. Org. Chem., 1999, 32, 219-265   | 2  |
| 150 | Org. Magn. Reson., 1971, 3(1), 7-22  | 2  |
| 151 | Acta Chem. Scand. B, 1976, 30, 101-108   | 9  |
| 152 | J. Heterocycl. Chem., 1995, 32, 771-776  | 9  |
| 153 | The Chemistry of Pyrroles: Organic Chemistry: A Series of Monograph, 2013, 34, 379-424 | 5  |
| 154 | J. Org. Chem., 1979, 44(15), 2798-2800   | 2  |
| 155 | Tetrahedron Lett., 1976, 17(29), 2507-2510   | 1  |
| 156 | J. Org. Chem., 1988, 53(9), 1841-1848  | 3  |
| 157 | Chem. Ber., 1978, 111, 2401 -2406  | 6  |
| 158 | J.Chem. Soc., 1962, 3090-3092  | 1  |
| 159 | J. Chem. Soc. B, 1970, 1761-1766   | 17 |
| 160 | Chem. Ber., 1974, 107, 2976-2984   | 7  |
| 161 | Helv. Chim. Acta, 2001, 84(12), 3796-3809  | 1  |
| 162 | J. Chem. Soc. C, 1970, 1251-1257   | 11 |
| 163 | J. Am. Chem. Soc., 1964, 86(19), 4152-4158   | 11 |
| 164 | J. Org. Chem., 1978, 43(23), 4469-4472   | 2  |
| 165 | Chem. Heterocycl. Compd., 1989, 25(1), 18-20   | 4  |

|     |  |    |
|-----|--|----|
| 166 | Tetrahedron Lett., 1980, 21(3), 281-284        | 4  |
| 167 | J. Org. Chem., 1983, 48, 3639-3644             | 2  |
| 168 | Inorg. Chem., 2019, 58(5), 3502-3508           | 2  |
| 169 | Angew. Chem. Int. Ed., 2008, 47(45), 8674-8677 | 8  |
| 170 | PhD diss., Durham University, 2017             | 17 |
| 171 | Bull. Chem. Soc. Jpn., 2012, 85(1), 110-123    | 25 |

**Table S2:** Distribution of solvents or their mixtures, or general experimental environments, by spectroscopic methods.

| Method             | Solvent or solvent mixture; or environment | Count |
|--------------------|--|-------|
| <sup>1</sup> H NMR | Chloroform                                 | 279   |
|                    | DMSO                                       | 227   |
|                    | Carbon tetrachloride                       | 79    |
|                    | TFA  | 77    |
|                    | Acetonitrile                               | 59    |
|                    | Acetone                                    | 38    |
|                    | Benzene                                    | 38    |
|                    | Cyclohexane                                | 34    |
|                    | Dichloromethane                            | 25    |
|                    | Neat liquid                                | 22    |
|                    | Methanol                                   | 14    |
|                    | DMSO, TFA                                  | 13    |
|                    | Tetrachloroethylene                        | 10    |
|                    | TFA, Chloroform                            | 9     |
|                    | Acetone, Water                             | 8     |
|                    | Toluene                                    | 7     |
|                    | Diethyl ether                              | 6     |
|                    | Dioxane                                    | 6     |
|                    | Pyridine                                   | 6     |
|                    | THF  | 6     |
|                    | Water                                      | 6     |
|                    | Polysol                                    | 5     |
|                    | Carbon disulfide                           | 4     |
|                    | DMF, Acetone                               | 4     |
|                    | DMSO, Water                                | 4     |
|                    | 1,1,2,2-Tetrachloroethane                  | 3     |
|                    | Ethanol                                    | 3     |
|                    | Aqueous hydrochloric acid                  | 2     |
|                    | Chloroform or DMSO                         | 2     |



|                           |   |     |
|---------------------------|---|-----|
|                           | DMF                                     | 2   |
|                           | DMSO or Acetonitrile                    | 2   |
|                           | Gas phase                               | 2   |
|                           | HMPA                                    | 2   |
|                           | 2,6-Dimethylpyridine                    | 1   |
|                           | Carbon tetrachloride or Dichloromethane | 1   |
|                           | Dichloromethane or Chloroform           | 1   |
|                           | Dimethyl sulfone                        | 1   |
|                           | Formamide                               | 1   |
|                           | nul                                     | 1   |
|                           | Tetramethylenesulfone                   | 1   |
|                           | Tetramethylurea                         | 1   |
|                           | Trichloroethylene                       | 1   |
| <b><sup>13</sup>C NMR</b> | Chloroform                              | 182 |
|                           | DMSO                                    | 84  |
|                           | Acetone                                 | 14  |
|                           | Methanol                                | 13  |
|                           | DMF                                     | 9   |
|                           | Acetonitrile                            | 9   |
|                           | Carbon tetrachloride                    | 9   |
|                           | Dioxane                                 | 7   |
|                           | HMPT                                    | 6   |
|                           | DMSO, Chloroform                        | 3   |
|                           | nul                                     | 1   |
|                           | THF                                     | 1   |
|                           | Dichloromethane                         | 1   |
|                           | THF, Dichloromethane                    | 1   |
| <b><sup>31</sup>P NMR</b> | Acetone                                 | 22  |
|                           | Toluene                                 | 20  |
|                           | Benzene                                 | 12  |
|                           | THF                                     | 11  |
|                           | Chloroform                              | 6   |
|                           | Hexane                                  | 6   |
|                           | Dichloromethane                         | 4   |
|                           | Diethyl ether                           | 3   |
|                           | DMF                                     | 3   |
|                           | Pyridine                                | 3   |
|                           | DMF or Benzene                          | 2   |
|                           | Methyl acetate                          | 2   |
|                           | Triethylamine                           | 2   |
|                           | Acetonitrile                            | 1   |
|                           | Bromobenzene                            | 1   |
|                           | Chloromethane                           | 1   |

|                           |                                |    |
|---------------------------|--------------------------------|----|
|                           | DMSO                           | 1  |
|                           | Ethyl acetate                  | 1  |
|                           | Nitrobenzene                   | 1  |
|                           | Tri-n-propylamine              | 1  |
| <b><sup>15</sup>N NMR</b> | DMSO                           | 20 |
|                           | DMSO                           | 20 |
|                           | Methanol                       | 3  |
|                           | Solid state                    | 2  |
|                           | 2,2,2-Trifluoroethanol         | 1  |
|                           | Formamide                      | 1  |
|                           | Chloroform or DMSO             | 1  |
| <b>UV</b>                 | Ethanol                        | 76 |
|                           | Water                          | 31 |
|                           | Water, Ethanol                 | 27 |
|                           | Dichloromethane                | 19 |
|                           | Cyclohexane                    | 16 |
|                           | Water, DMSO                    | 15 |
|                           | Water, Dioxane                 | 14 |
|                           | DMSO                           | 7  |
|                           | Dioxane                        | 6  |
|                           | Methanol                       | 6  |
|                           | Acetonitrile                   | 5  |
|                           | Chloroform                     | 5  |
|                           | Ethyl acetate                  | 4  |
|                           | Acetone                        | 4  |
|                           | Formamide                      | 4  |
|                           | Vapor phase                    | 4  |
|                           | DMF                            | 3  |
|                           | Isooctane                      | 3  |
|                           | H <sub>2</sub> SO <sub>4</sub> | 2  |
|                           | Gas phase                      | 1  |
|                           | Chloroform or Benzene          | 1  |
|                           | 2-Propanol                     | 1  |
| <b>UV/VIS</b>             | Methanol                       | 89 |
|                           | Acetonitrile                   | 8  |
|                           | Ethanol, Water                 | 7  |
|                           | Ethyl acetate                  | 5  |
|                           | Acetone                        | 5  |
|                           | Dichloromethane                | 5  |
|                           | Toluene                        | 5  |
|                           | Chloroform                     | 5  |
|                           | Ethanol                        | 4  |
|                           | DMSO, Water                    | 2  |

|                              |                                    |    |
|------------------------------|------------------------------------|----|
|                              | DMF                                | 2  |
|                              | DMSO                               | 2  |
| <b>Electronic absorption</b> | Ethanol                            | 18 |
|                              | Cyclohexane                        | 14 |
|                              | Methanol                           | 9  |
|                              | Ethanol or Methanol or Acetic acid | 1  |
|                              | Chlorobenzene                      | 1  |
|                              | Water                              | 1  |
|                              | 1,2,4-Trichlorobenzene             | 1  |
| <b>IR</b>                    | Chloroform                         | 59 |
|                              | Nujol                              | 38 |
|                              | KBr                                | 37 |
|                              | DMSO                               | 7  |
|                              | Solid state                        | 8  |
|                              | Tetrachloroethylene                | 4  |
|                              | Dichloromethane                    | 4  |
|                              | KBr or Nujol                       | 2  |
|                              | Hexane                             | 2  |
|                              | NaCl                               | 2  |
|                              | Vaseline (Mull)                    | 2  |
|                              | Methanol                           | 1  |
|                              | Carbon tetrachloride               | 1  |
|                              | Gas phase                          | 1  |
|                              | Hexachlorobutadiene (Mull)         | 1  |
|                              | Liquid phase                       | 1  |
|                              | Chloroform or DMSO                 | 1  |
|                              | Ar or N <sub>2</sub> matrix        | 1  |
|                              | Vapor phase                        | 1  |

**Table S3:** List of tautomeric SMIRKS<sup>a</sup>.

| <b>SMIRKS</b>  | <b>Rule number</b> |
|--|--------------------|
| [O,S,Se,Te;X1:1]=[Cz1H0:2][C:5]=[C:6][CX4z0,NX3:3][#1:4]>>[#1:4]             | PT_02_00           |
| [O,S,Se,Te;X2:1][Cz1:2]=[C:5][C:6]=[Cz0,N:3]                                 |                    |
| [#1,a,O:5][NX2:1]=[Cz{1-2}:2][CX4R{0-2}:3][#1:4]>>[#1,a,O:5][NX3:1]([#1:4])  | PT_03_00           |
| [Cz:2]=[C:3]   |                    |
| [Cz0R0X3:1]([C:5])=[C:2][Nz0:3][#1:4]>>[#1:4][Cz0R0X4:1]([C:5])[c:2]=[nz0:3] | PT_04_00           |
| [#1:4][N:1][C;e6:2]=[O,NX2:3]>>[NX2,nX2:1]=[C,c;e6:2][O,N:3][#1:4]           | PT_05_00           |
| [CX{2-3}z{0-1},N,n,S,s,O,o,Se,Te:1]=[NX2,nX2,CX3,c,P,p:2][N,n,S,O,Se,Te:3]   | PT_06_00           |
| [#1:4]>>[#1:4][CX4z{0-1},N,n,S,O,Se,Te:1][NX2,nX2,CX3z{0-                    |                    |

|  |          |
|--|----------|
| 1},c,P,p:2]=[N,n,S,s,O,o,Se,Te:3]  |          |
| [nX2,NX2,S,O,Se,Te:1]=[C,c,nX2,NX2:6][C,c:5]=[C,c,nX2:2]<br>[N,n,S,s,O,o,Se,Te:3][#1:4]>>[#1:4][N,n,S,O,Se,Te:1][C,c,nX2,NX2:6]=[C,c:5]<br>[C,c,nX2:2]=[NX2,S,O,Se,Te:3]   | PT_07_00 |
| [n,s,o:1]=[c,n:6][c:5]=[c,n:2][n,s,o:3][#1:4]>>[#1:4][n,s,o:1][c,n:6]=[c:5]<br>[c,n:2]=[n,s,o:3]   | PT_08_00 |
| [nX2,NX2,S,O,Se,Te,Cz0X3:1]=[c,C,NX2,nX2:6]<br>[C,c,NX2,nX2:5]=[C,c,NX2,nX2:2][C,c,NX2,nX2:7]=[C,c,NX2,nX2:8]<br>[N,n,S,s,O,o,Se,Te,CX4z0:3][#1:4]>>[#1:4][N,n,S,O,Se,Te,Cz0X4:1]<br>[C,c,NX2,nX2:6]=[C,c:5][C,c,NX2,nX2:2]=[C,c,NX2,nX2:7]<br>[C,c,NX2,nX2:8]=[NX2,S,O,Se,Te,CX3z0:3]   | PT_09_00 |
| [#1:1][n,N,O:2][c,nX2,C:3]=[c,nX2,C:4][c,nX2:5]=[c,nX2:6][c,nX2:7]=[c,nX2:8]<br>[c,nX2,C:9]=[n,N,O:10]>>[N,n,O:2]=[C,c,nX2:3][c,nX2:4]=[c,nX2:5]<br>[c,nX2:6]=[c,nX2:7][c,nX2:8]=[c,nX2:9][n,O:10][#1:1]   | PT_10_00 |
| [#1:1][n,N,O:2][c,nX2,C:3]=[c,nX2,C:4][c,nX2:5]=[c,C,nX2:6]<br>[c,C,nX2:7]=[c,C,nX2:8][c,nX2,C:9]=[c,C,nX2:10]<br>[c,C,nX2:11]=[nX2,NX2,O:12]>>[NX2,nX2,O:2]=[C,c,nX2:3]<br>[c,C,nX2:4]=[c,C,nX2:5][c,C,nX2:6]=[c,C,nX2:7][c,C,nX2:8]=[c,C,nX2:9]<br>[c,C,nX2:10]=[c,C,nX2:11][nX2,O:12][#1:1]   | PT_11_00 |
| [#1:1][n,N,O:2][c,nX2,C:3]=[c,nX2,C:4][c,nX2,C:5]=[c,nX2,C:6]<br>[c,nX2:7]=[c,C,nX2:8][c,C,nX2,NX2:9]=[c,C,nX2:10][c,nX2,C:11]=[c,C,nX2:12]<br>[c,C,nX2:13]=[nX2,NX2,O:14]>>[NX2,nX2,O:2]=[c,nX2,C:3]<br>[c,nX2,C:4]=[c,C,nX2:5][c,C,nX2:6]=[c,C,nX2,NX2:7]<br>[c,C,nX2:8]=[c,C,nX2,NX2:9][c,C,nX2:10]=[c,C,nX2:11]<br>[c,C,nX2:12]=[c,C,nX2:13][nX2,O:14][#1:1]   | PT_11_01 |
| [#1:1][n,N,O:2][c,nX2,C:3]=[c,nX2,C:4][c,nX2,C:5]=[c,nX2,C,NX2:6]<br>[c,nX2:7]=[c,C,nX2:8][c,C,nX2,NX2:9]=[c,C,nX2,NX2:10]<br>[c,nX2,C:11]=[c,C,nX2:12][c,C,nX2:13]=[c,C,nX2:14]<br>[c,C,nX2,NX2:15]=[nX2,NX2,O:16]>>[NX2,nX2,O:2]=[c,nX2,C:3]<br>[c,nX2,C:4]=[c,C,nX2:5][c,C,nX2,NX2:6]=[c,C,nX2,NX2:7]<br>[c,C,nX2:8]=[c,C,nX2,NX2:9][c,C,nX2,NX2:10]=[c,C,nX2:11]<br>[c,C,nX2:12]=[c,C,nX2:13][c,C,nX2:14]=[c,C,nX2,NX2:15][nX2,O,N:16][#1:1] | PT_11_02 |
| [#1:1][n,N,O:2][c,nX2,C:3]=[c,nX2,C:4][c,nX2,C:5]=[c,nX2,C:6]<br>[c,nX2:7]=[c,C,nX2:8][c,C,nX2,NX2:9]=[c,C,nX2:10][c,nX2,C:11]=[c,C,nX2:12]  | PT_11_03 |

|  |          |
|--|----------|
| [c,C,nX2:13]=[c,C,nX2:14][c,C,nX2:15]=[c,C,nX2:16]<br>[c,C,nX2:17]=[nX2,NX2,O:18]>>[NX2,nX2,O:2]=[c,nX2,C:3]<br>[c,nX2,C:4]=[C,c,nX2:5][c,C,nX2:6]=[c,C,nX2,NX2:7]<br>[c,C,nX2:8]=[c,C,nX2,NX2:9][c,C,nX2:10]=[c,C,nX2:11]<br>[c,C,nX2:12]=[c,C,nX2:13][c,C,nX2:14]=[c,C,nX2:15][c,C,nX2:16]=[c,C,nX2:17]<br>[nX2,O:18][#1:1]  |          |
| [#1:1][n,N,O:2][c,nX2,C:3]=[c,nX2,C:4][c,nX2,C:5]=[c,nX2,C,NX2:6]<br>[c,nX2:7]=[c,C,nX2:8][c,C,nX2,NX2:9]=[c,C,nX2,NX2:10]<br>[c,nX2,C:11]=[c,C,nX2:12][c,C,nX2:13]=[c,C,nX2:14]<br>[c,C,nX2,NX2:15]=[c,C,nX2:16][c,C,nX2:17]=[c,C,nX2:18]<br>[c,C,nX2:19]=[nX2,NX2,O:20]>>[NX2,nX2,O:2]=[c,nX2,C:3]<br>[c,nX2,C:4]=[C,c,nX2:5][c,C,nX2,NX2:6]=[c,C,nX2,NX2:7]<br>[c,C,nX2:8]=[c,C,nX2,NX2:9][c,C,nX2,NX2:10]=[c,C,nX2:11]<br>[c,C,nX2:12]=[c,C,nX2:13][c,C,nX2:14]=[c,C,nX2,NX2:15]<br>[c,C,nX2:16]=[c,C,nX2:17][c,C,nX2:18]=[c,C,nX2:19][nX2,O:20][#1:1] | PT_11_04 |
| [#1:1][O,S,N:2][C,c;z;r5:3]=[C,c;r5:4][c,C;r5:5]>>[O,S,N:2]=[Cz2r5:3]<br>[C&r5R{0-2}:4](#1:1)[C,c;r5:5]  | PT_12_00 |
| [O,S,Se,Te;X1:1]=[C:2]=[C:3][#1:4]>>[#1:4][O,S,Se,Te;X2:1][C:2]#[C:3]  | PT_13_00 |
| [#1:1][C:2][N+:3]([O-:5))=[O:4]>>[C:2]=[N+:3]([O-:5))([O:4][#1:1]  | PT_14_00 |
| [#1:1][C:2][N:3](=[O:5))=[O:4]>>[C:2]=[N:3](=[O:5))([O:4][#1:1]  | PT_15_00 |
| [#1:1][O;!R:2][N+0z1:3]=[CX3:4]>>[O;!R:2]=[N+0z1:3][CX4:4][#1:1]   | PT_16_00 |
| [#1:1][O:2][Nz1:3]=[C:4][C:5]=[C:6][C:7]=[O:8]>>[O:2]=[Nz1:3][c:4]=[c:5]<br>[c:6]=[c:7][O:8][#1:1]   | PT_17_00 |
| [#1:1][O:2][C:3]#[N:4]>>[O:2]=[C:3]=[N:4][#1:1]  | PT_18_00 |
| [#1:1][O,N:2][C:3]=[S,Se,Te:4]=[O:5]>>[O,N:2]=[C:3][S,Se,Te;v{2-4}:4][O:5]<br>[#1:1]   | PT_19_00 |
| [#1:1][C0:2]#[N0:3]>>[C-:2]#[N+:3][#1:1]   | PT_20_00 |
| [#1:1][O,NX3:2][P;v3:3]>>[O,NX2:2]=[P;v5:3][#1:1]  | PT_21_00 |
| [#1:1][CX4:2][NX2:3]=[CX3:4]>>[CX3:2]=[NX2:3][CX4:4][#1:1]   | PT_22_00 |
| [#1:1][O,S,NX3:2][cX3;z;r5:3]=[c;r5:4][c;r5:5]=[c;z{1-2};r5;R{1-2}:6]>>[O,S,NX2:2]=[CX3;z;r5:3][C;r5:4]=[C;r5:5][Cz{1-2};r5;R{1-2}:6][#1:1]  | PT_23_00 |
| [#1:1][OX2:2][N;z{1-2};X3!\$(N=O);H0:3]<br>[CX3,c,n,NX2;r5:4]=[n,NX2,CX3;r5:5]>>[O&-:2][N+;z{1-2};X3;H0:3]=[c,CX3,n,NX2;r5:4][n,NX3,CX4;r5:5][#1:1]  | PT_24_00 |
| [#1:1][OX2:2][NX3r5:3][c,C;r5:4]=[c,C;r5:5][CX3,NX2r5:6]=[NX2:7]>>[O&-&H0:2][NX3z2&+;r5:3]=[c,C;r5:4][c,C;r5:5]=[CX3,NX2r5:6][NX3:7][#1:1]   | PT_25_00 |

|  |          |
|--|----------|
| [#1:1][O:2][NX3r6:3][C;r6:4]=[C;r6:5][C;z1;r6:6]=[O,NX2,S:7]>>[O&-&H0:2]<br>[n&+,N&+,X3;z1;r6:3]=[c,C;r6:4][c,C;r6:5]=[c,C;z1;r6:6][O,NX3,S:7][#1:1]   | PT_26_00 |
| [H:12][OX2,CX4:1][c:2]1=[cR{2-}a3:3]([c:4])[cR{2-}a3:6]([c:5])=[c:7]<br>[cR{2-}a3:8](=[c:9])[cR{2-}a3:11]1=[c:10]>>[H:12][C:7]1[cR{2-}:6]<br>([c:5])=[cR{2-}:3]([c:4])[C:2](=[O,CX3:1])[cR{2-}:11](=[c:10])[cR{2-}:8]1=[c:9]     | PT_27_00 |
| [H:12][O:11][c:10]1=[c;a3;r5:9]([c,s;r5:6])[c;a3;r5:8]([c,s;r5:7])=[c:5][c;a3:4]<br>([c,s:1])=[c;a3:3]1[c,s:2]>>[H:12][C:5]1[c;a2:4]([c,s:1])=[c;a2:3]([c,s:2])[C:10]<br>(=[O:11])[c;a2;r5:9]([c,s;r5:6])=[c;a2;r5:8]1[c,s;r5:7] | PT_27_01 |
| [#1:1][CX4:2][c;r6:3]=[c;r6:4][c;r6:5]=[c;r6:6][N+:7]<br>([O-:9])=[O:8]>>[CX3:2]=[C;r6:3][C;r6:4]=[C;r6:5][C;r6:6]=[N+:7]([O-:9])[O:8]<br>[#1:1]   | PT_28_00 |
| [#1:1][CX4:2][c;r6:3]=[c;r6:4][NX3+:5]([O-:7])=[O:6]>>[CX3:2]=[C;r6:3]<br>[C;r6:4]=[NX3+:5]([O-:7])[O:6][#1:1]   | PT_29_00 |
| [#1:1][CX4:2][c;r6:3]=[c;r6:4][CX3:5]([H:7])=[OX1:6]>>[CX3:2]=[C;r6:3]<br>[C;r6:4]=[CX3:5]([H:7])[OX2:6][#1:1]   | PT_29_01 |
| [#1:1][N:2][N+:3]([O-:5])=[O:4]>>[N:2]=[N+:3]([O-:5])[O:4][#1:1]   | PT_30_00 |
| [#1:1][CX4z1:2]1[CX3:3]=[CX3:4][CX3:5]=[CX3;!a:6][SX4:7]1(=[O])<br>(=[O])>>[CX3z1;!a:2]1=[CX3:3][CX3:4]=[CX3:5][CX4:6]([#1:1])[SX4:7]1(=[O])<br>(=[O])   | PT_31_00 |
| [#1:1][CX4;\$([C][CX{3-4}]=,-[OX{1-2}]):2][CX2:3]#[NX1:4]>>[CX3;\$([C]<br>[CX{3-4}]=,-[OX{1-2}]):2]=[C:3]=[NX2:4][#1:1]  | PT_32_00 |
| [#1:1][CX4:2][CX3:3]=[C;\$([CX3][CX{2-3}]=,#[N,O]):4]<br>[CX2:5]#[NX1:6]>>[CX3:2]=[CX3:3][C;\$([CX3][CX{2-3}]=,#[N,O]):4]=[C:5]=[NX2:6][#1:1]  | PT_33_00 |
| [#1:1][CX4:2][PX4:3]=[C;\$([CX{2-3}z2]~[PX{3-4}]):4]>>[CX3:2]=[PX4:3][C;\$<br>([CX{3-4}z2]~[PX{3-4}]):4][#1:1]   | PT_34_00 |
| [Sv2X2:1][OX2:2][#1:3]>>[Sv4X3:1]([#1:3])=[OX1:2]  | PT_35_00 |
| [CX3:1]=[NX2:2][OX2:3][#1:4]>>[CX3:1]=[NX3+:2]([OX1-:3])[#1:4]   | PT_36_00 |
| [NX2:1]=[CX3z{2-3}:2][SX2:3][OX2:4][#1:5]>>[#1:5][NX3:1][CX3z{2-3}:2]=[SX2+:3][OX-:4]  | PT_37_00 |
| [#1:1][CX4;!a:2][CX3;!a:3]=[NX3+:4][SiX4-:5]([NX3:7])<br>([NX3:8])=[O:6]>>[CX3;!a:2]=[CX3;!a:3][NX3:4][SiX4:5]([NX3:7])([NX3:8])<br>[OX2:6][#1:1]  | PT_38_00 |
| [CX3,NX2:1]=[NX3+:2]([O-:3])[CX4:4][#1:5]>>[#1:5][CX4,NX3:1][NX3+:2]<br>([O-:3])=[CX3:4]   | PT_39_00 |
| [#1:1][PX4:2]=[C;\$([CX3][PX4+]):3][CX3z1:4]=[O:5]>>[PX3:2][C;\$([CX3]   | PT_40_00 |

|   |          |
|---|----------|
| [PX4+]):3]=[CX3z1:4][OX2:5][#1:1]   |          |
| [#1:1][SX2,NX3,OX2;!R:2][CX3,c;r{5-6}:3]=[NX3+r{5-6}:4]   | PT_41_00 |
| [OX1-:5]>>[SX1,NX2,OX1;!R:2]=[CX3,c;r{5-6}:3][NX3r{5-6}:4][OX2:5][#1:1]                             |          |
| [#1:1][CX4:4]1[NX3,O,S,Se:5][CX3:6](=[O:7])[CX3:2]=[CX3;a0:3]1>>[#1:1]                              | PT_42_00 |
| [CX4:2]1[CX3;a0:3]=[CX3:4][NX3,O,S,Se:5][CX3:6]1=[O:7]  |          |
| [#1:1][CX4:2][c:5]1=[c:9]2[c:8]=[c:7][c:6]=[c:11][c:10]2=[c:4][#8:3]1>>[#1:1]                       | PT_43_00 |
| [CX4:4]1[#8:3][CX3:5](=[CX3!c:2])[c:9]2=[c:10]1[c:11]=[c:6][c:7]=[c:8]2                             |          |
| [#1:7][CX4;\$([C][C]#[N]),\$([C][C](=[O])[O]):6)[c:5]1=[cR1:4][c:3]=[c:2]                           | PT_44_00 |
| [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] |          |
| [#1:1][CX4:2]([CH3:3])([CH3:4])[CX3R1r{5-8}!c;z0:5]=[CX3R1r{5-8}!c:6]                               | PT_45_00 |
| [CR{1-};!c:7]>>[CX3:2]([CH3:3])([CH3:4])=[CX3R1r{5-8};z0:5][CX4R1r{5-8}:6]([CR{1-}:7])[#1:1]        |          |
| [#1:8][CX4;\$([C][S](=[O])[O]):7][C:1]1=[C:6][C:5]=[NX2+0:4]  | PT_46_00 |
| [C:3]=[C:2]1>>[#1:8][N:4]1[C:3]=[C:2][C:1](=[CX3;\$([C][S](=[O])[O]):7))                            |          |
| [C:6]=[C:5]1  |          |
| [#1:10][CX4:8]1[#7X2:9]=[CX3:7][c:6]2=[c:5]1[c:4]=[c:3][c:2]=[c:1]2>>[#1:10]                        | PT_47_00 |
| [#7X3:9]1[c:7]=[c:6]2[c:1]=[c:2][c:3]=[c:4][c:5]2=[c:8]1  |          |
| [#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]                              | PT_48_00 |
| [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  |          |
| [#1:9][OX2:8][NX3R1r5:1]([aR{1-}r{5-}:2])[cR1r5:5]=[cR1r5:4]  | PT_49_00 |
| ([aR{1-}r{5-}:3])[CX3:6]=[O:7]>>[OX1-:8][NX3+R1r5:1]  |          |
| ([a,AR{1-}r{5-}:2)=[CR1r5:5][CX3R1r5:4]([a,AR{1-}r{5-}:3))=[CX3:6][OX2:7]                           |          |
| [#1:9]  |          |
| [#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*:4]~[*:7]~[R1:6]   | RC_03_00 |
| [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}:6][O,N,S,Se,Te;!R:5]                               | RC_03_01 |
| [#1:1]>>[#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*;R1:4]~[*:7]~[R:6]                                       |          |
| [O,N,S,Se,Te;R1:5]1   |          |
| [O,N,S,Se,Te:2]=[C;!R:3][!R:4]~[!R:7][R{0-1}:6][N,nx3:5][#1:1]>>[#1:1]                              | RC_03_02 |
| [O,N,S,Se,Te:2][#6R1;!c:3]1[*;R1:4]~[*:7]~[R:6][N,nx3:5]1   |          |
| [#1:1][OX2:2][BX3:3]([OX2])[cr6:4][cr6:5][CX3:6]=[OX1:7]>>[OX2:2]1[BX3:3]                           | RC_03_03 |
| ([OX2])[cr{5-6}:4][cr{5-6}:5][CX4:6]1[OX2:7][#1:1]  |          |

|  |          |
|--|----------|
| [#1:1][OX2:2][CX4;!R:3][CX4:4][CX4:5][CX3;!c:6]=[CX3!c;\$([C][CX3]<br>(=[OX1])[OX2]):7]>>[OX2:2]1[CX4:3][CX4:4][CX4:5][CX4:6]1[CX4;\$([C]<br>[CX3](=[OX1])[OX2]):7)][#1:1]   | RC_03_04 |
| [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]<br>[#1:1]>>[#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*;R1:4]~[*:7]~[*:8]~[R1:6]<br>[O,N,S,Se,Te;R:5]1   | RC_04_01 |
| [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]<br>[#1:1]>>[#1:1][O,N,S,Se,Te:2][#6R1;!c:3]1[*;R1:4]~[*;R1:7]~[*:8]~[R:6]<br>[O,N,S,Se,Te;R1:5]1   | RC_04_02 |
| [#1:1][NX3!R:2][SX4:3](=[O:4])(=[O:5])[c:6]=[c:7][NX3:8][CX3!c:9]=[CX3!<br>c:10]>>[NX3:2]1[SX4:3](=[O:4])(=[O:5])[c:6]=[c:7][NX3:8][CX4:9]1[CX4:10]<br>[#1:1]  | RC_04_04 |
| [#1:1][N;R1;X3:3]1[!a:4]~[R:6][O,N,S,Se,Te;R:5][#6R;z2;X4:2]1>>[C;!<br>R;z1;X3:2]=[N;!R,X2;+0:3][*:4]~[*:6][O,N,S,Se,Te;!R:5][#1:1]  | RC_09_00 |
| [#1:1][N;R1;X3:3]1[!a:4]~[*:7]~[*;R1:6][O,N,S,Se,Te;R:5][#6R;z2;X4:2]1>>[C;!<br>R;z1;X3:2]=[N;!R;+0:3][R{0-1}:4]~[*;R{0-1}:7]~[!R:6][O,N,S,Se,Te:5][#1:1]<br>[C;!R;z1;X3:2]=[N;!R;+0:3][!R:4]~[*;R{0-1}:7]~[R{0-1}:6][O,N,S,Se,Te;!R:5]<br>[#1:1]>>[#1:1][N;R1;X3:3]1[!a;R1:4]~[*:7]~[*:6][O,N,S,Se,Te;R1:5]<br>[#6R;z2;X4:2]1 | RC_10_00 |
| [#1:1]>>[#1:1][N;R1;X3:3]1[!a;R1:4]~[*:7]~[*:6][O,N,S,Se,Te;R1:5]<br>[#6R;z2;X4:2]1  | RC_10_01 |
| [OX2;R:2]1[*R:3]~[*R:4][NX3:5]([#1:1])[PX5R;z2:6]1>>[#1:1][O;!R:2]<br>[*:3]~[*:4][NX2;!R;+0:5]=[PX4;!R;z1:6]   | RC_12_00 |
| [OX:2]=[CX2;z2:3]=[NX2;!R:4][c;R{0-1};!\$(*=[#7,#8,#16]):5]~[c;R{0-1}:6]!<br>[C;R{0-1}:7][NX3;!R:8][#1:1]>>[O:2]=[CX3;z3;R:3]1[NX3;R:4]([#1:1])[c;R{1-<br>2};!\$(*=[#7,#8,#16]):5]~[c;R{1-3}:6]![C;R{1-2}:7][NX3;R:8]1   | RC_13_00 |
| [#1:1][NX3:2][CX{2-3}:3][NX3:4][CX3;R1:5]1[SX2;R1:6][NX3;R1:7]<br>[CX3;R1:8](=[O:9])[NX2:10]=1>>[NX3;R:2]1[CX{2-3};R:3][NX3;R:4]<br>[CX3;R:5](=[NX2:10][CX3:8](=[O:9])[NX3:7][#1:1])[SX2;R1:6]1  | RC_14_00 |
| [#1:1][NX3,OX2:2][CX4!R:3][CX4:4][CX4:5][CX3!c:6]=[NX3+:7]<br>[OX1-:8]>>[NX3,OX2:2]1[CX4:3][CX4:4][CX4:5][CX4:6]1[NX3:7][OX2:8]<br>[#1:1]  | RC_15_00 |
| [#1:1][OX2:2][CX4:3][PX3:4][CX4:5][OX2:6][BX3:7]>>[OX2:2]1[CX4:3]<br>[PX3:4][CX4:5][OX2:6][BX4-:7]1.[#1+1]   | RC_16_00 |
| [OX2:2]1[CX4:3][PX4;\$ (P=[O,S,Se]):4][CX4:5][OX2:6][BX4-:7]1.[#1:1]<br>[NX4+:8]>>[#1:1][OX2:2][CX4:3][PX4;\$ (P=[O,S,Se]):4][CX4:5][OX2:6]<br>[BX4-:7][NX4+:8]  | RC_17_00 |



|   |          |
|---|----------|
| [#1:1][OX2:2][CX4:3][c:4]=[c:5][P:6]=[OX1:7]>>[OX2:2]1[CX4:3][c:4]=[c:5]<br>[P:6]1[OX2:7][#1:1]   | RC_18_00 |
| [#1:1][CX4:2]([NX3+:3]([O-:5)=[O:4])[CX4:6][CX4:7]<br>[CX3:8]=[CX3:9]>>[CX3:2](=[NX3+:3]([O-:5])[O:4]1)[CX4:6][CX4:7]<br>[CX4:8]1[CX4:9][#1:1]  | RC_19_00 |
| [#1:1][OX2,NX3:2][CX4!R,CD4!R2:3][CX4:4][NX3+:5]<br>([OX1-:7)=[CX3:6]>>[OX2,NX3:2]1[CX4:3][CX4:4][NX3:5]([OX2:7][#1:1])<br>[CX4:6]1   | RC_20_00 |
| [#1:1][CX4:2]([NX3+:3]([O-:5)=[O:4])[CX4:6][CX4:7]<br>[CX3:8]=[CX3:9]>>[CX4:2]1([NX3+:3]([O-:5)=[O:4])[CX4:6][CX4:7]<br>[CX4:8]1[CX4:9][#1:1]   | RC_21_00 |
| [#1:1][OX2:2][NX2:3]=[CX3:4][CX4:5][NX3+:6]([OX1-:7)=[CX3:8]>>[OX1-:2]<br>[NX3+:3]1=[CX3:4][CX4:5][NX3:6]([OX2:7][#1:1])[CX4:8]1  | RC_22_00 |
| [#1:1][OX2:2][CX4:3][CX4:4][CX4:5][NX3+:6]<br>([OX1-:7)=[CX3:8]>>[OX2:2]1[CX4:3][CX4:4][CX4:5][NX3:6]([OX2:7][#1:1])<br>[CX4:8]1  | RC_23_00 |
| [#1:12][O:10][C:9][cr6:6]=[cr6:1][PX3z0:7]>>[#1:12][PX5z1:7]1[O:10][C:9]<br>[cr{5-6}:6]=[cr{5-6}:1]1  | RC_24_00 |
| [OX1:1]=[CX3R1:2][CX3R1:3]=[SX{1-2}z{0-1};!R:4]>>[OX2:1]1[c:2]=[c:3]<br>[SX{2-3}z{1-2}:4]1  | VT_01_00 |
| [SX1:1]=[CX3:2][CX3:3]=[SX1]>>[SX2:1]1[C:2]=[C:3][SX2:4]1   | VT_01_01 |
| [NX2,nX2:1]=[CX3,cz{2-3}:2]<br>[NX2z1:3]=[NX2z2+:4]=[NX1-:5]>>[NX3:1]1[Cz{2-3}:2]=[NX2z1:3]<br>[NX2z2:4]=[NX2z2:5]1   | VT_02_00 |
| [SX1:1]=[Cz2X2:2]=[NX2:3][cr6:4]=[cr6:5]<br>[NX2:6]=[NX2:7]>>[SX1:1]=[Cz3X3:2]1[NX2:3]=[C:4][C:5]=[NX2:6][NX3:7]1   | VT_03_00 |
| [NX3:1]1[N:2]=[CR2:3]2[C:4]=[C:5]<br>[C:6]=[CR2:7]2[N:8]=[N:9]1>>[NX2:1]=[NX2:2][C:3]1=[C:4][C:5]=[C:6]<br>[C:7]1=[NX2+:8]=[NX1-:9]   | VT_04_00 |
| [nX3;\$([n][C]#[N]),\$([n][N+](=[O])[O-]),\$([n][SX4](=[O])=[O]):1]1[c:2]=[c:3]<br>[nX2:4]=[nX2:5]1>>[NX2;\$([N][C]#[N]),\$([N][N+](=[O])[O-]),\$([N][SX4]<br>(=[O])=[O]):1]=[C:2][C:3]=[NX2+:4]=[NX1-:5] | VT_05_00 |
| [CX4,OX2,NX3:1]1[CX4:2]2[CX3:3]=[CX3:4][CX3:5]=[CX3:6]<br>[CX4:7]21>>[CX4,OX2,NX3:1]1[CX3:2]=[CX3:3][CX3:4]=[CX3:5]<br>[CX3:6]=[CX3:7]1   | VT_06_00 |

|  |          |
|--|----------|
| [PX4:1]=[CX3:2][NX3:3][CX3:4]=[OX1:5]>>[PX5:1]1[CX3-:2]<br>[NX3+:3]=[CX3:4][OX2:5]1  | VT_07_00 |
| [NX2:1]=[NX2:2][cr6:3][cr6:4][NX2+:5]#[NX1:6]>>[NX3+:1]1=[NX2:2][cr6:3]<br>[cr6:4][NX2:5]=[NX2:6]1   | VT_08_00 |
| [NX3:10][Pv3:9]([NX3:11])[N:8]=[CX3:7][c:5]1=[N:4][c:3]=[c:2]<br>[c:1]=[c:6]1>>[NX3:10][Pv5:9]1([NX3:11))=[N:8]<br>[CX3:7]=[C:5]2[N:4]1[C:3]=[C:2][C:1]=[C:6]2 | VT_09_00 |
| [NX2z1:9]=[NX2:8][c:6]1=[c:1]([c:2]=[c:3][c:4]=[c:5]1)<br>[PX3:7]>>[PX4+:7]1[Nz2:9][N-:8][c:6]2=[c:1]1[c:2]=[c:3][c:4]=[c:5]2                                  | VT_10_00 |

<sup>a</sup>Refer to CACTVS manual for details of attributes used in SMIRKS.

**Table S4:** Representative examples of chemical types of tautomerism, with most-applicable transform rule(s) added .

| Type of tautomerism  | Typical example |
|--|-----------------|
| Azo $\rightleftharpoons$ Hydrazone<br><br>Rules: PT_00_06, PT_00_07,<br>PT_00_09 |                 |
| Ring $\rightleftharpoons$ Chain <sup>a</sup><br><br>Rules: RC_13_00              |                 |
| Enol $\rightleftharpoons$ Keto<br><br>Rules: PT_00_06, PT_00_07                  |                 |

|  |  |
|--|--|
| <p>Oxo-enamine <math>\rightleftharpoons</math> Oxo-imine</p> <p>Rules: PT_03_00, PT_04_00, PT_06_00</p>  |  |
| <p>Diketo <math>\rightleftharpoons</math> Keto-enol</p> <p>Rules: PT_06_00</p>                           |  |
| <p>Enol-imine <math>\rightleftharpoons</math> Oxo-enamine</p> <p>Rules: PT_02_00, PT_07_00</p>           |  |
| <p>Amine <math>\rightleftharpoons</math> Imine</p> <p>Rules: PT_05_00, PT_06_00</p>                      |  |
| <p>Keto-enethiol <math>\rightleftharpoons</math> Thioketo-enol</p> <p>Rules: PT_07_00</p>                |  |
| <p>Azide <math>\rightleftharpoons</math> Tetrazole<sup>b</sup></p> <p>Rules: VT_02_00</p>                |  |
| <p>Chain <math>\rightleftharpoons</math> Ring<sup>b</sup> (Valence)</p> <p>Rules: VT_03_00, VT_04_00</p> |  |
| <p>Enamine <math>\rightleftharpoons</math> Imine</p> <p>Rules: PT_03_00, PT_04_00, PT_06_00</p>          |  |

|   |  |
|---|--|
| Oxo-enamine $\rightleftharpoons$ Phenol-imine<br><br>Rules: PT_02_00, PT_07_00                    |  |
| Pyridol $\rightleftharpoons$ Pyridone<br><br>Rules: PT_02_00, PT_05_00,<br>PT_07_00               |  |
| NH $\rightleftharpoons$ NH<br><br>Rules: PT_06_00, PT_07_00,<br>PT_08_00                          |  |
| Phenol-quinone $\rightleftharpoons$ Phenol-quinone<br><br>Rules: PT_07_00                         |  |
| Enol-imine $\rightleftharpoons$ Oxo-imine<br><br>Rules: PT_06_00                                  |  |
| Benzoxazaphospholine $\rightleftharpoons$<br>Iminophosphorane <sup>a</sup><br><br>Rules: RC_12_00 |  |
| CH $\rightleftharpoons$ NH<br><br>Rules: PT_03_00, PT_06_00                                       |  |
| Keto-enol $\rightleftharpoons$ Keto-enol<br><br>Rules: PT_07_00                                   |  |

|   |  |
|---|--|
| NH-imidazole $\rightleftharpoons$ NH-imidazole<br>Rules: PT_05_00, PT_06_00 |  |
| Lactam $\rightleftharpoons$ Lactim<br>Rules: PT_05_00, PT_06_00             |  |
| Amine-imine $\rightleftharpoons$ Amine-imine<br>Rules: PT_05_00, PT_06_00   |  |
| Cyclohexadienone $\rightleftharpoons$ Phenol<br>Rules: PT_27_00             |  |
| 3H-2-one $\rightleftharpoons$ 5H-2-one<br>Rules: PT_42_00                   |  |
| Enethiol $\rightleftharpoons$ Thioketo<br>Rules: PT_06_00                   |  |
| Diazaphosphazole $\rightleftharpoons$ Phosphinoimine<br>Rules: VT_09_00     |  |
| Phosphine $\rightleftharpoons$ Phosphonium salt<br>Rules: VT_10_00          |  |
| Isoindole $\rightleftharpoons$ Isoindolenine<br>Rules: PT_47_00             |  |

|  |  |
|--|--|
| 1,4-Dihydro $\rightleftharpoons$ 1,6-Dihydro<br>Rules: PT_05_00, PT_06_00              |  |
| Nitrone $\rightleftharpoons$ Nitro<br>Rules: PT_39_00                                  |  |
| N-hydroxide $\rightleftharpoons$ N-oxide<br>Rules: PT_26_00                            |  |
|  |  |
|  |  |
| Isopropylidenecycloalkane $\rightleftharpoons$ Isopropylcycloalkene<br>Rules: PT_45_00 |  |
| Thioamide $\rightleftharpoons$ Thioimidol<br>Rules: PT_06_00                           |  |
| Keteneimine $\rightleftharpoons$ Nitrile<br>Rules: PT_32_00, PT_33_00                  |  |
| Tropolone $\rightleftharpoons$ Tropolone<br>Rules: Multi steps                         |  |
| 2H $\rightleftharpoons$ 6H<br>Rules: PT_07_00  |  |
| Amide $\rightleftharpoons$ Imidol<br>Rules: PT_06_00                                   |  |
| Amino $\rightleftharpoons$ Imino<br>Rules: PT_23_00                                    |  |

|   |  |
|---|--|
| Arene-imine $\rightleftharpoons$ Azepine <sup>b</sup><br>Rules: VT_06_00            |  |
| Anaquinoid $\rightleftharpoons$ Paraquinonimine<br>Rules: PT_02_00, PT_07_00        |  |
| NH $\rightleftharpoons$ OH<br>Rules: PT_07_00                                       |  |
| 1,2-Dihydro $\rightleftharpoons$ 1,4-Dihydro<br>Rules: PT_03_00, PT_05_00, PT_06_00 |  |
| Carbamoylimino $\rightleftharpoons$ Guanidine <sup>a</sup><br>Rules: RC_14_00       |  |
| Thiol $\rightleftharpoons$ Thione<br>Rules: PT_02_00, PT_07_00, PT_06_00/PT_09_00   |  |
| Nitroso-enamine $\rightleftharpoons$ Oxim-imine<br>Rules: PT_07_00                  |  |

|  |  |
|--|--|
| 1,2-Dihydro $\rightleftharpoons$ 2,5-Dihydro<br>Rules: PT_06_00                    |  |
| Cycloheptatriene $\rightleftharpoons$ Norcaradiene <sup>b</sup><br>Rules: VT_06_00 |  |
| Pyrrole $\rightleftharpoons$ Pyrrolidine<br>Rules: PT_44_00                        |  |
| 1,4-Dihydro $\rightleftharpoons$ 4,6-Dihydro<br>Rules: PT_07_00                    |  |
| Nitrone $\rightleftharpoons$ Oxime<br>Rules: PT_36_00                              |  |
| Triazole $\rightleftharpoons$ Triazole<br>Rules: PT_05_00, PT_06_00                |  |
| 2H $\rightleftharpoons$ 4H<br>Rules: PT_03_00/PT_06_00                             |  |
| Enol-enamine $\rightleftharpoons$ Oxo-enamine<br>Rules: PT_06_00                   |  |
| Amino-thieno $\rightleftharpoons$ Imine-thieno<br>Rules: PT_03_00, PT_06_00        |  |
| Isobenzofuran $\rightleftharpoons$ Phthalan<br>Rules: PT_43_00                     |  |
| CH $\rightleftharpoons$ OH<br>Rules: PT_06_00                                      |  |



|   |  |
|---|--|
| 1,4-Dihydro $\rightleftharpoons$ 4,5-Dihydro<br>Rules: PT_06_00                 |  |
| N(1)H $\rightleftharpoons$ N(3)H<br>Rules: PT_06_00                             |  |
| Amine $\rightleftharpoons$ Zwitterion<br>Rules: no_transform                    |  |
|   |  |
| Selenol $\rightleftharpoons$ Selone<br>Rules: PT_06_00                          |  |
| Imine $\rightleftharpoons$ Imine<br>Rules: PT_22_00                             |  |
| Nitro $\rightleftharpoons$ aci-Nitro<br>Rules: PT_29_00                         |  |
| 5,6-Dihydro $\rightleftharpoons$ 5,6-Dihydro<br>Rules: PT_03_00, PT_06_00       |  |
| 5,6-Dihydro-2H $\rightleftharpoons$ 5,6-Dihydro-4H<br>Rules: PT_03_00, PT_06_00 |  |
| C1-H $\rightleftharpoons$ C3-H<br>Rules: Muti steps                             |  |
| Thiol $\rightleftharpoons$ Zwitterion<br>Rules: no_transform                    |  |
| Sulfenyl $\rightleftharpoons$ Sulfinyl<br>Rules: PT_35_00                       |  |
| Sila-hemiaminal $\rightleftharpoons$ Silanoic-amide<br>Rules: PT_38_00          |  |

|  |  |
|--|--|
| $\lambda^5$ -Phosphane $\rightleftharpoons\lambda^3$ -Phosphane<br>Rules: RC_24_00 |  |
| $1H \rightleftharpoons 2H$<br>Rules: PT_06_00                                      |  |
| $2H \rightleftharpoons 2H$<br>Rules: PT_31_00                                      |  |
| $1,6\text{-Dihydro} \rightleftharpoons 3,6\text{-Dihydro}$<br>Rules: PT_06_00      |  |
| $4H \rightleftharpoons 6H$   |  |
| Nitroso-imine $\rightleftharpoons$ Oxim-imine<br>Rules: PT_07_00                   |  |
| $C3\text{-H} \rightleftharpoons N(5)H$<br>Rules: PT_07_00                          |  |
| Oxo-thione $\rightleftharpoons$ nul <sup>c</sup><br>Rules: Two steps               |  |
| Pyridol $\rightleftharpoons$ Zwitterion<br>Rules: no_transform                     |  |

|   |  |
|---|--|
| $1H \rightleftharpoons 3H$<br><br>Rule: PT_03_00,PT_06_00             |  |
| N-nitronic acid $\rightleftharpoons$ Nitramide<br><br>Rules: PT_30_00 |  |
| Enol $\rightleftharpoons$ Ylide<br><br>Rules: PT_40_00                |  |
| S-oxide $\rightleftharpoons$ Sulfenyl<br><br>Rules: PT_37_00          |  |

<sup>a</sup> Ring-chain tautomerism type (the total count for ring-chain tautomerism of two tautomers including benzoxazaphospholine $\rightleftharpoons$ iminophosphorane, carbamoylimino $\rightleftharpoons$ guanidine, and  $\lambda^5$ -phosphane $\rightleftharpoons$  $\lambda^3$ -phosphane pairs is 368).

<sup>b</sup> Valence tautomerism type.

<sup>c</sup> "nul" indicates cases of tautomeric equilibria for which no name for one or the other or both tautomer was given in the references and we were not able to assign any specific name.

Supporting\_information MCN V7\_2019-11-01.docx (1.07 MiB)

[view on ChemRxiv](#) • [download file](#)

---