J Comput Aided Mol Des (2009) 23:693–704
DOI 10.1007/s10822-009-9303-2

PERSPECTIVE

Let's not forget tautomers

Yvonne Connolly Martin

Received: 28 July 2009 / Accepted: 7 September 2009 / Published online: 20 October 2009
© Springer Science+Business Media B.V. 2009

Abstract A compound exhibits tautomerism if it can be represented by two structures that are related by an intramolecular movement of hydrogen from one atom to another. The different tautomers of a molecule usually have different molecular fingerprints, hydrophobicities and p*K*a's as well as different 3D shape and electrostatic properties; additionally, proteins frequently preferentially bind a tautomer that is present in low abundance in water. As a result, the proper treatment of molecules that can tautomerize, ~25% of a database, is a challenge for every aspect of computer-aided molecular design. Library design that focuses on molecular similarity or diversity might inadvertently include similar molecules that happen to be encoded as different tautomers. Physical property measurements might not establish the properties of individual tautomers with the result that algorithms based on these measurements may be less accurate for molecules that can tautomerize—this problem influences the accuracy of filtering for library design and also traditional QSAR. Any 2D or 3D QSAR analysis must involve the decision of if or how to adjust the observed K_i or *IC50* for the tautomerization equilibria. QSARs and recursive partitioning methods also involve the decision as to which tautomer(s) to use to calculate the molecular descriptors. Docking virtual screening must involve the decision as to which tautomers to include in the docking and how to account for tautomerization in the scoring. All of these decisions are more difficult because there is no extensive database of measured tautomeric ratios in both water and non-aqueous solvents and there is no consensus as to the best computational method to calculate tautomeric ratios in different environments.

Keywords Tautomer · Drug design · Hydrophobicity · Protein recognition

Y. C. Martin (✉)
Martin Consulting, 2230 Chestnut St., Waukegan, IL 60087, USA
e-mail: yvonnecmartin@comcast.net

Introduction

Molecules that can exist as different tautomers are chameleons. By virtue of a proton hopping from one polar atom to another and the rearrangement of double bonds or ring opening or closing, a particular atom changes from a hydrogen-bond donor to an acceptor while another atom in the molecule changes from a hydrogen-bond acceptor to a hydrogen-bond donor. Tautomeric reactions in which a heterocyclic ring is opened and closed also change the shape of the molecule.

Small changes in molecular structure or solvent environment can dramatically change the ratio of tautomers: Such changes complicate the assignment of a physical property measurement to a specific chemical structure, the identification of the bioactive species from a tautomeric mixture, and the probability that a "minor" species is the one recognized by a macromolecule.

Although there are many reasons for not carefully considering tautomers in computer assisted drug design, the time has come to take up the challenge. This perspective is not a comprehensive review, but rather a sampling of the experimental information available on tautomers, the implications of these observations, and possible approaches to a more reliable consideration of tautomers in drug design. Although others have also highlighted the issue of tautomers [1–6], the full impact of tautomerism has not

received comprehensive attention from the computer-aided drug design community.

Experimental observations of tautomers

Rate of tautomerization

In general, if the tautomerism involves moving a proton from one heteroatom to another, the reaction is fast, particularly in aqueous solutions [7]. In these cases, NMR studies see both tautomers [8] and experimental measurements of log *P*, log *D*, or p*K*a contain contributions from all tautomers unless the analytical detection method has been specifically designed to detect only one. On the other hand, tautomerization may be slow if it involves a ring-chain equilibrium or if it involves moving a proton from a heteroatom to carbon atom.

Examples of the relationship between structure, solvent, and the tautomer ratio

The ratio of tautomers of any compound is highly dependent on the structure of the solute well as the solvent [7, 9]. For example, crystallization conditions may induce different tautomers of the same molecule or the two forms might co-exist in a single crystal [10–13].

Figure 1 shows examples of tautomeric equilibria in water [9]. Note that the equilibrium between 4-hydroxypyridine and 4-pyridone is affected by the solvent, by intramolecular hydrogen bonding, and by the electronic effects of substituents. In water the thione form of 4-mercaptopyridine predominates, but the equilibrium switches to the thiol form for 2-mercaptothiophene. The absence of numbers for some of the equilibrium constants in Fig. 1 indicates that although it was possible to establish the predominant tautomer, it was not possible to quantitate the concentration of the minor form.

Figure 2 shows an example of the change in tautomer ratio as a function of solvent and of structure [14]. The replacement of one of a pair of enolizable hydrogens by a methyl group increases the proportion of the NH form in all solvents and increases the proportion of the OH form in both non-polar solvents. Note that tautomerization would also racemize the chiral carbon of Structure 2.04.

Ring-chain tautomerism is well established in carbohydrates, but it also occurs in other molecules such as warfarin, Fig. 3 [15]. An example of the substituent effect on this type of equilibrium is shown in Fig. 4 [16]. Substitution of an ortho hydrogen with a nitro group favors the open form, whereas substitution with an amino or hydroxy group favors the cyclic form. The equilibrium constant for ring closure follows a Hammett relationship.

Clearly if one were comparing the biological properties of the compounds in Figs. 1, 2, 3 and 4, it would be important to be alert to the possibility that tautomerism might complicate the structure-activity relationships.

Examples of ligand tautomer preferences of macromolecules

Often the resolution of a protein crystal structure cannot clearly establish the tautomer of the bound ligand. However, there are several documented cases where the bound

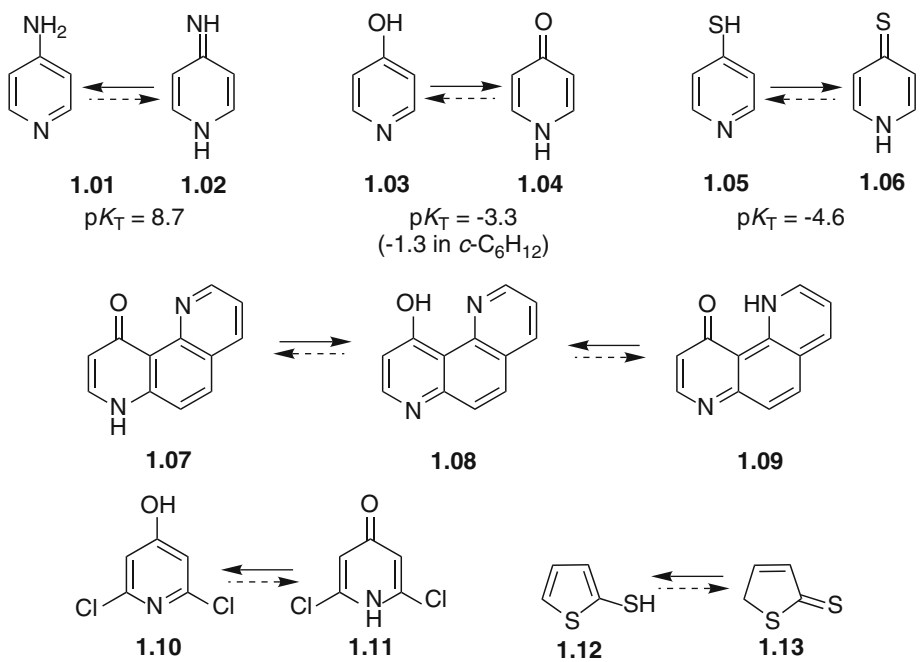
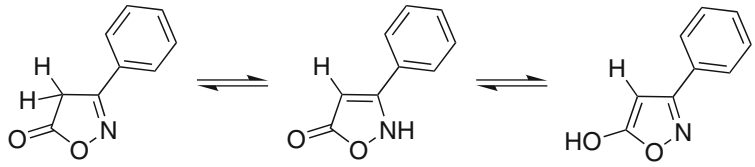
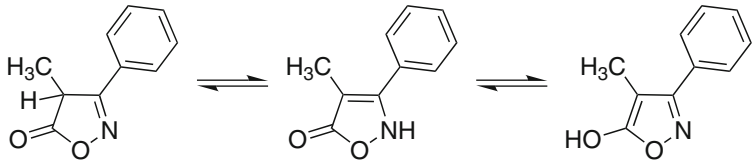


Fig. 1 Structure-activity relationships of tautomeric equilibria in water for some simple substituted heterocycles [9]



	2.01	2.02	2.03
Chloroform	100%	0%	0%
Dioxane	90-100%	0-3%	0-7%
Water	70%	30%	0%



	2.04	2.05	2.06
Chloroform	40%	60%	19-24%
Dioxane	37-39%	53-59%	4-9%
Water	0%	100%	0%

Fig. 2 The effect of changes in structure and solvent on tautomeric equilibria of simple heterocycles [14]

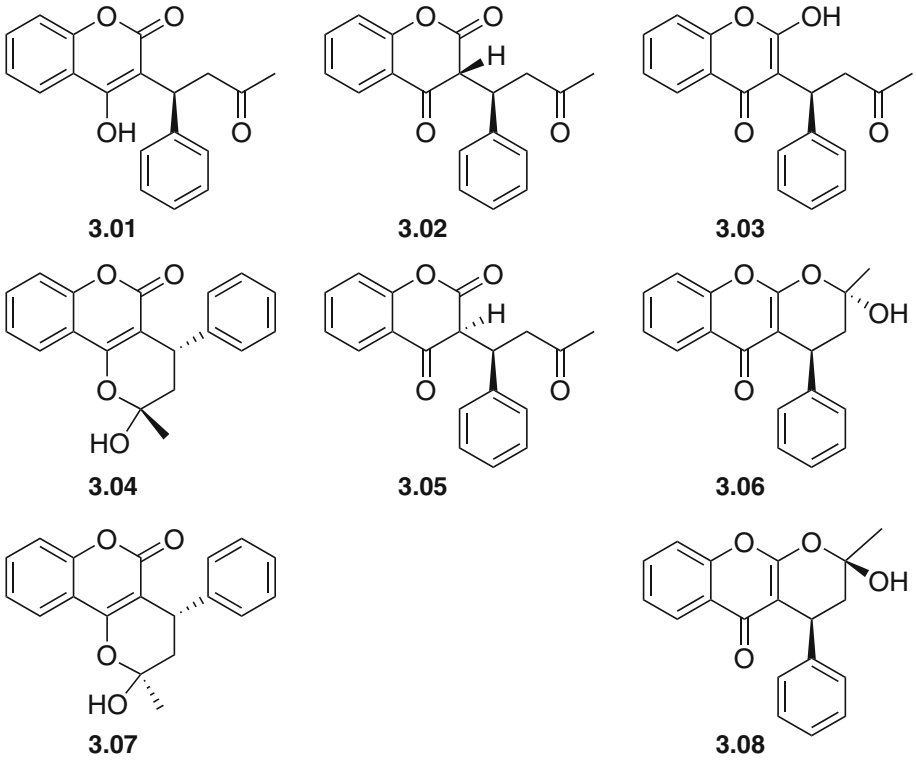
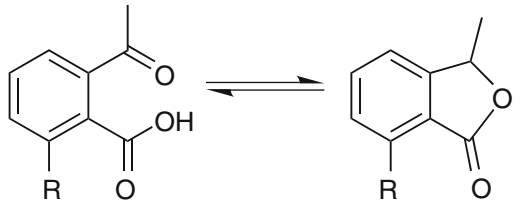


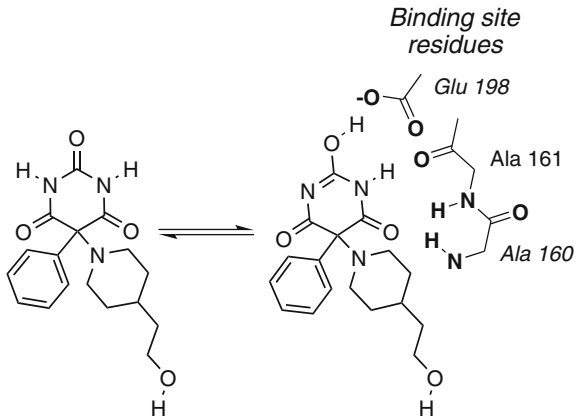
Fig. 3 The tautomers of warfarin [15]

tautomer has been unambiguously established. Figure 5 illustrates the contrast between the solution structure of a barbiturate analogue and that in a 1.8 Å crystal structure as bound to matrix metalloproteinase 8 [17]. Others have shown with SCRF-HF/6-31G** calculations that the tautomer of unsubstituted barbituric acid that corresponds to the bound tautomer is 20.05 kcal/mol less stable in polar medium [18]. Figure 6 shows the tautomer of pterin bound



R	id	fraction	id	fraction
H	4.01	equal	4.05	equal
NO_2	4.02	major	4.06	minor
NH_2	4.03	minor	4.07	major
OH	4.04	minor	4.08	major

Fig. 4 Structure-activity relationships for a chain-ring tautomerization [16]



5.01 Preferred form in water — **5.02** Form bound

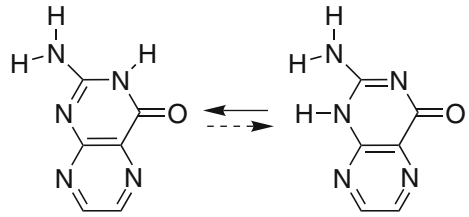
Calculated Octanol-water log *P*

CLOGP	1.35	2.96
KowWin	1.16	1.55

Fig. 5 The recognition of a minor tautomer by matrix metalloproteinase 8 [17]

to the 2.3 Å structure of ricin toxin A-chain. It is 3 kcal/mol higher in energy (AMSOL in AM1-SM2 Hamiltonians) in solution than the favored tautomer [19].

In some cases more than one tautomer is bound to the protein. For example, Fig. 7 shows the two tautomers that are bound with equal occupancy in a 1.53 Å structure of CDK [20]. This result contrasts with crystal structures of similar compounds in KDR [21] and PDGF [22], two other kinases, in which only the 2,4-dihydroindeno tautomer, the left structure, is observed. Macrophage migration inhibitory factor (MIF) catalyzes phenylpyruvate tautomerization, Fig. 8 [23]. It catalyzes the reaction in both directions, and hence binds both tautomers, although the enol-keto direction is preferred.

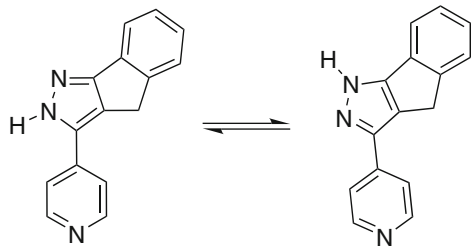


6.01 Tautomer in water — **6.02** Tautomer bound

Calculated Octanol-water log *P*

CLOGP	-1.60	-1.54
KowWin	-1.03	-1.89

Fig. 6 The recognition of a minor tautomer by the A-chain of ricin toxin [19]

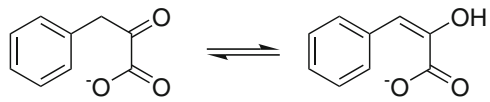


7.01 **7.02**

Calculated Octanol-water log *P*

CLOGP	3.18	3.18
KowWin	2.89	2.89

Fig. 7 Two tautomers recognized by CDK [20]



8.01 **8.02**

K_m	4.9 X 10^{-3} M	1.3 X 10^{-4} M
$k_{cat}(s^{-1})$	285	290
k_{cat}/K_m ($M^{-1}s^{-1}$)	5.8 X 10^4	2.2 X 10^6

Calculated Octanol-water log *P*

CLOGP	0.90	-0.74
KowWin	0.46	1.5

Fig. 8 Two tautomers recognized by macrophage inhibitory factor [23]

Enzymes can also select one species from a ring-chain equilibrium. For example, Fig. 9 shows the tautomers of chlorthalidone, a carbonic anhydrase inhibitor. The crystal structure of the carbonic anhydrase II-chlorthalidone complex shows that it is not bound as the amide form, but rather as an unusual lactim tautomer [24].

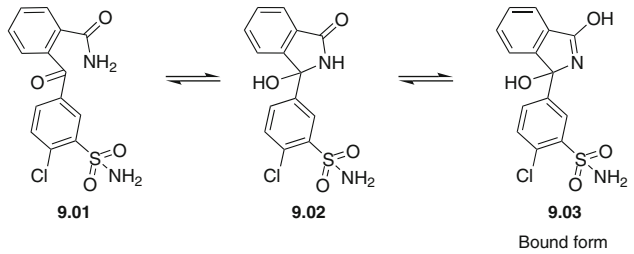


Fig. 9 The recognition of a minor tautomer by carbonic anhydrase II [24]

Proteins can bind different tautomers of related compounds. For example, glucose is a substrate for xylose isomerase and xylitol is an inhibitor, Fig. 10. Interestingly, the 0.95 Å crystal structures show that glucose is bound as a ring tautomer, not the chain form as expected from the structure of xylitol [25, 26].

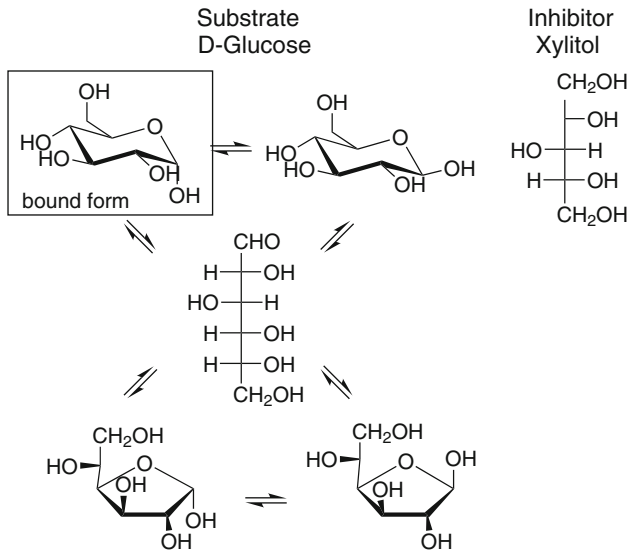


Fig. 10 The recognition of sugars by xylose isomerase [26]

A slightly more complex process is involved with the anti-tuberculosis drug isoniazid. It first forms an adduct with NAD(P); this adduct then inhibits a long-chain enoyl-acyl carrier protein reductase (InhA) [27]. Figure 11 summarizes the structures involved. Measurements on model compounds show that in contrast to the bound structure, in water the ring tautomer is favored by a factor of 2 [28, 29].

Complementary hydrogen bonds of bases in DNA lead to the formation of the characteristic double helix of DNA. When the base-pair mimics shown in Fig. 12 form a double helix with complementary DNA, the analogue that positions the tautomerizable group in the major groove is in the keto-amino tautomer [30]. However, the analogue that binds in the minor groove is in the syn-enol tautomer. The differences in tautomer preferences reflect the differences in the character of the major and minor grooves.

Frequency of molecules that can tautomerize

A summary of one program's enumeration of tautomers [31] of marketed drugs [32] is shown in Fig. 13. Of the 1,791 compounds, 1,334 or 74% exist as only one tautomer—put another way, 26% exist as an average of three tautomers. For this dataset and enumeration program 2,949

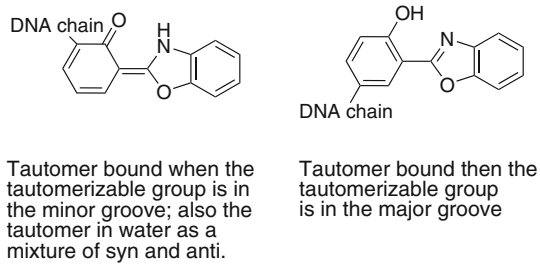
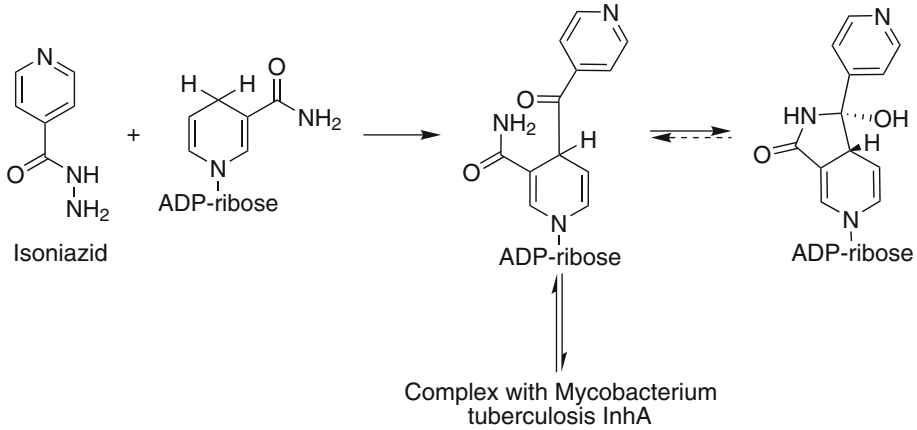


Fig. 12 Different tautomers bound to the minor and major grooves of DNA [30]

Fig. 11 The inhibition of InhA by an adduct of isoniazid [27]



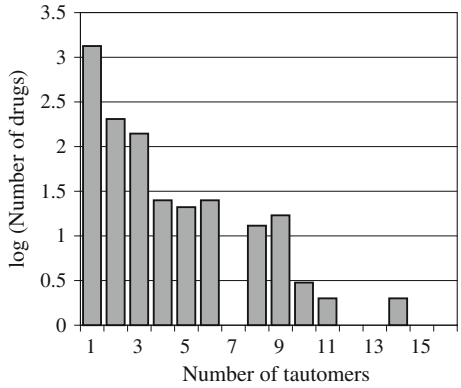


Fig. 13 The frequency distribution of tautomers of marketed drugs

tautomers are found; this increases the size of the dataset by 1.64-fold. Using a different tautomer generating program, others have found similar or slightly more increases in the size of a database [3]. Hence, although consideration of tautomers will increase the number of structures considered for virtual screening, the increase should be manageable.

Calculated properties of tautomers

p*K*a Differences between tautomers

Because the tautomers of a molecule have different structures, they differ in their ability to gain or lose a proton; their p*K*a values. In the simple case of an ionizable molecule that has two tautomeric forms, the tautomeric ratio is a function of the p*K*a's of the tautomers. For example, consider the tautomeric and ionic equilibria of 6-chloro-2-pyridone in water, Fig. 14. Algebraically $K_t = K_a^{OX}/K_a^{OH}$. Hence, one can calculate the value of any one of these equilibrium constants from values of the other two.

The observed p*K*a of a tautomerizable molecules is a composite of several individual microscopic ionization constants and the tautomeric equilibrium constant(s) [33]. For example, the protonated form tetracycline (Structure 10) can be present as any one of nine tautomers, and the neutral form by ten [34]. Each of these 19 species could contribute to the observed p*K*a as well as the biological properties and octanol-water log *D* of the molecule. Similarly, 8-oxoguanine (Structure 11) can exist as one or more of 100 neutral or anionic tautomers. This complicates investigations into its mechanism of mutagenicity [33].

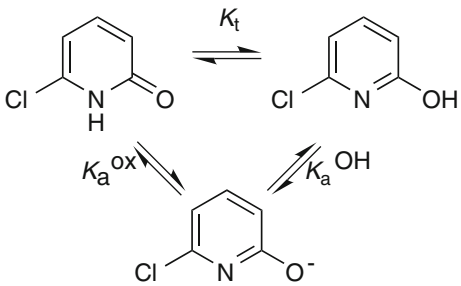
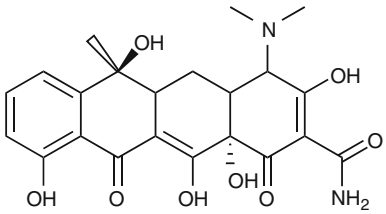
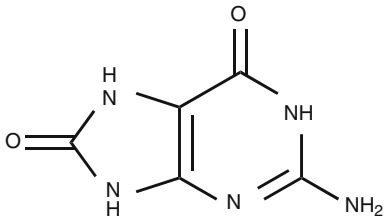


Fig. 14 The relationship between the equilibrium constant for tautomerization and the p*K*a's of the tautomers



10



11

Calculation of the tautomer ratio in solution

Although many workers have investigated the relative stabilities of tautomers in different liquid phases, because of the difficulty of measuring the equilibrium constants there is no publically available comprehensive database of this data. This lack hinders the development of empirical methods to predict the ratios of tautomers of a molecule. The implications of the lack of experimental data are described in detail in an article on predicting p*K*a [35], a less complex equilibrium constant.

If the tautomerization involves only the movement of a proton between sites, the tautomer equilibrium constant can be calculated from the p*K*a of each tautomer. This relationship holds because deprotonation of the tautomers lead to resonance structures of a common structure. Hammett-type [9] or empirical charge [36] relationships can be used to calculate the p*K*a's of the tautomers and hence the tautomeric ratio. However, even these calculations have errors in the range of 0.8 log units [35].

More elaborate, but not necessarily more accurate, calculations involve free-energy perturbation [37] or quantum chemical calculations [18, 19, 28, 33, 38–48]. To date there appears to be no consensus as to the most appropriate method.

Calculated octanol-water log *P* of tautomers

Usually the tautomers of a molecule have different hydrophobicities. Because small changes in structure or solvent can dramatically change the tautomeric ratio, ignoring the possibility of tautomerism leads to complications in assigning the specific molecular structure of a substance for which octanol-water log *P* has been measured. Indeed, usually the tautomer ratio in each phase has not been established. This ambiguity in turn results in inaccuracies of computational models to predict log *P*. For example we [49] and others [50] showed empirically that programs that calculate octanol-water log *P* are less accurate for molecules that can tautomerize.

Calculated log *P* values are often used to filter compounds for virtual screening, presumably because of its inverse correlation with water solubility [51–53] or permeability [52]. Such relationships have not been investigated to see if they also apply to molecules that can tautomerize.

In addition, calculated log *P* values might be used to predict brain to blood ratio using the simple equation that includes terms for log *P* and polar surface area, PSA [54]. Although PSA is quite similar for tautomers, the figures in this report show that tautomers of a molecule usually have different hydrophobicities. The question then becomes, which log *P* value should be used in the brain penetration calculation—should we assume that blood is like water and use the log *P* of the dominant form in water, or do we recognize that tautomerization is fast and use the log *P* of the more hydrophobic form to simulate brain tissue?

Figures 5, 6, 7 and 8 contain values of octanol-water log *P* calculated by two popular programs. Note that not only do the values calculated from the different programs seldom agree, but often they do not even agree as to which tautomer is more hydrophobic. As another example, Table 1 lists the calculated octanol-water log *P* of the tautomers of sildenafil (Viagra) and phenobarbital. Although the programs suggest little difference between Tautomers 1 and 3 of sildenafil, KowWin predicts that the enol form, Tautomer 2, is the least hydrophobic, whereas CLOGP and ALOGP suggest that it is the most hydrophobic of the three. As a consequence, CLOGP and ALOGP predict that Tautomer 2 is the predominant form in the water-saturated octanol phase, whereas KowWin predicts that it is the minor form in this phase. Similar contradictions are seen with the calculated log *P* of phenobarbital tautomers: CLOGP predicts that Tautomer 1, the tautomer most highly populated in water, is also the most hydrophobic tautomer, whereas ALOGP predicts that it is the least hydrophobic tautomer.

Cheminformatics issues with tautomers

Identifying if a molecule is in a database

This problem has been discussed by others [3, 55, 56]. Because the tautomers of a molecule do not have the same molecular structure, they will usually be encoded differently in the bitmaps or fingerprints that are used to discover if a particular molecule is in a database. An example of different tautomers registered in different databases is seen with sildenafil: Although Tautomer 3 (Table 1) has been reported to be more stable than Tautomer 1 and it is the one associated with a Chemical Abstracts [57] Number, Tautomer 1 is listed as the structure in PubChem [58] and ChemSpider [59].

The usual solution to this problem is to use a special algorithm to generate a unique tautomer, usually one assumed to predominate in water [3, 55]. Unfortunately, different software vendors use slightly different algorithms with the result that the same compound can be represented differently in different databases.

Substructure searching and identification

Substructure search queries that will identify tautomers need to be constructed with this possibility in mind. For example, if one uses Structure 1.03 as a search query, if the ring is specified to be aromatic, then molecules that contain Substructure 1.04, perhaps as the *N*-methyl derivative, would not be found.

Many cheminformatic investigations involve an analysis of the substructures present in the molecules under consideration. For example, QSARs or recursive partitioning may be based on the relative frequency of certain substructures in active versus inactive compounds: Clearly, such investigations are compromised if they do not include the substructures that are present in any (or most abundant?) tautomer of the molecule. The examples in Figs. 5, 6, 7, 8, 9, 10, 11 and 12 show that one cannot focus exclusively on the "major" tautomer.

Similarity searching

Table 2 shows Tanimoto similarities calculated with ECFP4 fingerprints [31] and the probability, based on the similarity, that the two compounds will have potency within 10-fold of each other [60]. The columns on the left list the similarities and probabilities between tautomers; the columns to the right list these values for the most similar molecule in this small dataset. Note that in most cases the most similar molecule is not a tautomer of the query molecule. Only if the query structure is rather complex is the tautomer similar. Note the low similarity between

Table 1 Calculations of octanol-water log *P* of different tautomers of viagra and phenobarbital

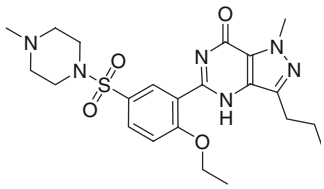
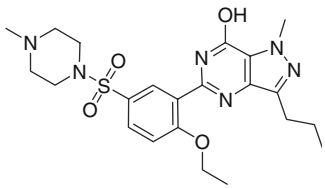
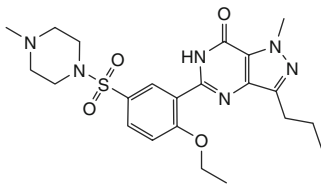
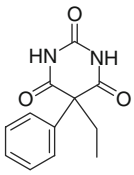
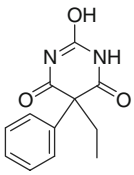
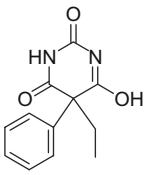
Compound	Tautomer	Structure	Octanol-water log P Program		
			CLOGP [74]	KowWin [75]	ALOGP [31]
Viagra	1		2.22	2.30	2.25
Viagra	2		3.56	1.60	3.06
Viagra	3		1.98	2.47	2.25
Phenobarbital	1		1.36	1.33	1.32
Phenobarbital	2		0.67	0.26	2.00
Phenobarbital	3		0.67	0.93	2.00

Table 1 continued

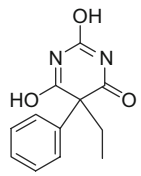
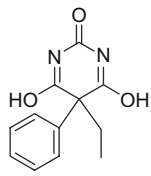
Compound	Tautomer	Structure	Octanol-water log P Program		
			CLOGP [74]	KowWin [75]	ALOGP [31]
Phenobarbital	4		0.67	1.27	2.68
Phenobarbital	5		0.67	1.27	2.68

Table 2 Tanimoto similarity comparisons of structures (ECFP_4 Fingerprints [31])

Structure	Tautomer			Most similar structure		
	Tautomer	Similarity	Probability of equal potency (%)	Structure	Similarity	Probability of equal potency (%)
101	102	0.07	0.14	103	0.43	18.66
103	104	0.07	0.14	105	0.43	18.66
105	106	0.07	0.14	103	0.43	18.66
107	108	0.06	0.12	105	0.19	0.89
107	109	0.06	0.12	105	0.19	0.89
108	109	0.06	0.12	106	0.23	1.62
110	111	0.04	0.09	104	0.22	1.40
112	113	0.15	0.49	111	0.33	6.57
201	202	0.30	4.42	204	0.33	6.57
203	201	0.26	2.52	206	0.33	6.57
202	203	0.22	1.40	205	0.35	8.42
204	205	0.27	2.91	201	0.33	6.57
204	206	0.25	2.18	205[a]	0.27	2.91
205	206	0.25	2.18	202	0.35	8.42
401	405	0.15	0.49	404	0.43	18.66
402	406	0.21	1.20	404	0.49	26.65
403	407	0.23	1.62	404	0.55	25.46
404	408	0.23	1.62	403	0.55	32.07
501	502	0.42	17.23	502[a]	0.42	17.23
601	602	0.34	7.45	602[a]	0.34	7.45
701	702	0.56	32.70	702[a]	0.56	32.70
801	802	0.38	11.81	802[a]	0.38	11.81
901	902	0.26	2.52	902[a]	0.26	2.52
901	903	0.27	2.91	903[a]	0.27	2.91
902	803	0.47	24.20	903[a]	0.47	24.20

[a] The most similar structure is a tautomer

Structures 5.01 and 5.02. This result shows that even simple similarity searching can be misleading if one ignores tautomerization.

Because similarity calculations form the basis for clustering and diversity selection, incorrect handling of tautomers can result in erratic results.

Tautomer enumeration programs

Cheminformatics software vendors recognize the problems that tautomers cause. As a result, most supply a tautomer enumeration program, generally only heterocyclic tautomers. To date, there has been no comparison of the different programs, probably because there is no recognized database. The users interested in using a database for virtual screening must then decide if they will enumerate all possible tautomers or just a few that are likely to be the most abundant in water.

Implications of tautomerization for QSAR

Figures 1, 2 and 3 remind us that within a series the ratio of tautomers in either the water or a non-aqueous phase is not constant. Because QSARs correlate the total concentration of a molecule with some biological effect, tautomerization has the effect of adding equilibria in addition to those for drug-target and drug-distribution. For example, correcting the observed concentration to that of "bioactive" tautomer in the aqueous phase does not account for the differential partitioning of tautomers of the various analogues to inert nonaqueous and receptor phases or that the target biomolecule may recognize a minor tautomer.

As noted above, for substructure-based QSARs, the first issue is to decide which tautomers should be included in the analysis. The second issue is how the algorithm allows the model to ignore some of the tautomers of a molecule.

Tautomerization complicates the calculation of molecular descriptors for traditional 2D QSAR [61]. For example, it may be ambiguous which calculated log *P* values to use as a molecular descriptor. Hence, the reliability of QSAR analyses that use hydrophobicity as a descriptor may suffer. In addition, because tautomers of a molecule have different p*K*a's, assigning a physical property to a specific molecular structure is especially challenging if the molecule can also ionize at pHs of interest [62]. On the other hand for 3D-QSARs, one must decide which tautomer as well as which conformer to use for the analysis.

Implications of tautomerization for docking molecules

High throughput docking programs are generally imprecise enough that one can attempt to dock all reasonable tautomers of a molecule. If the objective of the study is to identify compounds for experimental testing, if any tautomer of a molecule has a high score, validation is provided by experimental testing.

On the other hand, if the objective of the docking is to propose the structure of the protein-ligand complex, the preliminary docked structures would then be refined to optimize the fit and provide a prediction of affinity. This optimization would involve exploring the conformation of the ligand and the protein active site as well as the protonation and tautomeric state of both.

One strategy is to optimize and calculate the energy of every possible tautomeric and protonation state of the system, both in water and in the active site. This can be done with molecular mechanics force-fields [6, 19, 63, 64], with quantum mechanics [25, 29, 64–66], or a combination of the two [67]. At the current time, no method is particularly accurate—errors of 0.7–1.0 log units for each of the components are not uncommon [35, 67–69]. A quantum mechanical or QM/MM structure optimization would reveal the bound tautomer of both the ligand and the protein [67, 70]. For such calculations one would have to decide the level of theory necessary and whether the whole complex will be treated quantum mechanically or, if not, how the boundary between the quantum and molecular mechanics will be handled. Because a thermodynamic cycle is involved, the use of any method requires that it can reliably predict the ratio of tautomers in aqueous systems.

Directions for the future

The need for more experimental data

This review emphasizes the need for more experimental information on the tautomeric ratio of diverse molecules in water and various solvents. Such observations would form the basis for methods to predict the tautomeric ratio and a test bed to compare the accuracy of the various empirical and quantum chemical methods. Unfortunately, these measurements are difficult to design and often require synthesis of model compounds in hopes that they accurately mimic the properties of the corresponding tautomer.

Careful measurement of the impact of tautomerization on p*K*a and water solubility would provide information that would improve the predictions of these properties.

The need for cheminformatic databases that can maintain information about tautomers

Once a body of information is available, it might be discovered that enhancements must be made to the current architecture for storing chemical structures and information

[56]. For example, consider the problem of a database that would store all of the tautomers of Structures 10 or 11. Such a database would need to store not only the canonical tautomer but also structures and available properties of each individual tautomer, the measured or calculated equilibrium constants between the tautomers, and the properties of the compound itself.

The need for computer programs that predict ring-chain tautomerization

Rules for ring formation in organic synthesis have been formulated by Baldwin [71]. These would provide a starting point for a program that would enumerate ring-chain tautomers, a capability absent from the current tautomer generation programs.

The need for validation of the various computational methods

Although the various methods to explore the structure and energetics of enzyme-ligand complexes are interesting, for such methods to be useful they must be validated. For example is the continuum solvent assumption sufficient, or is it important to include explicit water molecules in the calculation? Before QM/MM calculations can be used in routine investigations of protein-ligand complexes, they will need to run faster and with less human interaction. A point to be examined would be whether a semi-empirical method [72, 73] might be sufficient for the quantum mechanical portion of the calculation and, indeed, whether the whole system can be accurately calculated with semi-empirical methods.

Summary

Tautomerization equilibria present a continuing challenge to computer-aided molecular design, affecting everything from library design to SAR to docking and scoring protein-ligand interactions. The absence of experimental data and validated computational methods make tautomerization easy to ignore but overwhelming to consider.

References

1. Pospisil P, Ballmer P, Scapozza L, Folkers G (2003) J Recept Signal Transduct 23:361–371
2. Masek BB, Clark RD, Yu Y, Smith K, Pearlman RS (2005) Hi fidelity chemistry. Tripos, Inc. http://www.bagim.org/presentation_files/06-03-15_Brian-Masek/HiFi-Chemistry.pdf. Accessed 3 Jan 2009
3. Oellien F, Cramer J, Beyer C, Ihlenfeldt W-D, Selzer PM (2006) J Chem Inf Model 46:2342–2354
4. Subramaniam S, Mchrotra M, Gupta D (2008) Bioinformation 3:14–17
5. Milletti F, Stochi L, Sforna G, Cross S, Cruciani G (2009) J Chem Inf Model 49:68–75
6. Todorov NP, Monthoux PH, Alberts IL (2009) J Chem Inf Model 46:1134–1142
7. Elguero J, Marzin C, Katritzky A, Linda P (1976) Advances in heterocyclic chemistry. Supplement 1: the tautomerism of heterocycles. Academic Press, New York
8. Shcherbakova I, Elguero J, Katritzky AR (2000) Adv Heterocycl Chem 77:51–113
9. Katritzky A (1994) Lecture 6. Heterocyclic tautomerism. http://ufark12.chem.ufl.edu/Lectures/L6.pdf. Accessed 15 April 2009
10. McConnell JF, Sharma BD, Marsh RE (1964) Nature 203:399–400
11. Steiner T, Koellner G (1997) Chem Commun 1997:1207–1208
12. Kubicki M (2004) Acta Crystallogr B 60:191–196
13. Wu Z-H, Ma J-P, Wu X-W, Huang R-Q, Dong Y-B (2009) Acta Crystallogr Sect C Cryst Struct Commun 65:o128–o130
14. Elgureo J, Marzin C, Katritzky AR, Linda P (1976) Advances in heterocyclic chemistry supplement 1: the tautomerism of heterocycles. Academic Press, New York, pp 302–303
15. Porter WR Tautomers of warfarin. Personal communication. August 2009
16. Finkelstein J, Williams T, Toome V, Traiman S (1976) J Org Chem 32:3229–3230
17. Brandstetter H, Grams F, Glitz D, Lang A, Huber R, Bode W, Krell H-W, Engh RA (2001) J Biol Chem 276:17405–17412
18. Senthilkumar K, Kolandaivel P (2002) J Comput-Aided Mol Des 16:263–272
19. Yan XJ, Day P, Hollis T, Monzingo AF, Schelp E, Robertus JD, Milne GWA, Wang SM (1998) Proteins Struct Funct Genet 31:33–41
20. Furet P, Meyer T, Strauss A, Raccuglia S, Rondeau J-M (2002) Bioorg Med Chem Lett 12:221–224
21. Dinges J, Ashworth KL, Akritopolou-Zanze I, Arnold LD, Baumeister SA, Bousquet PF, Cunha GA, Davidsen SK, Djuric SW, Gracias VJ, Michaelides MR, Rafferty P, Sowin TJ, Stewart KD, Xia Z, Zhang HQ (2006) Bioorg Med Chem Lett 16:4266–4271
22. Ho CY, Ludovici DW, Maharoof USM, Mei J, Sechler JL, Tuman RW, Strobel ED, Andraka L, Yen H-K, Leo G, Li J, Almond H, Lu H, DeVine A, Tominovich RM, Baker J, Emanuel S, Gruninger RH, Middleton SA, Johnson DL, Galemmo RA Jr (2005) J Med Chem 48:8163–8173
23. Taylor AB, Johnson WH, Czerwinski RM, Li H-S, Hackert ML, Whitman CP (1999) Biochemistry 38:7444–7452
24. Temperini C, Cecchi A, Scozzafava A, Supuran CT (2009) Bioorg Med Chem 17:1214–1221
25. Garcia-Viloca M, AlHambra C, Truhlar DG, Jiali G (2003) J Comput Chem 24:177–190
26. Fenn TD, Ringe D, Petsko GA (2004) Biochemistry 43:6464–6474
27. Rozwarski DA, Grant GA, Barton DHR, Jacobs WRJ Jr, Sacchettini JC (1998) Science 279:98–102
28. Delaine T, Bernardes-Genisson V, Stigliani J-L, Gornitzka H, Meunier B, Bernadou J (2007) Eur J Org Chem 2007:1624–1630
29. Stigliani J-L, Arnaud P, Delaine T, Bernardes-Génisson V, Meunier B, Bernadou J (2008) J Mol Graphics Modell 27:536–545
30. Dupradeau FY, Case DA, Yu C, Jimenez R, Romesberg FE (2005) J Am Chem Soc 127:15612–15617
31. Pilot P (2008) Scitegic chemistry components Accelrys. http://accelrys.com/products/datasheets/chemistry-component-collection.pdf. Accessed 4 May 2008

32. Proudfoot JR (2005) Bioorg Med Chem Lett 15:1087–1090
33. Jang YH, Goddard WA III, Noyes KT, Sowers LC, Hwang S, Chung DS (2002) Chem Res Toxicol 15:1023–1035
34. Duarte HA, Carvalho S, Paniago EB, Simas AM (1999) J Pharm Sci 88:111–120
35. Lee AC, Crippen GM (2009) J Chem Inf Model
36. Szegezdi J, Csizmadia F (2007) Tautomer generation. pKa based dominance conditions for generating dominant tautomers. American Chemical Society Fall meeting, Aug 19-23, 2007. http://www.chemaxon.com/conf/Tautomer_generation_A4.pdf. Accessed 30 April 2009
37. Worth GA, Richards WG (1994) J Am Chem Soc 116:239–250
38. Rashin AA, Rabinowitz JR, Banfelder JR (1990) J Am Chem Soc 112:4133–4137
39. Kleinpeter E, Thomas S, Fischer G (1995) J Mol Struct 355:273–285
40. Cramer CJ, Truhlar DG (1996) In: Mezey PG, Tapia O, Bertrán J (eds) Solvent effects and chemical reactivity. Springer, Dortrecht, pp 1–80
41. Karelson M, Maran U, Katritzky AR (1996) Tetrahedron 52: 11325–11328
42. Koskinen JT, Koskinen M, Mutikainen I, Mannfors B, Elo H (1996) Z Naturforsch A Phys Sci 51:1771–1778
43. Maran U, Karelson M, Katritzky AR (1996) Int J Quantum Chem 60:41–49
44. Maran U, Karelson M, Katritzky AR (1996) Int J Quantum Chem 60:1765–1773
45. Danilov VI, Stewart JJP, Les A, Alderferd JL (2000) Chem Phys Lett 328:75–82
46. Rogstad KN, Jang YH, Sowers LC, Goddard WA III (2003) Chem Res Toxicol 16:1455–1462
47. Godsi O, Turner B, Suwinska K, Peskin U, Eichen Y (2004) J Am Chem Soc 126:13519–13525
48. Podolyan Y, Gorb L, Leszczynski J (2005) J Phys Chem A 109:10445–10450
49. Martin YC (2007) We can't predict log P, so why should we expect to predict affinity? Open eye. http://eyesopen.com/about/events/cup8/martin_yvonne/Y.Martin-CUPVIII.pdf. Accessed 27 July 2009
50. Machatha S, Yalkowsky S (2005) Comparison of the octanol-water partition coefficients calculated by CLOGP, ACDLOGP and KowWin to experimentally determined values. http://www.aapsj.org/abstracts/AM_2005/AAPS2005-000448.pdf. Accessed 15 April 2009
51. Meylan WM, Howard PH, Boethling RS (1996) Environ Toxicol Chem 15:100–106
52. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Adv Drug Deliv Rev 23:3–25
53. Sanghvi T, Jain N, Yang G, Yalkowsky SH (2003) QSAR Comb Sci 22:258–262
54. Clark DE (1999) J Pharm Sci 88:815–821
55. Sayle R, Delaney J (1999) Canonicalization and enumeration of tautomers. http://www.daylight.com/meetings/emug99/Delany/taut_html/sld001.htm. Accessed 31 August 2009
56. Kenny PW, Sadowski J (2005) In: Oprea T (ed) Chemoinformatics in drug discovery. Wiley-VCH Verlag GmbH & Co, Weinheim, pp 271–285
57. Dittmar PG, Stobaugh RE, Watson CE (1976) J Chem Inf Comput Sci 16:111–121
58. (2009) Pubchem. National Institutes of Health. http://pubchem.ncbi.nlm.nih.gov/. Accessed 25 July 2009
59. (2009) Chemspider. http://www.chemspider.com/. Accessed 25 July 2009
60. Muchmore SW, Debe DA, Metz JT, Brown SP, Martin YC, Hajduk PJ (2008) J Chem Inf Model 48:941–948
61. Hansch C, Leo A (1995) Exploring QSAR: fundamentals and applications in chemistry and biology. American Chemical Society, Washington DC
62. Martin YC, Hackbarth JJ (1976) J Med Chem 19:1033–1039
63. Rastelli G, Thomas B, Kollman PA, Santi DV (1995) J Am Chem Soc 117:7213–7227
64. Perakyla M, Kollman PA (1997) J Am Chem Soc 119:1189–1196
65. Alex A, Finn P (1997) J Mol Struct 398:551–554
66. Hart JC, Burton NA, Hillier IH, Harrison MJ, Jewsbury P (1997) Chem Commun 1431–1432
67. Wang W, Donini O, Reyes CM, Kollman PA (2001) Annu Rev Biophys Biomol Struct 30:211–243
68. Khandelwal A, Lukacova V, Kroll DM, Comez D, Raha S, Balaz S (2004) QSAR Comb Sci 23:754–766
69. Riccardi D, Schaefer P, Cui Q (2005) The Journal of Physical Chemistry B 109:17715–17733
70. Gilson MK, Zhou H-X (2007) Annu Rev Biophys Biomol Struct 36:21–42
71. Baldwin JE (1976) J Chem Soc Chem Commun (18):734–736
72. Stewart J (1997) J Mol Struct 401:195–205
73. Stewart JJP (2001) Mopac 2002 manual. Fujitsu Ltd. http://www.cache.fujitsu.com/mopac/Mopac2002manual/index.html. Accessed 27 July 2009
74. Leo A, Hoekman D (2000) Perspect Drug Discovery Des 19–38
75. Meylan WM, Howard PH (2000) Perspect Drug Discovery Des 19:67–84