JOURNAL OF CHEMICAL INFORMATION AND MODELING

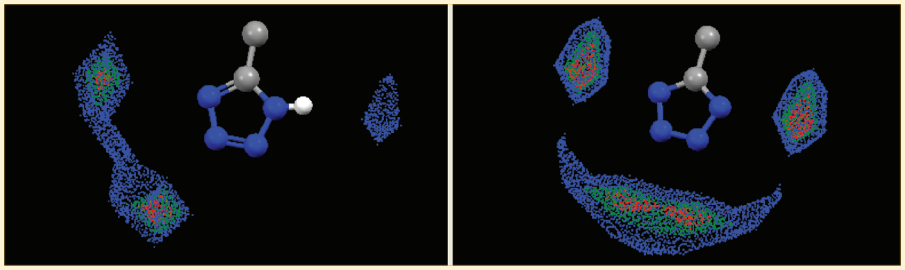
Article

pubs.acs.org/jcim

The Hydrogen Bond Environments of 1*H*-Tetrazole and Tetrazolate Rings: The Structural Basis for Tetrazole–Carboxylic Acid Bioisosterism

Frank H. Allen,* Colin R. Groom, John W. Liebeschuetz, David A. Bardwell, Tjelvar S.G. Olsson, and Peter A. Wood

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom



ABSTRACT: Bioisosterism involving replacement of a carboxylic acid substituent by 1*H*-tetrazole, yielding deprotonated carboxylate and tetrazolate under physiological conditions, is a well-known synthetic strategy in medicinal chemistry. To improve our overall understanding of bioisosterism, we have used this example to study the geometrical and energetic aspects of the functional group replacement. Specifically, we use crystal structure informatics and high-level ab initio calculations to study the hydrogen bond (H-bond) energy landscapes of the protonated and deprotonated bioisosteric pairs. Each pair exhibits very similar H-bond environments in crystal structures retrieved from the CSD, and the attractive energies of these H-bonds are also very similar. However, by comparison with −COOH and $-COO^-$, the H-bond environments around 1*H*-tetrazole and tetrazolate substituents extend further, by about 1.2 Å, from the core of the connected molecule. Analysis of pairs of PDB structures containing ligands which differ only in having a tetrazole or a carboxyl substituent and which are bound to the same protein indicates that the protein binding site must flex sufficiently to form strong H-bonds to either substituent. A survey of DrugBank shows a rather small number of tetrazole-containing drugs in the 'approved' and 'experimental' drug sections of that database.

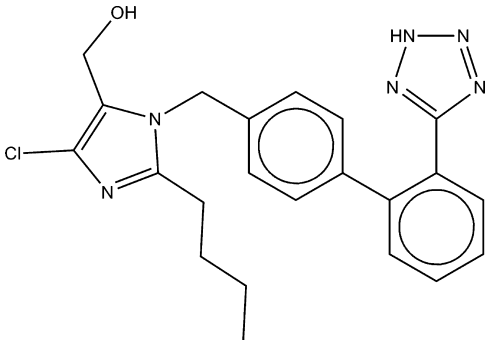
INTRODUCTION

A strategy commonly used by medicinal chemists to improve the efficacy of an existing drug, or in developing a new drug candidate, is to replace one or more substituent groups on the main molecular framework with groups that are isosteric with the original. The hope is that these group replacements will improve the physicochemical and biological properties of the original molecule.[1,2] Thornber[3] presents a history of isosterism in drug design and stresses that isosteric groups should maintain the shape characteristics of the molecule and conserve its key intermolecular interactions with the protein active site. This latter point was reinforced by Taylor et al.[4] in a study of carbonyl–nitro isosterism using nonbonded contact information from crystal structures. Over time, databases and software systems for selecting isosteric groupings have also been developed.[5−8] Very recently, Meanwell[9] has presented a major review of recent tactical applications of bioisosterism in drug design and highlights the difference between isosterism and bioisosterism, stressing that effective bioisosteres are more alike in their biological properties rather than their physical properties and bioisosteres that prove effective in one biological setting may not necessarily translate directly to another.

A frequently mentioned synthetic modification in drug design applications is to replace a carboxylic acid group with 1*H*-tetrazole, an acidic group with a pK_a of ≈5 for NH proton loss, a value which is similar to the pK_a for −COOH. An early example was the synthesis of a tetrazole derivative of the nonsteroidal anti-inflammatory indomethacin[10] that has similar activity to the parent carboxylic acid compound. Recent bioisosteric applications of tetrazole have been summarized by Myznikov et al.,[11] while Herr[12] reviews the specific medicinal chemistry and synthetic aspects of −COOH/1*H*-tetrazole bioisosterism, including physicochemical data and some information from absorption, distribution, metabolism, and toxicity (ADMET) studies. The review also provides case histories of several successful applications, including the antihypertensive losartan (I).[13] Here, compounds having −COOH substituents at the C2- or C3-position on the terminal phenyl ring were shown to be active when injected but had minimal activity when administered orally. Additional carboxylic acid derivatives were synthesized in an attempt to improve potency and bioavailability,

Received: October 31, 2011
Published: February 3, 2012

but the breakthrough came when it was found that a 1*H*-tetrazole substituent at C2 conferred a 10-fold increase in oral bioavailability compared with the −COOH compound.

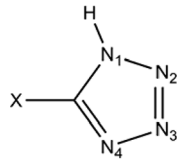


Biot et al.[14] provide a further detailed example encountered during the design of glutathione reductase inhibitors as antimalarials and address the structural basis of the isosterism through computational modeling of both carboxyl and 1*H*-tetrazole derivatives. These studies included full geometry optimization and the calculation of atomic point charges and eletrostatic potentials. Similar calculations for parent 1*H*-tetrazole have also been published.[15−17] However, none of these studies have made use of the conformational and intermolecular interaction data that are available from crystal structures stored in the Cambridge Structural Database (CSD)[18] or the Protein Data Bank (PDB).[19]

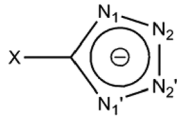
To improve our overall understanding of bioisosterism, we have used the carboxylate/1*H*-tetrazole example to study the geometrical and energetic aspects of the functional group replacement. Specifically, we use small-molecule crystal structure information from the CSD to compare the conformational preferences and intermolecular interactions of −COOH and the deprotonated −COO^- with those of 1*H*-tetrazole and the deprotonated tetrazolate substituent, respectively. It is important to include the deprotonated pairing in the analysis, since these species are most likely to be encountered under physiological conditions. Specifically, we have compared the conformations adopted by −COOH and −COO^- substituents with respect to the biphenyl framework in losartan (I) and related antihypertensives with those adopted by 1*H*-tetrazole and tetrazolate substituents, respectively, and carried out pairwise comparisons of the hydrogen-bond (H-bond) environments of −COOH and 1*H*-tetrazole with carboxylate and tetrazolate. The CSD H-bond analysis is augmented by H-bond energies for systems involving 1*H*-tetrazole and tetrazolate obtained using intermolecular perturbation theory (IMPT),[20] together with similar energy data for H-bonds involving −COOH and −COO^- taken from the IsoStar library of intermolecular interactions.[21] We also examine the prevalence of tetrazole containing ligands and tetrazole/carboxyl isosteres in the PDB,[19] and additionally we have retrieved tetrazole containing drugs among the FDA-approved and experimental drugs contained in DrugBank.[22]

MATERIALS AND METHODS

Cambridge Structural Database Analyses. These analyses were carried out using CSD version 5.32 (November, 2010) plus one distributed update. Substructure searches and the location of H-bonds used the CSD system program ConQuest[23] with the following secondary search criteria: (i) atomic coordinates error free after CSD checks; (ii) no disorder in the crystal structure; (iii) no catena bonding; (iv) no powder studies; (v) organic structures only according to CSD definitions; and (vi) a crystallographic R-factor <0.075 (unless stated otherwise). Hydrogen bonds from N−H and O−H donors were located and geometrically characterized using neutron-normalized H-atom positions,[24] using mean N−H and O−H bond distances of 1.015 and 0.993 Å, respectively, as determined from accurate neutron structures in the CSD.[25] Substructure searches required for the conformational analyses were straightforward, but the H-bond analyses required some care, and additional notes are provided here.



(II)



(III)

Substructure searches were used to locate 1*H*-tetrazole rings (II) and deprotonated tetrazolate rings (III) in structures having at least one N−H or O−H donor (this requirement is obviously satisfied internally in 1*H*-tetrazole itself). Throughout these searches, any substituent atom X was allowed to be connected to the ring C_5 atom via an acyclic bond. Nonbonded contact searches were then carried out to locate H-bonds of the following types (where N/O means N or O as either acceptor or donor): (i) N1−H···N/O in 1*H*-tetrazoles; (ii) N2 or N3 or N4···H−N/O in both 1*H*-tetrazoles and in 1Z-tetrazoles (i.e., tetrazoles having any substituent Z = C/N/O at N1 and where N1−Z is acyclic); and (iii) N1 or N2···H−N/O in the topologically symmetrical tetrazolates. N···H H-bonds were required to be less than the sum of van der Waals radii, taken as N = 1.55 Å, H = 1.20 Å[26,27] and with the H-bond angle N···H−N/O greater than 120°.[28] On many occasions the donor hydrogen, D-H, was within the sum of van der Waals radii ($\sum$vdW) of two neighboring tetrazole or tetrazolate N-acceptors but exhibits a short contact to one of these N-atoms and a much longer one to its N-neighbor in the ring. The shortest D-H···N distance and its associated angular geometry were retained as describing the true H-bond.

Use of the IsoStar Library of Intermolecular Interactions. IsoStar[21] is a library of visual and numerical information about nonbonded interactions. IsoStar provides 3D scatterplots showing the distribution of a chemical contact group (e.g., a H-bond donor or other functional group) around a chemical central group (e.g., a cyclic or acyclic unit). IsoStar data are derived from (i) the CSD (>20 000 scatterplots), (ii) protein−ligand complexes in the PDB having a resolution better than 2 Å (>5500 scatterplots), and (iii) ab initio theoretical energy minima calculated using IMPT[20] for more than 1500 key interactions. Currently, IsoStar contains structural information for 48 contact groups and 300 central groups, and this information is updated annually. The library also provides geometry statistics, sliders to adjust nonbonded distance ranges, and hyperlinking of all scatterplot points back to the original CSD or PDB structures. Where a specific contact group−central group pairing is not available within the precomputed library, the relevant scatterplots can be computed directly from the CSD using the associated software module, IsoGen.

Ab initio Calculations of H-Bond Energies. Energies for H-bonds involving 1*H*-tetrazole and tetrazolates have been

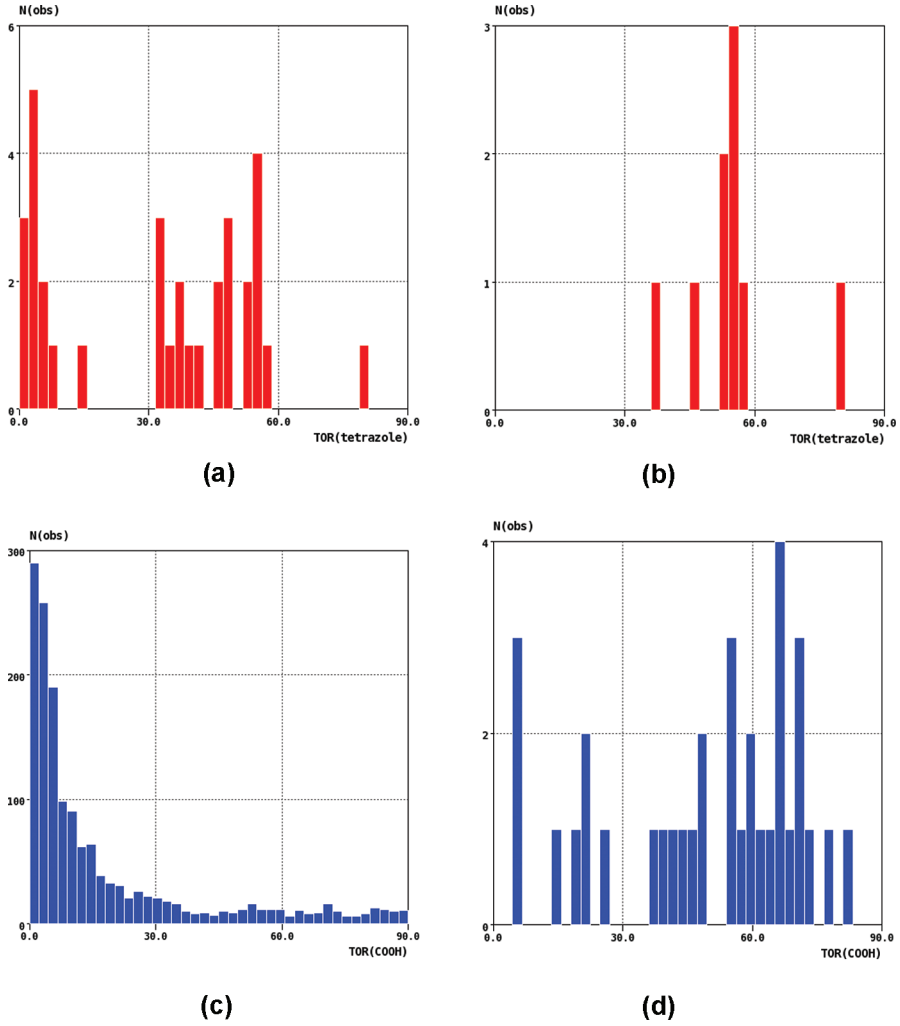


Figure 1. CSD[16] conformational distributions of 1*H*-tetrazole substituents attached (a) to any phenyl ring and (b) at ortho-positions in biphenyl systems (as in losartan (I) and its analogues). These distributions are compared with those for analogous −COOH substituents, i.e., (c) to any phenyl ring and (d) at ortho-positions in biphenyl systems.

calculated using the IMPT procedure,[20] developed at the University of Cambridge and implemented in the CADPAC6.5 program package.[29] The IMPT methodology yields separate energy components (first order: electrostatic and exchange−repulsion energies and second order: polarization, charge-transfer and dispersion energies) which sum to a total interaction energy (E_t) that is free of basis set superposition errors.[30] Various basis sets are available, and the 6-31G** basis set was used here. Calculations used model systems with the O−H of methanol and the O = C of acetone as model donors and acceptors as required.

Electrostatic Potential Surfaces. These surfaces were calculated using MOPAC[31] accessed via the Mercury program[32] and using the AM1 Hamiltonian. Display levels of electrostatic potential are given in the legend to Figure 2.

DrugBank and PDB Searches. DrugBank[22] was searched in June 2011, when the database contained 1437 FDA-approved small molecule drugs and 5174 experimental drugs. The PDB[19] was searched using the graphical ligand substructure search facility within Relibase+[33] (version 3.1 of June 2011),[34] a database search, retrieval, and analysis system for protein−ligand complex structures.

RESULTS AND DISCUSSION

Conformational Analysis. The CSD was used to locate 1*H*-tetrazole groups bonded to phenyl rings. The histogram of the inter-ring torsion angle, τ(tetrazole) in Figure 1a, shows two broad conformational groupings having (i) the tetrazole almost coplanar with the phenyl ring and (ii) τ(tetrazole) in the range 30−60°. However, when the search was further restricted to tetrazole substituents in an ortho-position on a biphenyl framework, as in losartan (I), Figure 1b shows that torsion angles exclusively adopt the noncoplanar conformation (ii). The crystal structure of losartan itself (CSD code OCAHAC)[35] gives τ(tetrazole) = 52°. A similar CSD analysis for −COOH attached to a phenyl ring (Figure 1c) not only shows the expected strong preference for coplanarity of −COOH with the ring but also that other conformations are possible. However, when the CSD search is restricted to *ortho*-COOH substituents on biphenyl, the picture changes dramatically (Figure 1d), and values of τ(COOH) > 20° now predominate, giving −COOH conformations that are highly comparable with tetrazole conformation (ii) (Figure 1b). The crystal structure of the −COOH analogue of losartan is also present in the CSD as VURTIL,[36] and the −COOH here is twisted by 24.3° with respect to the

dx.doi.org/10.1021/ci200521k | *J. Chem. Inf. Model.* 2012, 52, 857−866

phenyl ring. Further out of plane twisting of −COOH is, of course, possible with rather small energy penalties.[37]

CSD data show that this overall conformational similarity is also apparent for the deprotonated tetrazolate and carboxylate substituents, which are likely to be operational under physiological conditions. The torsional distributions for $-COO^-$ attached to phenyl rings and as ortho substituents on biphenyl frameworks, respectively, are almost identical to the distributions for −COOH (Figures 1c,d) and are not shown here. Data for tetrazolate attached to phenyl rings are sparse, with just 22 instances. Of these, only 2 rings are ortho-substituents on biphenyl frameworks and have τ(tetrazole) of 49.6° and 53.7°; the remaining 20 examples have τ(tetrazole) <22° with the majority (18) being <10°.

H-Bonding of 1*H*−Tetrazole and Tetrazolate Groups in Crystal Structures. While the H-bonding characteristics of many N-heterocyclic five-membered rings in crystal structures have been studied previously,[38] and while rings containing up to three N atoms are covered as central groups in IsoStar,[21] the H-bonding capabilities of rings containing four hetero-N atoms do not appear to have been analyzed using CSD information.

There are 80 CSD structures containing 105 independent examples of 1*H*-tetrazole. Of these, 74 structures contain N1−H···(N/O) H-bonds and 6 contain N1−H···Cl^- H-bonds (one structure appears anomalous and is omitted from the analysis). Thus, 100% of the highly acidic N1−H protons in 1*H*-tetrazole form H-bonds, as shown in Table 1. This table also summarizes

Table 1. Percentage of Acceptor N and Donor N−H Forming H-bonds in 1*H*-Tetrazole (N1−H) and its N1−Z Derivatives (Z = C, N, O) and in Tetrazolates[a]

atom	*N*(frag)	*N*(HB)	%(HB)
N1−H-tetrazoles			
N1−H	105	105	100%
H···N2	105	16	15.2%
H···N3	105	43	41.0%
H···N4	105	66	62.9%
N1−Z-tetrazoles (Z = C, N, O)			
H···N2	78	8	10.3%
H···N3	78	26	33.3%
H···N4	78	53	67.9%
N1−H and N1−Z tetrazoles (overall)			
H···N2	183	24	13.1%
H···N3	183	69	37.7%
H···N4	183	119	65.0%
Tetrazolates			
H···N1	232	228	98.3%
H···N2	232	218	94.0%

[a] *N*(frag) is the number of unique fragments (acceptors or donors) that could form H-bonds and *N*(HB) and %(HB) are the number and the percentage, respectively, of these unique fragments that actually form H-bonds. Data are from crystal structures in the CSD.[16]

the H-bond acceptor ability of the other three N-atoms in 1*H*-tetrazole determined from the CSD. This analysis shows clear differences in their acceptor abilities, which follow the order N4 (65.0% of which form H-bonds overall) > N3 (37.7%) > N2 (13.1%). We also analyzed the number of H-bonds accepted by the N2, N3, and N4 atoms of individual 1*H*-tetrazole rings and found that all three acceptors were used by only 11.7% of rings, two acceptors were used by 31.2% of rings, one acceptor (almost always N4) by 45.5% of rings, and that 11.7% of rings did not accept H-bonds at all.

In Table 2, we summarize the mean crystal structure geometry of the H-bonds formed (i) to N or O by the N1−H donor of 1*H*-tetrazole, (ii) the H-bonds formed by (N/O)−H donors to the N2, N3, N4 acceptors of 1*H*-tetrazole, and (iii) the H-bonds formed by (N/O)−H donors to the two symmetry independent N-acceptors of tetrazolate. The *d*(H···N) distances involving 1*H*-tetrazole in Table 2 decrease in the order N2 > N3 ≈ N4 > N1 for both N−H and O−H donors. The crystallographic evidence presented here clearly indicates that energies for the H-bonds from and to 1*H*-tetrazole are likely to be in the order N1−H···N/O > N/O−H···N4 ≈ N/O−H···N3 > N/O−H···N2, with H-bonds involving N2 being very much weaker than the others. This conclusion is reinforced by the H-bond angles (θ, the angles at H in each bond). These angles prefer to be linear (θ = 180°) for the strongest bonds, with H-bond energies decreasing quite rapidly with decreasing θ.[28] Thus, in Table 2, the strongest bonds all have median θ values >165°, while the median θ for the obviously weaker bonds to N2 is only 145.5°.

A total of 91 CSD structures contain tetrazolate rings and an O−H or N−H donor, and these structures contain 116 independent rings. The symmetry of the ring means that there are 232 available N1 (≡N4) or N2 (≡N3) acceptors. Table 1 shows that very high percentages of both the N1 and N2 acceptors form N−H···N and O−H···N bonds within the van der Waals radii limit. The analysis showed two reasons why N1 or N2 of tetrazolate do not accept H: (i) There are other competing strong acceptors, such as >C═O, nitro-O, etc., and/or (ii) only one or two donor (N/O)−H are available in the structure. H-bond distances to the symmetry-independent N1 and N2 atoms are included in Table 2. The distances are similar to each other and to the distances involving the N3 and N4 acceptors in 1*H*-tetrazole, indicating similar energies for all of these H-bonds. As expected, bonds to both of the tetrazolate acceptors have a high median θ value >165° (Table 2). We have also determined how many N-acceptor atoms in each tetrazolate ring actually form H-bonds simultaneously in crystal structures: All four N-acceptors form H-bonds in 52.6% of rings, and the percentages of rings having 3, 2, 1, and 0 active N-acceptors are 21.2%, 16.8%, 7.3%, and 2.2% respectively. Thus, provided there are sufficient donors, tetrazolate is a highly effective acceptor of H-bonds.

For completeness, we note that the tautomeric 2*H*-tetrazole ring occurs in just 12 crystal structures in the CSD. In all cases, the N2−H donor forms H-bonds, either to other ring N-atoms or to O- acceptors in other parts of the structure. However, the data for H-bonding to the N-acceptors of the ring are too sparse for meaningful analysis. 2*H*-tetrazole substituents will, nevertheless, deprotonate *in vivo* in exactly the same way as 1*H*-tetrazole.

Electrostatic Potential Surfaces. We began the computational part of the study by calculating electrostatic potential surfaces using MOPAC/AM1[31] for carboxyl/carboxylate and 1*H*-tetrazole/tetrazolate. The surfaces are compared in Figure 2. Given that H-bonds have a high-electrostatic component, it is clear from Figure 2 that the charged species (b,d) contain more potent H-bond acceptors than their neutral counterparts. Further, there is excellent qualitative agreement between the H-bond analysis from crystal structures, presented above, and the computed surfaces. Thus, the surfaces suggest that: (i) the N1−H···O bonds from 1*H*-tetrazole are stronger than the O−H···O bonds formed by the carboxyl group, (ii) the strength

Table 2. H-Bonded Interactions of 1*H*-Tetrazoles and Tetrazolates in Crystal Structures in the CSD.[16] Table gives mean H-bond geometry involving X = N/O acceptors and N/O−H donors in 1*H*-tetrazoles (N1−H), N1−Z-tetrazoles (Z = C, N, O, H) and in tetrazolates. For each H-bond, the columns show *d* (the mean H···acceptor distance in Å using neutron-normalized H-atom positions as discussed in the text), *D* (the mean N···N/O distance in Å), and *θ* (the median H-bond angle at H in N−H···N/O or N/O−H···N in degrees. Separate *d* and *D* values are given for N and O, respectively (with esd's and numbers of observations in parentheses), but the median *θ* is given as an overall value for both N and O as acceptor or donor

	d(X = N)	*d*(X = O)	*D*(X = N)	*D*(X = O)	Θ(X = N/O)
N1−H Tetrazoles					
N1−H···N/O	1.85(0.08,65)	1.77(0.15,32)	2.83(0.06,65)	2.75(0.11,32)	166.6
N1−Z Tetrazoles					
N/O−H···N2	2.29(0.21,14)	2.24(0.24, 5)	3.15(0.19,14)	3.07(0.14,5)	145.5
N/O−H···N3	1.98(0.12,39)	1.92(0.06,24)	2.96(0.11,39)	2.88(0.06,24)	164.4
N/O−H···N4	1.94(0.13,82)	1.86(0.09,33)	2.92(0.12,82	2.84(0.09,33)	166.6
Tetrazolates					
N/O−H···N1	1.99(0.14,139)	1.89(0.12,63)	2.96(0.12,139)	2.86(0.11,63)	166.5
N/O−H···N2	2.03(0.15,141)	1.90(0.15,54)	3.00(0.12,141)	2.87(0.13,54)	165.6

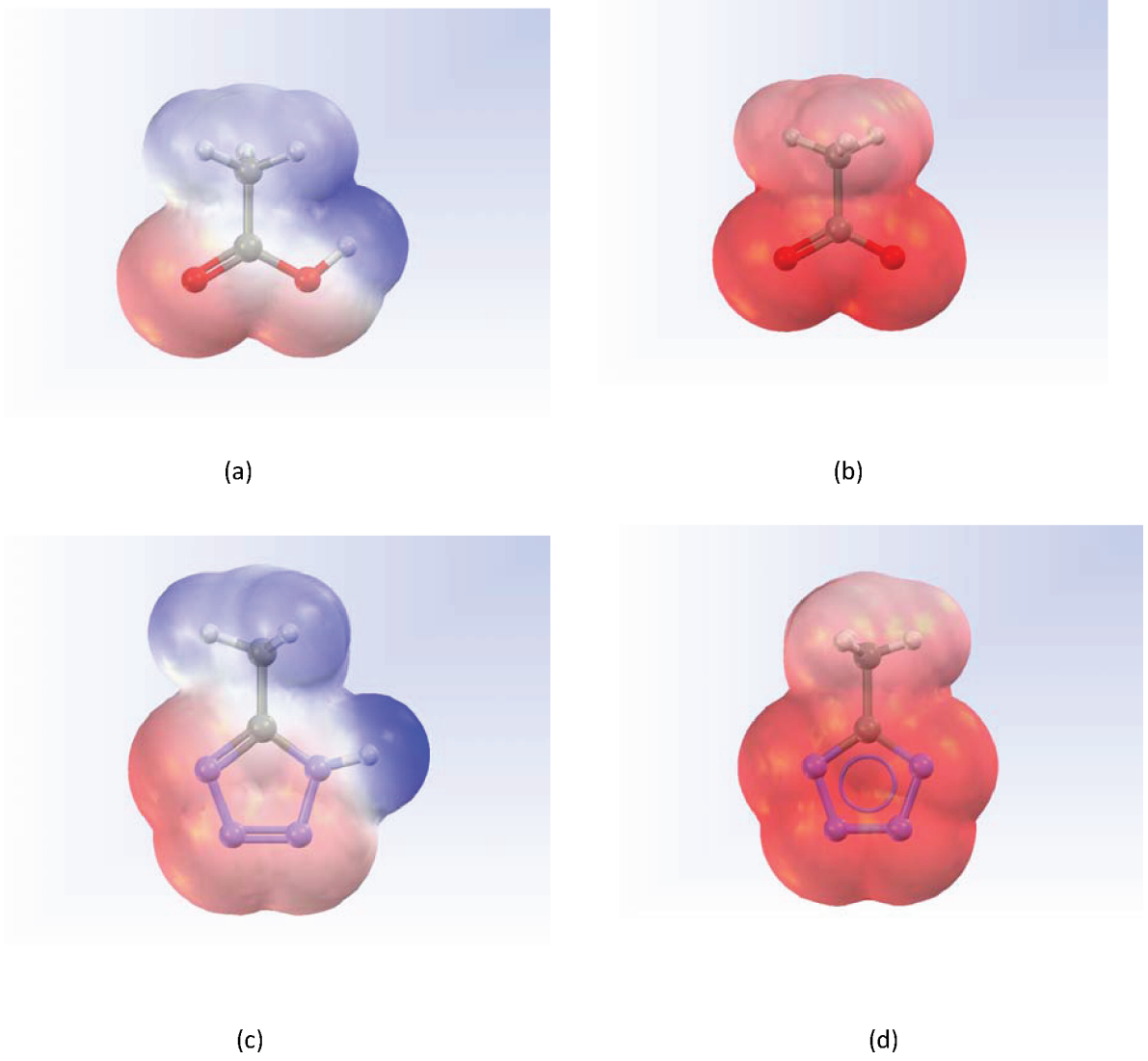


Figure 2. Electrostatic potential surfaces computed using MOPAC/AM1[29] for (a) −COOH and (b) $-COO^-$ groups are compared with the surfaces for (c) 1*H*-tetrazole and (d) tetrazolate moieties. For neutral species (a,c) the color distribution runs from +0.1e (blue) through 0 (white) to −0.1e (red). For charged species (b,d) the color distribution runs from +0.3e (blue) through 0 (white) to −0.3e (red).

of H-bonds accepted by the 1*H*-tetrazole N-atoms will be different and will likely be in the order N4 > N3 > N2, as indicated by the crystal structure data, and (iii) the two independent N-acceptors of tetrazolate will form strong H-bonds that are similar in strength to those formed by the carboxylate O-acceptors.

dx.doi.org/10.1021/ci200521k | *J. Chem. Inf. Model.* 2012, 52, 857−866

Ab initio Calculations of H-Bond Energies for 1*H*-Tetrazole and Tetrazolates. H-bond energies were calculated using the IMPT procedure[20] applied to the following model systems: (i) (1*H*-tetrazole)N1−H···*O* = C(acetone) and (methanol)O−H···N2,3,4(1*H*-tetrazole) and (ii) (methanol)-O−H···N1,2(tetrazolate, symmetric). The pairs of molecules involved in each calculation were first optimized, and the geometry of the H-bonded dimers was specified using mean values for H-bond distances and angles determined from the analysis of crystal structure data. The final calculated H-bond energies (E_t) are presented in Table 3 together with their IMPT energy components.

For 1*H*-tetrazole, the energy values of Table 3 closely mimic the H-bonding propensities deduced from crystal structure data (Table 1) for the N-donor and N-acceptors of the heterocycle. Thus, the N1−H donor in 1*H*-tetrazole is exceptionally potent, with N1−H···O═C having an energy of −44.0 kJ·mol^{-1} and a 100% propensity for H-bond formation. The N-acceptors, N4, N3, N2, which have propensities for H-bond formation of 63%, 41%, and 15% respectively, have energies that closely follow this decreasing trend at −27.3, −19.0, and −17.2 kJ·mol^{-1}. For the deprotonated tetrazolate, the H-bond energies for the two symmetry-independent ring nitrogens, N1 and N2, are very similar and very strong at −62.4 and −58.8 kJ·mol^{-1} respectively, values which correspond to the crystallographically observed propensities of H-bond formation of 98% and 94%. We also note the close agreement between the qualitative indications provided by the electrostatic potential surfaces (Figure 2) and the quantitative IMPT data given in Table 3. Given the rather small number of crystal structures in the CSD containing 1*H*-tetrazole and the deprotonated tetrazolate, the level of alignment of the experimental observations with the MOPAC/AM1 surfaces and the ab initio calculations is highly encouraging.

Table 3. IMPT (ref 18) Energies for H-Bonds Involving 1*H*-Tetrazole and Tetrazolate Together with IMPT H-Bond Energies Involving −COOH and −COO$^-$ Taken from the IsoStar Knowledge Base (ref 19)[a]

H-bond	E_{es}	E_{er}	E_{pol}	E_{ct}	E_{disp}	E_t
1*H*-Tetrazole						
N1−H···O═C (acetone)	−63.6	53.3	−9.1	−8.1	−16.5	−44.0
N2···H−O (methanol)	−45.7	50.8	−4.2	−5.8	−12.0	−17.2
N3···H−O (methanol)	−46.4	51.1	−5.3	−6.4	−11.9	−19.0
N4···H−O (methanol)	−54.5	53.1	−5.9	−6.9	−13.2	−27.3
Tetrazolate						
N1···H−O (methanol)	−65.2	33.1	−10.8	−9.7	−9.9	−62.4
N2···H−O (methanol)	−62.0	32.1	−10.8	−9.0	−9.1	−58.8
carboxyl −COOH						
CO−H···O (acetone)	−49.2	42.1	−7.0	−5.9	−14.0	−33.9
C−O···H−O (methanol)	−21.1	16.5	−1.7	−1.8	−6.2	−14.3
C═O···H−O (methanol)[b]	−23.4	18.5	−3.1	−2.7	−7.0	−17.7
C═O···H−O (methanol)[c]	−35.1	26.4	−3.5	−3.4	−8.3	−24.0
carboxylate −COO$^-$						
C═O···H−O (methanol)[b]	−92.0	69.6	−18.3	−15.6	−15.3	−71.6
C═O···H−O (methanol)[c]	−93.5	64.1	−19.5	−15.2	−13.6	−77.7

[a]IMPT energies are in kJ·mol^{-1}; E_t is the total energy arising from the separate components: electrostatic (E_{es}), exchange−repulsion (E_{er}), polarization (E_{pol}), charge-transfer (E_{ct}), and dispersion (E_{disp}). [b]H-bond directed toward the *E* lone pair. [c]H-bond directed toward the *Z* lone pair.

H-Bonding Characteristics of −COOH and −COO$^-$ Groups. Because of their chemical ubiquity, H-bonds involving the carboxylic acid and carboxylate groups were among the earliest to be studied[39−42] using both crystal structure data and computational techniques. Regularly updated distributions and mean values of H-bond geometry involving these groups are available within the annual releases of the IsoStar knowledge base.[21] IsoStar also contains IMPT energies for selected model systems involving −COOH and −COO$^-$ groups, and these data are included in Table 3 for comparison with the tetrazole and tetrazolate H-bond energy data determined above. Again, the qualitative electrostatic potential surfaces of Figure 2 are reinforced by the quantitative IMPT data.

Comparative H-Bond Environments for 1*H*-Tetrazole/−COOH and Tetrazolate/−COO$^-$. Comparisons of H-bond environments in crystal structures have been carried out using the IsoStar/Isogen system[21] summarized in Materials and Methods Section. We have examined and compared the H-bonded interactions of both the protonated isosteric pair 1*H*-tetrazole/carboxyl and the deprotonated tetrazolate/carboxylate forms. The latter pair is most likely to be physiologically relevant. The distributed IsoStar library covers the interactions of the common −COOH and −COO$^-$ central groups with a wide variety of hydrogen-bonding contact groups, but it does not cover 1*H*-tetrazole or tetrazolate. Thus, for these heterocyclic central groups, we have used the IsoGen software to generate scatterplots depicting the spatial distributions of H-bonds with (N/O)-H donors, since their acceptor and donor capabilities are likely to be responsible for their principal intermolecular interactions, e.g., within protein active sites.

Figure 3 shows comparative contoured IsoStar plots for N−H and O−H donor groups around −COOH and −COO$^-$ (a,b) and IsoGen plots for N−H and O−H around 1*H*-tetrazole and tetrazolate (c,d). Peaks also occur in Figure 3a,c which show donor-H···(N/O) for the carboxyl and 1*H*-tetrazole groups. The plots highlight the key H-bond interaction areas and the preferred directions of (N/O)-H approach to each central group. There are clear similarities in the general distributions of the donor (N/O)-H groups around the protonated species (compare Figure 3a,c). H-bonds to the carbonyl-O1 of −COOH form along O1-lone pair directions (Figure 3a), and this distribution is mimicked by the H-bonds to N4 and N3 of 1*H*-tetrazole in Figure 3c. If we consider tetrazole-N4 as the structural mimic of carboxyl-O1, then one of the H-bonds to tetrazole-N4 corresponds directly to the H-bonds to the E-lone pair of carboxyl-O1 (the leftmost peak in Figure 3a,c). The second H-bond to carboxyl-O1 (to its Z-lone pair at the bottom of Figure 3a) is then mimicked by the H-bond to tetrazole-N3. However, the peak for this H-bond in Figure 3c is now ≈1.2 Å further from N4, due to the replacement of the single acceptor, carboxyl-O1, by N4, N3 of the tetrazolate acting as a 'double acceptor'.

There is also energetic similarity between the two H-bonding environments. Thus, the strongest H-bonds formed by

dx.doi.org/10.1021/ci200521k | J. Chem. Inf. Model. 2012, 52, 857−866

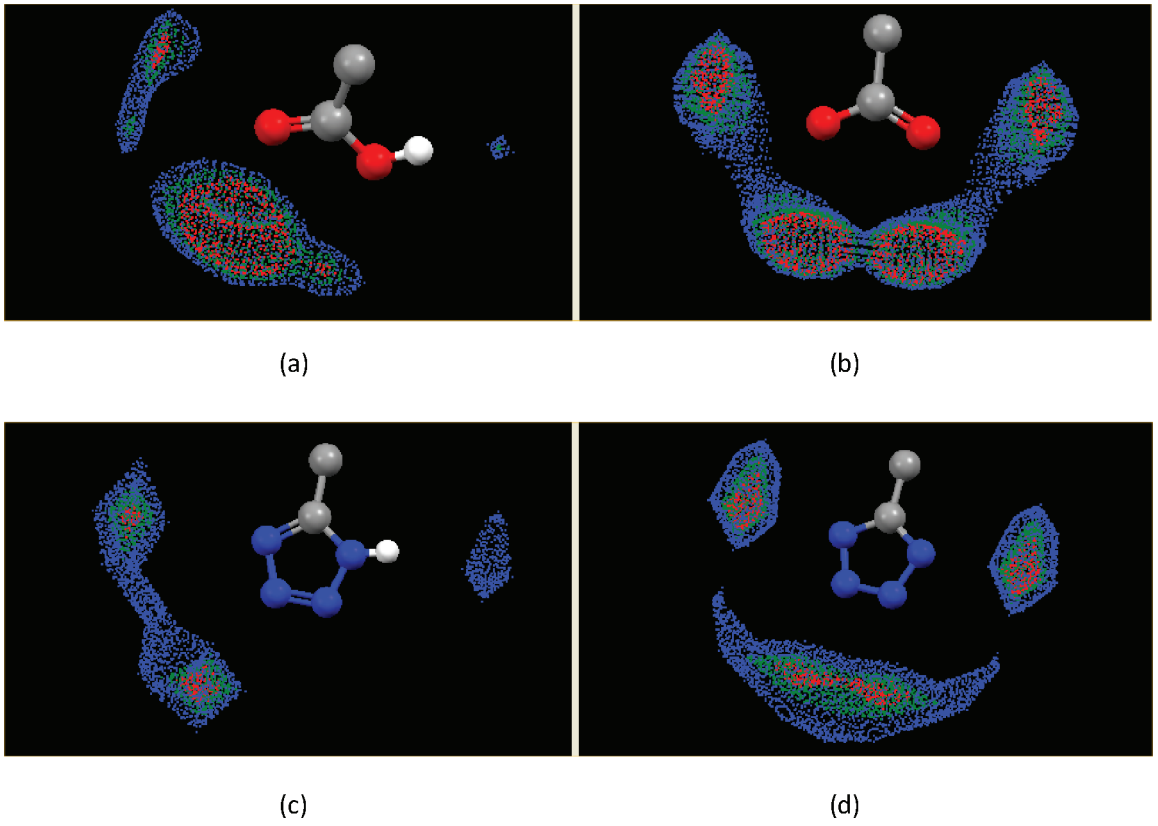


Figure 3. Distributions of N−H and O−H donors around (a) −COOH and (b) $-COO^-$ groups (CSD data from IsoStar),[19] compared with distributions of N−H and O−H donors around (c) 1*H*-tetrazole and (d) tetrazolate moieties generated from the CSD using IsoGen software distributed with the IsoStar library.

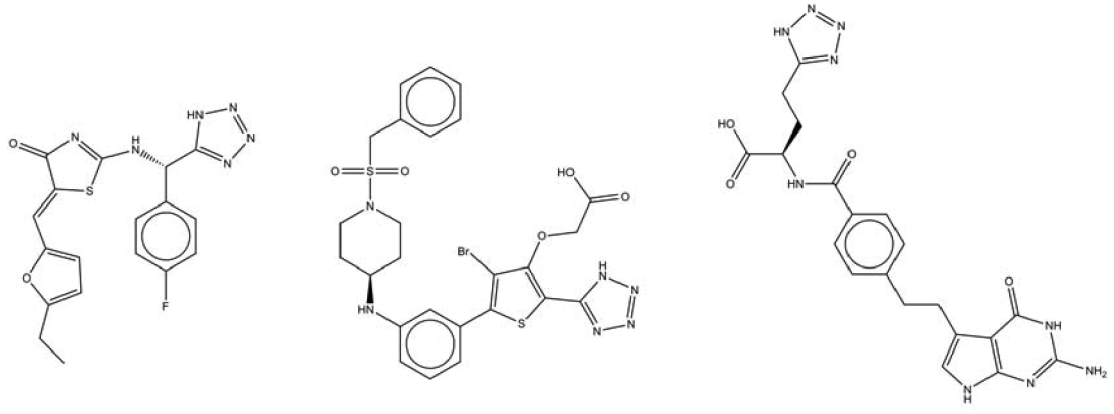


Figure 4. Chemical diagrams of the 1*H*-tetrazole containing ligands in the PDB for which a carboxyl analogue also occurs in the PDB and bound to the same protein.

1*H*-tetrazole and −COOH involve their N1−H and O−H donors, respectively, although we note that bonds involving N1−H (tetrazole) are significantly stronger (by about 10 $kJ \cdot mol^{-1}$) than those involving the carboxyl O−H donor. The energies of the H-bonds to the tetrazole N4 and N3 atoms are very similar to the energies of H-bonds to the C═O of the carboxyl group.

The comparison of H-bonding environments around the deprotonated species (Figure 3b for carboxylate and Figure 3d for tetrazolate) is even more compelling. Here, the double acceptor pairs N1, N2 and N1′, N2′ of the tetrazolate can be regarded as structural mimics for the O1 and O2 atoms of $-COO^-$, as described above for the protonated species. Once again, one of the H-bonds accepted by each of the tetrazolate N1 and N1′ atoms is a direct mimic of the H-bonds accepted by the E-lone pairs of carboxylate O1, O2 (i.e., the leftmost and rightmost peaks in Figure 3d). However, the H-bonds accepted by each of the tetrazolate N2 and N2′ atoms extend the tetrazolate

dx.doi.org/10.1021/ci200521k | J. Chem. Inf. Model. 2012, 52, 857−866

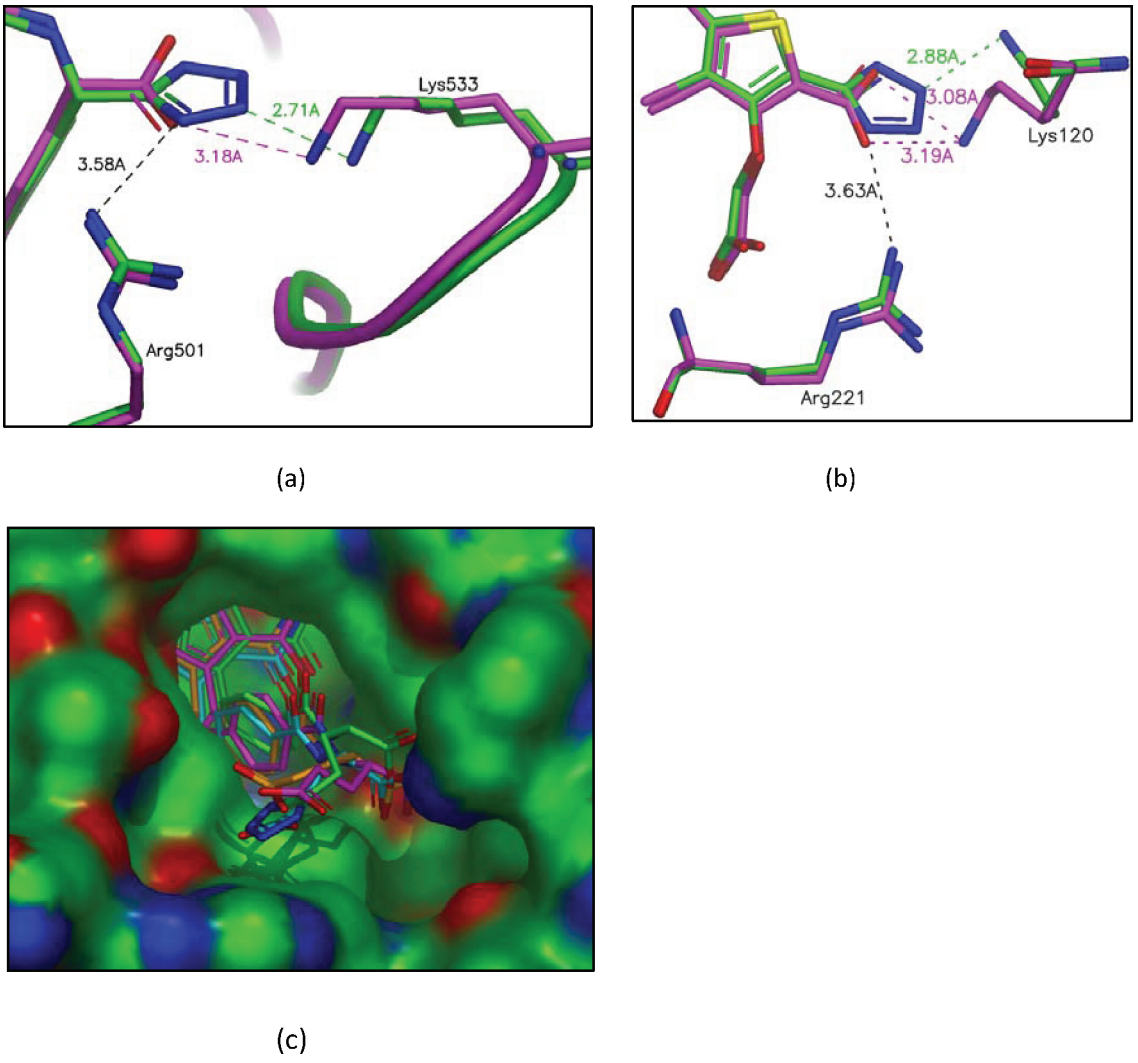


Figure 5. Superimpositions of the ligand binding sites for the tetrazolate:carboxylate ligand pairs (see text) (a) VXR/VRX bound to hepatitis C NS5B polymerase, (b) 902/527 bound to tyrosine phosphatase 1B, and (c) LY3/LYA bound to *E.coli* thymidylate synthase.

H-bond environment (downward in Figure 3d) by ≈1.2 Å when compared to the H-bond environment of the carboxylate group. In energetic terms, the H-bonds to the carboxylate O-atoms and the tetrazolate N-atoms are all very strong, with a minimum attractive energy of around −60 kJ·mol^{-1}, but Table 3 shows that the bonds to carboxylate-O are stronger than those to tetrazolate-N by around 15−18 kJ·mol^{-1}.

Thus, the H-bonded interaction environments of the 1*H*-tetrazole/carboxyl and the deprotonated tetrazolate/carboxylate pairs are topologically and energetically very similar, and this similarity forms the structural basis for the observed isosterism. However, there is also an expansion in the H-bonding environment in going from carboxyl or carboxylate to 1*H*-tetrazole or tetrazolate, respectively. Thus, while the disposition of H-bond acceptors and donors within a protein active site may well accommodate either isostere, it is possible that certain sites may have inbuilt specificity for one or other of the isosteric groups. We have investigated this possibility using data from the PDB.[19]

PDB Search and analysis. Substructure searches of PDB ligands using Relibase+[33,34] located 28 ligands having either an unsubstituted 1*H*- or 2*H*-tetrazole substituent, 16 being drawn as 1*H*- and 12 as 2*H*-. In three cases, we located structures of the same protein liganded with both tetrazolate and carboxylate, and we have compared the ligand binding in all three cases. The three pairs of ligands, for which chemical diagrams are shown in Figure 4, are as follows (tetrazolate and carboxylate ligand codes and PDB structure code(s) in parentheses): (a) VXR (in 2i1r),[43] VRX (in 2hwi);[44] (b) 902 (in 2nt7),[45] 527 (in 2qbp);[46] and (c) LY3 (in 1jtq),[47] LYA (in 1juj and 1ju6).[47] The proteins involved in these interactions are (a) hepatitis-C NS5B polymerase, (b) tyrosine phosphatase 1B, and (c) *E.coli* thymidylate synthase.

Superimpositions of the binding of the ligand pairs (a−c) in the active sites in the respective proteins are illustrated in Figure 5. In Figure 5a,b the principal interactions between the proteins and the ligands are indicated, and in both cases, these data show that the protein is able to flex in order to accommodate the larger tetrazolate ligand, whose H-bond environment is extended by ≈1.2 Å by comparison to its carboxylate analogue, as discussed above. In both cases, it would appear that by flexing in this way the protein is able to form stronger H-bonds to the tetrazolate than to the carboxylate. In the third example (Figure 5c), the ligands clearly bind through their pyrrolo[2,3]pyrimidine function at the opposite end of the molecule to the tetrazolate/carboxylate substituent.

dx.doi.org/10.1021/ci200521k | *J. Chem. Inf. Model.* 2012, 52, 857−866

DrugBank Searches. DrugBank[22] reports nine FDA approved drugs that contain 1*H*- or 2*H*-tetrazole substituents. Seven of these drugs are antihypertensives with closely related chemical structures: valsartan, olmesartan, losartan, candesartan, irbesartan, forasartan, and tasosartan. The other two are the antiallergic drugs: pemirolast and pranlukast. A further 12 compounds were retrieved from the experimental drugs section of the database and have a variety of specificities. When the search was extended to include N-substituted tetrazoles, a further 11 approved drugs were retrieved. Ten of these are second or third generation β-lactam-containing cephalosporin antibiotics, while the other is a drug to treat intermittent claudication. A further nine experimental drugs contain N-substituted tetrazoles. By comparison, DrugBank shows that 250 approved drugs contain a −COOH function, while there are 748 instances of −COOH in the experimental drug section of the database. Although these statistics might not appear to provide strong support for the use of tetrazole as a bioisostere, we note that many of the −COOH compounds are older and more well established drugs, and any therapeutic benefits of tetrazole analogues may not be sufficient to warrant development. Other factors too, such as cost of process, could also mitigate against the tetrazole analogues. We also note that the tetrazole group is not present in nature. With many drugs having their origins in endogenous ligands, this might also explain this observation. Thus, the existence of even as few as nine FDA-approved tetrazole-containing drugs indicates the value of this particular bioisostere in drug discovery, and the statistics derived from DrugBank may indicate that this bioisosterism may indeed be underused.

■ CONCLUSION

This paper has used crystal structure informatics and high-level ab initio calculations to study the H-bond landscapes of the bioisosteric functional group pairs 1*H*-tetrazole and COOH and their deprotonated analogues tetrazolate and carboxylate. To our surprise, this is the first comprehensive report of the H-bond landscapes for this N_4-heterocyclic system. Both bioisosteric pairs exhibit very similar H-bond environments in crystal structures from the CSD, and the attractive energies of these H-bonds are very similar. However, the H-bond environments around 1*H*-tetrazole and tetrazolate substituents extend further away, by ≈1.2 Å, from the core of the connected molecule than for the −COOH and $-COO^-$ analogues. Analysis of pairs of PDB structures, which contain ligands which differ only in having a tetrazolate or a carboxylate substituent and which are bound to the same protein, indicates that the protein binding site is able to flex sufficiently to form strong H-bonds to either substituent. This flexibility appears to be crucial in improving the binding of tetrazolate ligands, by comparison with carboxylate ligands, in the available examples in the PDB. The ≈1.2 Å expansion in the H-bond landscape in tetrazoles suggests that a similar effect might also be produced in carboxylates by use of an additional substituent 'spacer' atom (X), e.g., to give −X−COOH as a substituent, so as to mimic the position of the tetrazole and to take better advantage of active-site flexibility. Nevertheless, despite the fact that 1*H*-tetrazole/carboxyl bioisosterism has been frequently mentioned in the drug discovery literature, a brief survey of DrugBank shows an apparently rather small number of tetrazole-containing small molecules in the 'approved' and 'experimental' drug sections of that database. Perhaps carboxylate/tetrazolate isosterism is underused in current drug discovery strategies?

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: allen@ccdc.cam.ac.uk.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Friedman, H. L. Influence of Isosteric Replacements upon Biological Activity. *NAS-NRS Publication No. 206*; NAS-NRS: Washington, DC, 1951; Vol. *206*, pp 295−358.

(2) Wasserman, A. M.; Bajorath, J. Identification of target family directed bioisosteric replacements. *Med. Chem. Commun.* **2011**, *2*, 601−606.

(3) Thornber, C. W. Isosterism and molecular modification in drug design. *Chem. Soc. Rev.* **1979**, *8*, 563−580.

(4) Taylor, R.; Mullaley, A.; Mullier, G. W. Use of crystallographic data in searching for isosteric replacements: composite crystal-field environments of nitro and carbonyl groups. *Pestic. Sci.* **1990**, *29*, 197−213.

(5) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. N. RECAP − Retrosynthetic Combinatorial Analysis Procedure: A powerful new technique for identifyung privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511−522.

(6) Ertl, P. Cheminformatics analysis of organic substituents: Identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like isosteric groups. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374−380.

(7) Maass, P.; Schulz-Gasch.; Stahl, M.; Rarey, M.; Recore, A Fast and Versatile Method for Scaffold Hopping Based on Small Molecule Crystal Structure Conformations. *J. Chem. Inf. Model.* **2007**, *47*, 390−399.

(8) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic rings of the future. *J. Med. Chem.* **2009**, *52*, 2952−2963.

(9) Meanwell, N. A. Synopsis of some recent tactical applications of bioisosteres in drug design. *J. Med. Chem.* **2011**, *54*, 2529−2591.

(10) Juby, P. F.; Hudyma, T. W. Preparation and anti-inflammatory properties of some 1-substituted 3-(5-tetrazolylmethyl)indoles and homologs. *J. Med. Chem.* **1969**, *12*, 396−401.

(11) Myznikov, L. V.; Hrabalek, A.; Koldobskii, G. I. Drugs in the tetrazole series. *Chem. Heterocycl. Compd.* **2007**, *43*, 1−9.

(12) Herr, R. J. 5-Substituted-1*H*-tetrazoles as carboxylic acide isosteres: medicinal chemistry and synthetic methods. *Bioorg. Med. Chem.* **2002**, *10*, 3379−3393.

(13) Carini, D. J.; Duncia, J. V.; Aldrich, P. E.; Chiu, A. T.; Johnson, A. L.; Pierce, M. E.; Price, W. A.; Santella, J. B. III; Wells, G. Nonpeptide angiotensin II receptor antagonists: the discovery of a series of N-(biphenylylmethyl)imidazoles as potent, orally active antihypertensives. *J. Med. Chem.* **1991**, *34*, 2525−2547.

(14) Biot, C.; Bauer, H.; Schirmer, R. H.; Davioud-Charvet, E. 5-Substituted tetrazoles as bioisosteres of carboxylic acids. Bioisosterism and mechanistic studies on glutathione reductase inhibitors as antimalarials. *J. Med. Chem.* **2004**, *47*, 5972−5983.

(15) da Silva, G.; Moore, E. E.; Bozzelli, J. W. Quantum chemical study of the structure and thermochemistry of the five-membered nitrogen-containing heterocycles and their anions and radicals. *J. Phys. Chem. A* **2006**, *110*, 13979−13988.

(16) Chermahini, A. N.; Ghaedi, A.; Teimouri, A.; Momenbeik, F.; Dabbagh, H. A. Density functional theory study of intermolecular interactions of cyclic tetrazole dimers. *J. Mol. Struct. Theochem* **2008**, *867*, 78−84.

(17) Zhu, W.; Xiao, H. First-principles study of electronic structure, absorption spectra and thermodynamic properties of crystalline 1*H*-tetrazole. *Struct. Chem* **2010**, *21*, 847−854.

(18) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, *B58*, 380−388.

(19) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(20) Hayes, I. C.; Stone, A. J. An intermolecular perturbation theory for the region of moderate overlap. *Mol. Phys.* **1984**, *53*, 83−105.
(21) Bruno, I. J.; Cole, J. C.; Lommerse, J. P. M; Rowland, R. S.; Taylor, R.; Verdonk, M. L. IsoStar: A library of information about intermolecular interactions. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525−537.
(22) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39* (Suppl. 1), D1035−D1041.
(23) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr.* **2002**, *B58*, 389−397.
(24) Allen, F. H. A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron difraction. *Acta Crystallogr.* **1986**, *B42*, 515−522.
(25) Allen, F. H.; Bruno, I. J. Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. *Acta Crystallogr.* **2010**, *B66*, 380−386.
(26) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.
(27) Rowland, R. S.; Taylor, R. Intermolecular nonbonded contact distances in organic crystal structures: comparison with distances expected from van der Waals radii. *J. Phys. Chem.* **1996**, *100*, 7384−7391.
(28) Wood, P. A.; Allen, F. H.; Pidcock, E. Hydrogen-bond directionality at the donor H atom: Analysis of interaction energies and database statistics. *CrystEngComm.* **2009**, *11*, 1563−1571.
(29) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Cohen, A. J.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Laming, G. J.; Lee, A. M.; Maslen, P. E.; Murray, C. W.; Palmieri, P.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M.-D.;Tozer, D. J. *CADPAC 6.5. The Cambridge Analytical Derivatives Package: A suite of quantum chemistry programs*; Chemistry Department, Cambridge University: Cambridge, U.K., 1998.
(30) Stone, A. J. Computation of charge-transfer energies by perturbation theory. *Chem.Phys.Lett.* **1993**, *211*, 101−109.
(31) Stewart, J. J. P. *MOPAC2009*; Stewart Computational Chemistry: Colorado Springs, CO, 2009.
(32) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. Mercury CSD2.0 - new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **2008**, *41*, 466−470.
(33) Hendlich, M.; Bergner, A.; Guenther, J.; Klebe, G. Relibase - Design and development of a database for comprehensive analysis of protein - ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607−620.
(34) *Relibase+*; Cambridge Crystallographic Data Centre: Cambridge, U.K.
(35) Tessler, L.; Goldberg, I. Losartan, an antihypertensive drug. *Acta Crystallogr.* **2004**, *E60*, o1830−o1832.
(36) Bradbury, R. H.; Allott, C. P.; Dennis, M.; Fisher, E.; Major, J. S.; Masek, B. B.; Oldham, A. A.; Pearce, R. J.; Rankine, N.; Revill, J. M.; Roberts, D. A.; Russell, S. T. New nonpeptide angiotensin II receptor antagonists. 2. Synthesis, biological properties, and structure-activity relationships of 2-alkyl-4-(biphenylylmethoxy)quinoline derivatives. *J. Med. Chem.* **1992**, *35*, 4027−4038.
(37) Weng, Z. F.; Motherwell, W. D. S.; Allen, F. H.; Cole, J. M. Conformational Variability of Molecules in Different Crystal Environments: a Database Study. *Acta Crystallogr.* **2008**, *B64*, 348−362.
(38) Nobeli, I.; Price, S. L.; Lommerse, J. P. M; Taylor, R. Hydrogen bonding properties of oxygen and nitrogen acceptors in aromatic heterocycles. *J. Comput. Chem.* **1997**, *18*, 2060−2074.
(39) Taylor, R.; Kennard, O. Hydrogen-bond geometry in organic crystals. *Acc. Chem. Res.* **1984**, *17*, 320−326.
(40) Görbitz, C. H.; Etter, M. C. Hydrogen bonds to carboxylate groups. Syn/anti distributions and steric effetcs. *J. Am. Chem. Soc.* **1992**, *114*, 627−631.
(41) Görbitz, C. H.; Etter, M. C. Hydrogen bonds to carboxylate groups. The question of three-centre interactions. *J. Chem. Soc., Perkin Trans. 2* **1992**, pp131−135.
(42) Steiner, T. Competition of hydrogen-bond acceptors for the strong carboxyl donor. *Acta Crystallogr.* **2001**, *B57*, 103−106.
(43) Yan, S.; Larson, G.; Wu, J. Z.; Appleby, T.; Ding, Y.; Hamatake, R.; Hong, Z.; Yao, N. Novel thaizolones as HCV NS5B polymerase allosteric inhibitors: further designs, SAR and X-ray complex structure. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 63−67.
(44) Yan, S.; Appleby, T.; Larson, G.; Wu, J. Z.; Hamatake, R.; Hong, Z.; Yao, N. Structure-based design of a novel thiazolone scaffold as HCV NS5B polymerase allosteric inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5888−5891.
(45) Wan, Z.-K.; Follows, B.; Kirincich, S.; Wilson, D.; Binnun, E.; Xu, W.; Joseph-McCarthy, D.; Wu, J.; Smith, M.; Zhang, Y.-L.; Tam, M.; Erbe, D.; Tam, S.; Saiah, E.; Lee, J. Probing acid replacements of thiophene PTP1B inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2913−2920.
(46) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-based optimization of protein tyrosine phaphatase 1B inhibitors: from the active site to the second phosphotyrosine binding site. *J. Med. Chem.* **2007**, *50*, 4681−4698.
(47) Sayre, P. H.; Finer-Moore, J. S.; Fritz, T. A.; Biermann, D.; Gates, S. B.; MacKellar, W. C.; Patel, V. F.; Stroud, R. M. Multi-targeted antifolates aimed at avoiding drug resistance form covalent closed inhibitory complexes with human and *Escherischia coli* thymidylate synthases. *J. Mol. Biol.* **2001**, *313*, 813−829.