

# Charting the Chemical Space of Target Sites: Insights into the Binding Modes of Amine and Amidine Groups

Antonio Macchiarulo, Roberto Nuti, Gokcen Eren, and Roberto Pellicciari\*

Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, via del Liceo 1, 06127 Perugia, Italy

Received November 10, 2008

Nowadays there is growing awareness that the translation of the increasing number of lead compounds into clinical candidates is still a slow and often inefficient process. In order to facilitate the lead optimization procedure, due consideration must be given to the use of the right bioisosteric replacements. Very recently, we reported that exploring a chemical space of binding sites is a more effective strategy for studying the bioisosteric relationships existing among functional groups. As a continuation of our work in this field, we report herein the construction of a chemical space covered by binding sites of small molecules containing diverse amine and amidine groups. The analysis of the differences in some properties of the binding sites of these functional groups allow for gaining insights into the binding modes of positively charged groups. In addition, this study pinpoints that different types of interactions and bioisosteric relationships exist among primary, secondary, tertiary, quaternary amine, and amidine moieties.

## INTRODUCTION

The translation of lead compounds into drug candidates is one of the main challenges in drug-discovery pipelines. A successful translation is the result of multiproperty optimization procedures wherein lead compounds have to acquire high selectivity for a therapeutically relevant target and good bioavailability and have to avoid off-target promiscuous activities which are commonly associated with unwanted side effects. However, it should be mentioned that this process is still slow and often inefficient<sup>1</sup> as evidenced by the decline in the production of new chemical entities (NCEs) despite the ever increasing R&D investment on the part of both industry and governments.<sup>2</sup>

The bioisosteric replacement is the technique of choice used by medicinal chemists to improve the properties of a lead molecule in order to selectively retain its original activity to the therapeutically relevant target while removing unwanted off-target effects and poor pharmacokinetic profiles.

In the quest for bioisosteric groups, diverse computational methodologies have been developed to chart the functional group space with the aim of identifying moieties with similar biological properties. These approaches have been recently reviewed by Ertl<sup>3</sup> and comprise the calculation of property descriptors, mining of literature data and analysis of molecular database.

More recently, we have revisited the concept of bioisosterism by pursuing the idea that binding sites with similar geometrical and/or chemical properties would recognize similar functional groups or molecules. Thus we have introduced the concept of a “binding site-centric” chemical space as complementary to the “ligand-centric” chemical space.<sup>4</sup>

So far, the analysis of binding sites has been exploited for the identification of preferred interaction regions of atom

probes,<sup>5–7</sup> the definition of different types of molecular recognitions,<sup>8,9</sup> and the prediction of protein functions.<sup>10,11</sup> Very recently, properties of binding sites have been studied with the aim to infer about the geometry of interactions between catalytic residues and substrates and to assess the specificity and/or promiscuity of ligand recognition in protein families.<sup>12–15</sup>

In our first endeavor of charting the chemical space of target sites for bioisosterism, we have analyzed the chemical properties of binding sites recognizing small molecules that contained acidic groups such as carboxylic, sulfonic, or phosphonic groups.<sup>4</sup> The identification of areas in the space diversely populated by the binding sites of these three types of acidic groups provided insights into the features of a selective binding mode of carboxylic, sulfonic, and phosphonic moieties.

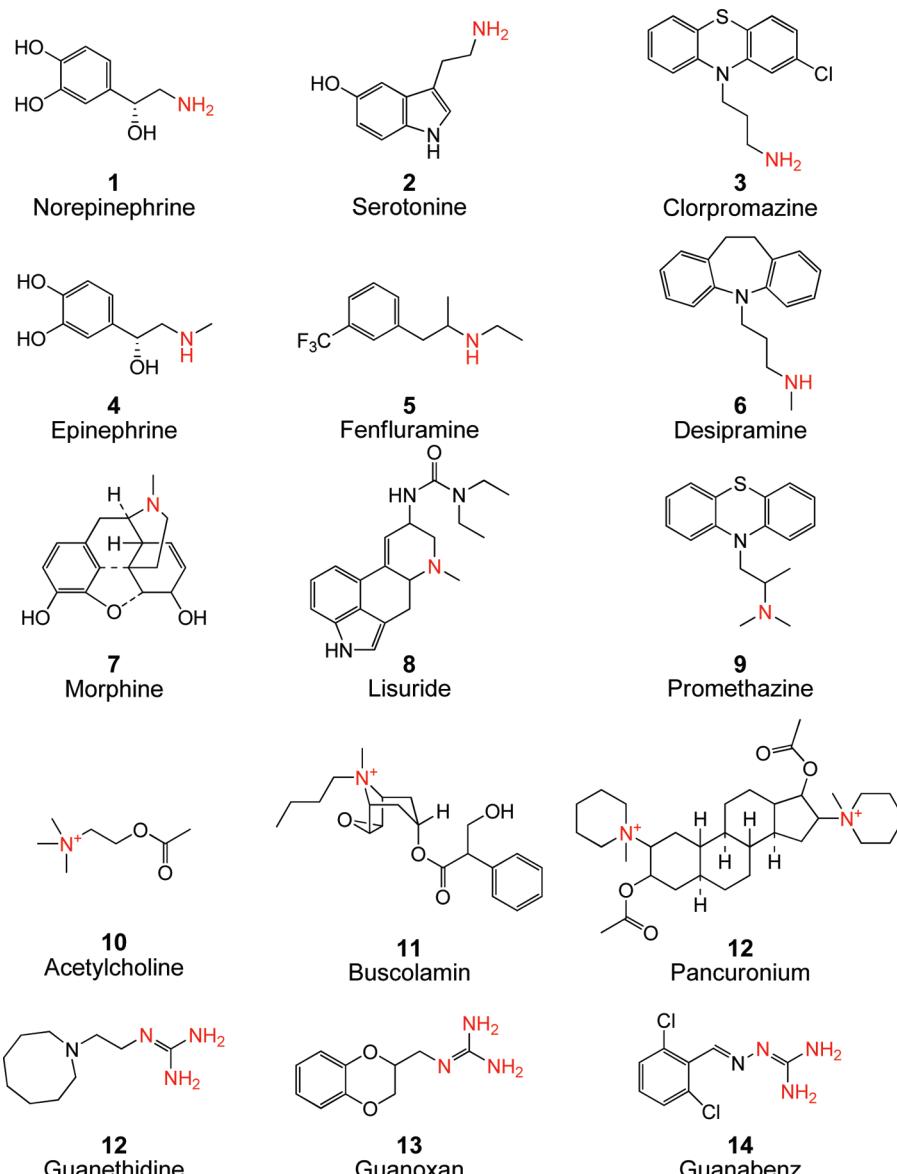
As a continuation of our effort in the study of bioisosteric relationships of functional groups by using the analysis of their relative binding sites, herein we report the construction of the chemical space covered by binding sites of small molecules containing amine and amidine groups such as primary, secondary, tertiary, quaternary amines, and guanidine moieties.

Although these groups are found in endogenous biologically active substances and have been widely used to develop successful drug candidates in many therapeutic areas (Figure 1), few studies have dealt with their bioisosteric relationships.

In one of the earlier attempts of structure–activity relationships of amidine derivatives, Fastier reported the possibility of an enhanced binding at the biological target of the guanidinium group compared to the ammonium groups.<sup>16</sup> The author ascribed such a feature to the presence of two electrostatic enforced hydrogen bonding in the *sin* orientations according to the planar geometry of the guanidinium group.<sup>17–19</sup>

Later on, the analysis of a number of parameters such as basicity, electronic properties, and steric hindrance led Cereda

\* Corresponding author phone: ++39 075 585 5120; fax: ++39 075 585 5124; e-mail: rp@unipg.it.



**Figure 1.** Examples of endogenous biologically active substances and drugs containing different amine and amidine groups.

and co-workers to successfully use amidine moieties as bioisosteres of quaternary and tertiary ammonium groups in a series of antimuscarinic compounds.<sup>20</sup>

More recently, by analyzing the pharmacological profile of 3138 compounds tested in at least 50 targets, Azzaoui and co-workers reported that compounds bearing bulky and hydrophobic amines such as quaternary and tertiary amines preferentially show promiscuous binding at the biological targets.<sup>21</sup>

While several studies have been reported concerning the geometries of hydrogen bonding interactions of amine groups at target sites,<sup>22–26</sup> in this work we investigate the distribution of some properties of the binding sites of small molecules bearing amine and amidine groups. The aim is to shed light on whether specific properties of the binding sites may underlie selective interactions of these functional groups. In the last part of the work, we address the issue of bioisosteric relationships between amine and amidine groups in the context of the chemical space of binding sites. Here, we compare the regions occupied by different classes of binding sites that recognize such groups in a reduced

multidimensional space. The aim is to uncover and understand possible relationships between amine and amidine groups that may ultimately prove useful in the selection of the right bioisosteric replacement in the ligand design and optimization, especially in cases where the structure of the biological target is known.

## METHODS

**Definition and Collection of Data Set.** The criteria used to define and collect the data set of binding sites from the RCSB protein database (PDB) were those reported in ref 4. Briefly, structures with a sequence similarity below 90% and containing small molecules bearing the amine or amidine (guanidine) moieties were retrieved from the PDB database. In cases of entry with multimeric chains and/or ligands with more than one amine functional group, all the unique binding sites present in the crystal structure were stored into the data set. The collected protein complexes were submitted to the addition of polar hydrogens optimizing the resulting hydrogen bond network. The MolProbity server was used at this aim.<sup>27</sup>

After the above operations, the following five data sets were collected: (i) 128 binding sites in complex with small molecules containing primary amines; (ii) 58 complexes with small molecules bearing secondary amines; (iii) 160 binding sites in complex with small molecules containing tertiary amines; (iv) 68 complexes with molecules containing the quaternary amine moiety; and (v) 59 binding sites containing ligands with guanidine groups. The average and standard deviation of the resolution factors of the crystal structures was  $2.13 \pm 0.43$ . For each entry of the data set, the binding site of the bioisosteric group was defined as the chemical environment contained in a sphere of radius  $3.5\text{\AA}$  centered on each nitrogen atom of the amine and guanidine group (see the Supporting Information, Figure S1).<sup>4</sup>

**Statistical Analysis.** A number of descriptors (frequency distribution of residue types, formal charge, hydrogen bonding, number of water molecules and structural water molecules, hydrophobicity, flexibility, bulkiness) were calculated for each data set of binding sites. The formal charge of the binding site was calculated according to ref 4 by summing the charged residues (lysine, arginine, aspartate, glutamate) and metal ions present in the chemical environment. Although the protonation state of charged residues may involve histidines and varies according to the pH of the environment and the presence of cofactors, we thought that this approximation furnishes a crude estimation of the propensity of the binding site to make electrostatic interactions. Hydrogen bonding property was calculated counting the number of hydrogen bond acceptors and donors according to the properties of side chains and backbone of residues defining the binding site. In particular, following the convention of atom types in pdb files, hydrogen bond donors were expressed as the sum of “OH” (tyrosine side chain), “OG” (serine and threonine side chains), and “N” atom types (in the side chain and backbone of all residues of binding site), while hydrogen bond acceptors were the sum of oxygen atom types including “OH” and “OG” (in the side chain and backbone of all residues of binding site). Thus, the side chains of serine, threonine, and tyrosine were considered both as hydrogen bond acceptors and donors in order to take into account bifurcated hydrogen bonds. Since histidine is endowed with a tautomeric equilibrium between the proximal and distal nitrogen of the imidazole ring, the nitrogen atoms were considered as hydrogen donor or acceptor according to the presence or absence of the polar hydrogen atom after the hydrogen bond optimization procedure carried out with MolProbity.

Oxygen type of crystal water was considered apart from hydrogen bonding property to reflect the specific ability of this molecule to adapt its hydrogen accepting or donating role upon the binding of basic groups. Although different residue type classification can be adopted, in this study we classified residue type according to the polar character of the side chain. Thus, polar residues were arginine, lysine, histidine, asparagine, glutamine, serine, threonine, cysteine, tyrosine, aspartate, and glutamate. Furthermore, we counted the presence of aromatic residues (tryptophane, tyrosine, and phenylalanine) in the binding site. The presence of structural water molecules in the binding sites was assessed using the program Dowser.<sup>28</sup> Briefly, this program calculates the energy of interaction of a crystal water molecule with the surrounding atoms and defines it as structural water if

the interaction energy is below  $-12$  kcal/mol. The presence of ions in the binding sites was taken into account for the analysis only if the ions were endowed with functional roles in the biology of the macromolecule and discarded when their presence derived from the experimental conditions was used to obtain the crystals.

Other descriptors were calculated on the basis of the profile produced by amino acid scales from the literature. Briefly, amino acid scale are tables of numerical values representing different chemical and physical properties of each type of amino acid.<sup>29</sup> In particular, we used the Kyte-Doolittle scale to evaluate the hydropathicity of the binding site;<sup>30</sup> the average flexibility index of amino acid residues to determine the propensity of the site to conformational changes upon ligand binding;<sup>31</sup> and an amino acid scale to calculate the size and bulkiness distribution of binding sites.<sup>32</sup> Property distributions were numerically represented using the mean, median, standard deviation, and confidence limits. The type of distribution was assessed using the Pearson coefficient of asymmetry. The coefficient is calculated as three times the ratio of the difference between the mean and median over the standard deviation. If the coefficient is  $<-0.5$ , the distribution is negatively skewed; if the coefficient is comprised between  $-0.5$  and  $0.5$ , the data are approximately normally distributed; if the coefficient is  $>0.5$ , the distribution is positively skewed. Since most distributions were not normal distributions, the median was considered as more representative than the mean.<sup>33</sup> The above descriptors were also instrumental to perform a principal component analysis (PCA). All the statistical analyses were carried out using the XLSTAT add-on module of Microsoft-Excel. Images of binding sites at the edges of the “other side” of chemical space were prepared using VMD.<sup>34</sup>

## RESULTS

**Properties of Amine and Amidine Binding Sites.** Five data sets were collected from the protein databank comprising structures in complex with small molecules bearing amine and amidine groups as detailed in the Methods section. In the case of amidines, we selected the guanidine moiety as the representative group of the entire class. For each binding site of the five data sets, the following properties were calculated: frequency distribution of residue type and atom type, formal charge, hydrogen bonding, hydrophobicity, flexibility, and bulkiness. The statistical analysis of these properties is reported in Table 1.

**Residue Types, Metal Ions, and Water Molecules.** The analysis of the presence and number of polar, nonpolar, and aromatic residues unveils some interesting properties of the binding sites that interact with amine and guanidine groups. First, the mean values of the number of polar residues are generally higher than those of nonpolar and aromatic residues, indicating the polar character of the interaction at these binding sites.

Interestingly, this difference is less marked in the case of binding sites recognizing quaternary amines. In view of these results, the majority of the binding sites of quaternary amines, that are almost normally distributed on the number of aromatic residues, uses one aromatic residue (mean =  $0.88 \pm 1.02$ ; median = 1) to anchor the relative ligand by exploiting  $\pi$ -cation interactions rather than electrostatic interactions involving acidic residues.

**Table 1.** Statistical Analysis of Molecular Descriptors

			st.	conf.	Pearson		
descriptors	data set		mean	dev	95%	median	asymmetry
hydrophilicity (KYTE-Doolittle)	primary amine	-3.73	3.92	0.65	-3.50	-0.17	
	secondary amine	-3.54	3.62	0.91	-3.50	-0.03	
	tertiary amine	-2.33	3.56	0.53	-3.25	0.77	
	quaternary amine	-1.63	3.46	0.78	-1.80	0.15	
bulkiness	guanidine	-4.21	5.11	1.26	-3.90	-0.18	
	primary amine	30.02	17.09	2.83	27.02	0.53	
	secondary amine	26.61	13.80	3.46	26.77	-0.03	
	tertiary amine	24.26	14.74	2.18	20.60	0.74	
flexibility	quaternary amine	29.51	16.98	3.82	25.80	0.66	
	guanidine	39.90	21.14	5.22	35.95	0.56	
	primary amine	1.00	0.56	0.09	0.91	0.46	
	secondary amine	0.80	0.35	0.09	0.77	0.25	
polar residues	tertiary amine	0.74	0.41	0.06	0.53	1.53	
	quaternary amine	0.76	0.41	0.09	0.73	0.23	
	guanidine	1.32	0.61	0.15	1.33	-0.06	
	primary amine	1.67	1.11	0.18	1.50	0.46	
nonpolar residues	secondary amine	1.36	0.87	0.22	1.00	1.24	
	tertiary amine	1.11	0.87	0.13	1.00	0.37	
	quaternary amine	1.10	0.85	0.19	1.00	0.36	
	guanidine	1.83	1.04	0.26	2.00	-0.49	
aromatic residues	primary amine	0.45	0.66	0.11	0.00	2.02	
	secondary amine	0.41	0.65	0.16	0.00	1.91	
	tertiary amine	0.41	0.62	0.09	0.00	1.97	
	quaternary amine	0.31	0.55	0.12	0.00	1.67	
water molecules	guanidine	0.76	0.93	0.23	0.00	2.45	
	primary amine	0.37	0.64	0.11	0.00	1.73	
	secondary amine	0.41	0.73	0.18	0.00	1.71	
	tertiary amine	0.39	0.59	0.09	0.00	1.99	
structural water	quaternary amine	0.88	1.02	0.23	1.00	-0.35	
	guanidine	0.46	0.73	0.18	0.00	1.89	
	primary amine	0.57	0.79	0.13	0.00	2.16	
	secondary amine	0.47	0.82	0.21	0.00	1.70	
metal ions	tertiary amine	0.57	0.89	0.13	0.00	1.92	
	quaternary amine	0.49	1.03	0.23	0.00	1.41	
	guanidine	1.15	1.30	0.32	1.00	0.35	
	primary amine	0.05	0.23	0.06	0.00	0.71	
formal charge	secondary amine	0.02	0.13	0.03	0.00	0.39	
	tertiary amine	0.02	0.14	0.02	0.00	0.41	
	quaternary amine	0.00	0.00	—	0.00	—	
	guanidine	0.05	0.29	0.07	0.00	0.53	
HB acceptors	primary amine	0.07	0.26	0.06	0.00	0.83	
	secondary amine	0.00	0.00	—	0.00	—	
	tertiary amine	0.01	0.08	0.01	0.00	0.24	
	quaternary amine	0.00	0.00	—	0.00	—	
HB donors	guanidine	0.00	0.00	—	0.00	—	
	primary amine	-0.53	1.00	0.17	0.00	-1.59	
	secondary amine	-0.31	0.88	0.22	0.00	-1.05	
	tertiary amine	-0.29	0.87	0.13	0.00	-1.02	
	quaternary amine	-0.19	0.53	0.12	0.00	-1.09	
	guanidine	-0.78	0.95	0.23	-1.00	0.70	
	primary amine	2.27	1.60	0.40	2.00	0.51	
	secondary amine	1.57	1.23	0.31	1.00	1.39	
	tertiary amine	1.59	1.19	0.18	1.00	1.50	
	quaternary amine	1.49	1.38	0.31	1.00	1.06	
	guanidine	3.08	1.96	0.48	3.00	0.13	
	primary amine	0.60	0.63	0.15	1.00	-1.92	
	secondary amine	0.69	0.82	0.21	0.50	0.69	
	tertiary amine	0.40	0.64	0.09	0.00	1.89	
	quaternary amine	0.50	0.66	0.15	0.00	2.28	
	guanidine	0.54	0.62	0.15	0.00	2.60	

While the number of metal ions in the binding sites of all the data set is negligible, according to the electrostatic incompatibility of their formal charge with the amine and guanidine groups, one water molecule is almost constantly present around the guanidine moieties ( $\text{mean} = 1.15 \pm 1.30$ ;  $\text{median} = 1$ ). Water molecules in the binding sites may belong to the solvation shell of the ligand and the protein or, alternatively, can have a role in the protein machinery (structural water). While the formers affect the affinity of the target binding site for a ligand by determining the extent of the desolvation energy,<sup>35</sup> structural water is generally involved in ligand recognition by forming hydrogen bonds that bridge the ligand to the binding site. In this latter case,

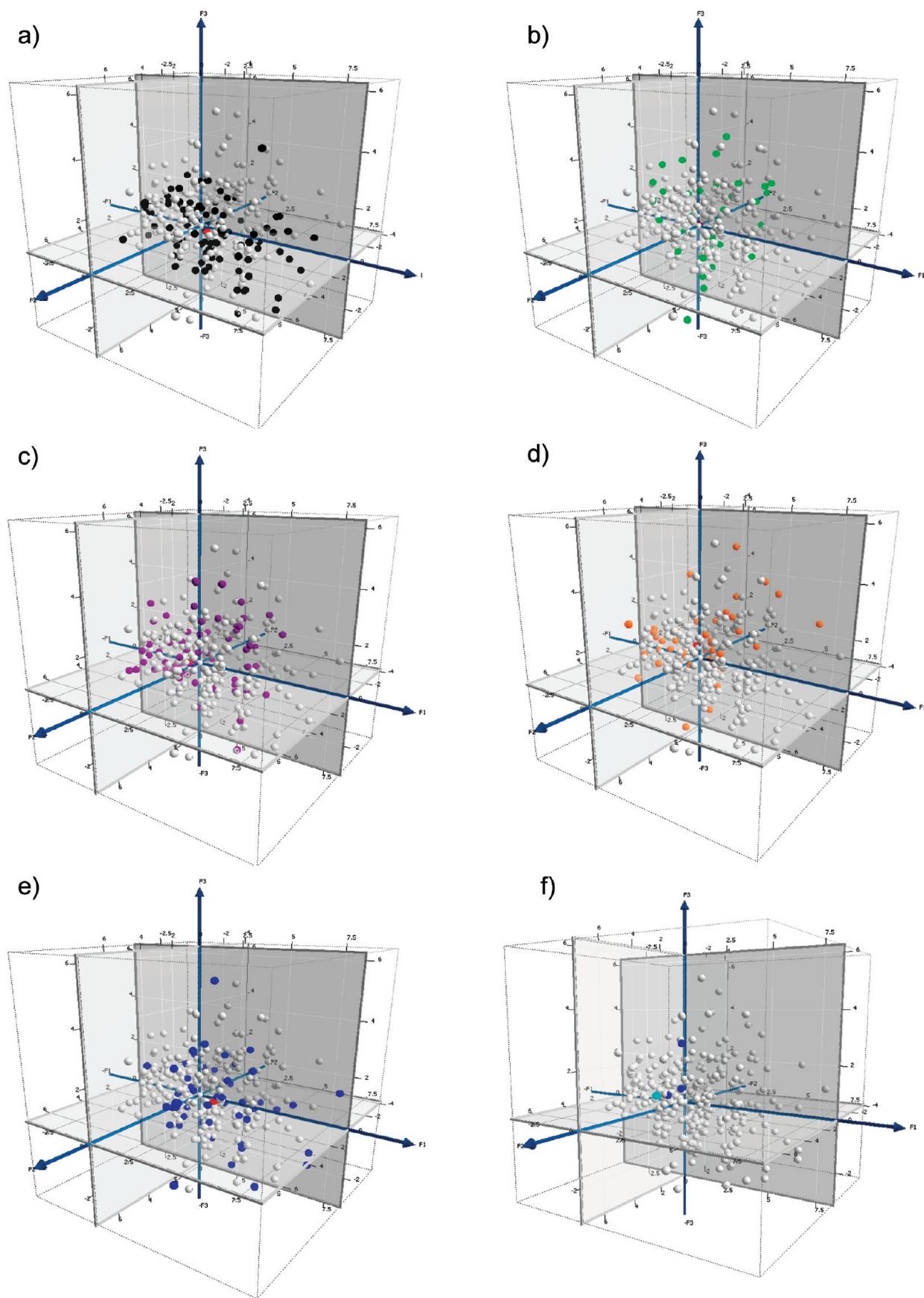
it occupies energetically favorable and conserved positions within the binding sites. Since the presence of structural water in the binding sites of the five data sets is negligible, the nearly constant occurrence of water molecules into the binding sites of guanidine moieties may be ascribed to the strong desolvation energies of these ligands and protein sites, as evidenced by other descriptors such as the hydrophilicity, bulkiness, and the high number of hydrogen bond acceptors.

**Formal Charge.** The mean and the median of the distributions of the formal charge reveal that the binding sites of guanidines are generally endowed with a negative formal charge that is almost normally distributed in the relative data set. This is in contrast to the distributions of primary, secondary, tertiary, and quaternary amines that show on average an uncharged binding site. Nevertheless, the inspection of the Pearson coefficient of asymmetry reveals that the distribution of the formal charge within the primary amine binding sites is skewed at negative values (Pearson coefficient of asymmetry = -1.59). This observation points out that there is a subset of pockets in the set of primary amine binding sites that is endowed with a negative formal charge.

Thus, while the contribution to the formal charge of guanidine binding sites and part of primary amine binding sites is given by the presence of negative charged residues, namely aspartic and glutamic acids, these latter are generally missing in the binding sites of the secondary, tertiary, and quaternary amines. This feature nicely fits the high number of polar hydrogen atoms that are present on the primary amine and guanidine moieties. Accordingly, the five polar hydrogen atoms of the guanidine group and the three polar hydrogen atoms of the primary amine in their relative full protonation state are fully exploited by the relative binding site through the formation of electrostatically enforced hydrogen bonding interactions that involve aspartic acid and glutamic acid residues. Conversely, while the binding sites of quaternary amines adopt  $\pi$ -cation interactions to anchor such groups, those of the remaining secondary and tertiary amine moieties mainly rely on hydrogen bonding and van der Waals interactions.

**Hydrogen Bonding.** Hydrogen bonding is a directional interaction that provides specificity to the binding of functional groups. The hydrogen bonding propensity of the different amine and guanidine binding sites was evaluated comparing the number of hydrogen bond acceptor and donor groups as defined in the Methods section. Guanidine binding sites display a nearly normal distribution of hydrogen bond acceptor groups according to the low Pearson coefficient of asymmetry. They have the greatest median and average number of hydrogen bond acceptors ( $\text{mean} = 3.08 \pm 1.96$ ;  $\text{median} = 3$ ) followed by the primary amines ( $\text{mean} = 2.27 \pm 1.60$ ;  $\text{median} = 2$ ). Secondary, tertiary, and quaternary amine binding sites show on median only one hydrogen bond acceptor, though their distributions are positively skewed.

The poor median and average values of hydrogen bond donor groups in all the data set indicate the shortage to donate hydrogen bonding on part of these binding sites. It is noteworthy, however, that a portion of primary amine binding sites display the presence of hydrogen donor groups ( $\text{mean} = 0.60 \pm 0.63$ ;  $\text{median} = 1$ ; Pearson coefficient of asymmetry = -1.92) when compared to the remaining amine and guanidine binding sites. This observation may tentatively



**Figure 2.** Plots of the first three principal components for different amine and guanidine binding sites and the average of the distributions (red dots): a) primary amine groups (black dots); b) secondary amine groups (green dots); c) tertiary amine groups (magenta dots); d) quaternary amine groups (orange dots); and e) guanidine groups (blue dots). f) Guanidine binding sites are identified in the mitochondrial aldehyde dehydrogenase. The cyan dot shows the binding site analyzed in detail.

be explained with the presence of a subset of binding sites in the data set of primary amines that host ligands with an uncharged nitrogen atom. Hence, the presence of an hydrogen bond donor in these binding sites allows for the exploiting of the free lone pair of the uncharged primary amine to achieve a finer binding of such a functional group.

**Hydrophobicity, Flexibility, and Bulkiness.** Amino acid scales of chemical and physical properties of residues are frequently used for the evaluation of the relative profiles of proteins. In this study, a score of hydrophobicity (Kyte-Doolittle scale), flexibility, and bulkiness was assigned to each binding site as the sum of the scale values for the residues that define the site.

As far as it concerns the hydrophobicity descriptor, the guanidine, primary, and secondary amine binding sites display on average a strong polarity. In agreement with previous data, tertiary and quaternary amine binding sites are less polar with the mean and median values of the latter (mean =  $-1.63 \pm 3.46$ ; median =  $-1.80$ ) being significantly lower than those of primary amine binding sites (mean =  $-3.73 \pm 3.92$ ; median =  $-3.50$ ).

The molecular recognition of ligands on the part of binding sites is highly influenced by the conformational flexibility of the side chains of residues. The larger flexibility shown by the binding sites of the guanidine data set (mean =  $1.32 \pm 0.61$ ; median =  $1.33$ ) suggests that these protein pockets may take advantage of their plasticity to achieve nice complementarities of shape and interaction with the guanidine moiety, engaging the polar hydrogen atoms of this latter in large hydrogen bond networks. On the other side, the poor flexibility of secondary, tertiary, and quaternary amine binding sites accounts for a less conformational demanding molecular recognition of these groups that ensues from the dearth of directional interactions, namely hydrogen bonding, in these pockets. The bulkiness descriptor encodes for the size of the binding site. It ensues that guanidine binding sites are the larger pockets (mean =  $39.90 \pm 21.14$ ; median =  $35.95$ ), followed by primary (mean =  $30.02 \pm 17.09$ ; median =  $27.02$ ), quaternary (mean =  $29.51 \pm 16.98$ ; median =  $25.80$ ), and secondary (mean =  $26.61 \pm 13.80$ ; median =  $26.77$ ) amine binding sites. Tertiary amine binding sites show the smaller recognition pocket (mean =  $24.26 \pm 14.74$ ; median =  $20.60$ ).

**Principal Component Analysis (PCA).** In this part of the study, we have applied a Principal Component Analysis (PCA) on the autoscaled values of the descriptors in order to qualitatively map the amine and guanidine binding sites in the chemical space as represented by the first three principal components (Figure 2).

These components explain about the 61% of the variance of the original data set. The inspection of the loadings (Table 2) of the original variables into the first three components reveals that the first component (F1) encodes the effects of polarity, flexibility, and size (bulkiness) of the binding site in determining the binding of amine and guanidine groups. On the other side, the second component (F2) explains the propensity of the binding site to make hydrogen bonding interactions and the desolvation energies of the ligand and protein site. This effect, in particular, is ascribed to the positive high loadings of the variables accounting respectively for the number of hydrogen bond accepting groups and water molecules in the binding site. Likewise, the third

**Table 2.** Loadings of the Molecular Descriptors onto the Three Principal Components

descriptors	F1	F2	F3
hydrophobicity (Kyte-Doolittle)	-0.645	0.394	<b>0.559</b>
bulkiness	<b>0.810</b>	-0.056	0.465
flexibility	<b>0.930</b>	0.028	0.172
polar residues	<b>0.890</b>	-0.302	-0.226
nonpolar residues	0.097	0.520	<b>0.634</b>
aromatic residues	0.305	-0.258	<b>0.556</b>
water molecules	0.171	<b>0.686</b>	-0.136
structural water	0.154	0.337	-0.049
metal ions	-0.071	-0.195	-0.001
formal charge	-0.513	-0.328	0.425
HB acceptors	0.694	<b>0.592</b>	-0.095
HB donors	0.376	-0.510	0.288

**Table 3.** Averages and Standard Deviations of the Distribution of Binding Sites in the Three Principal Components

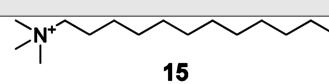
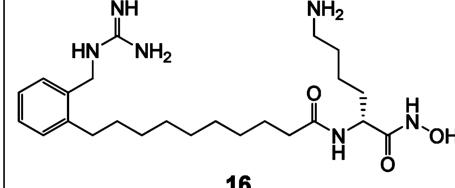
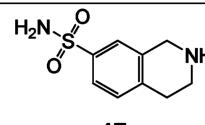
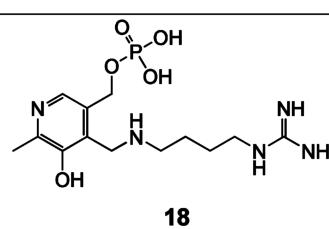
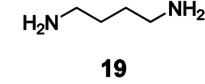
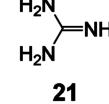
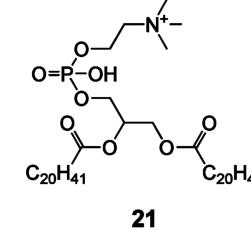
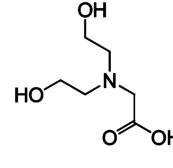
constellation	F1	F2	F3
primary amines	$0.42 \pm 2.12$	$0.00 \pm 1.40$	$-0.20 \pm 1.15$
secondary amines	$-0.19 \pm 1.54$	$-0.35 \pm 1.28$	$-0.05 \pm 1.47$
tertiary amines	$-0.61 \pm 1.58$	$0.00 \pm 1.15$	$-0.06 \pm 1.19$
quaternary amines	$-0.46 \pm 1.62$	$-0.31 \pm 1.29$	$0.45 \pm 1.25$
guanidine moieties	$1.46 \pm 2.19$	$0.70 \pm 1.88$	$0.11 \pm 1.45$

component (F3) accounts for the van der Waals and  $\pi$ -cation interactions that affect the binding of amine and guanidine groups. In particular, by inspecting the sign of the coefficients of the number of nonpolar and aromatic residues, it can be pointed out that these interactions occur at binding sites located at high positive values of the third component. Figure 2 shows the distributions of the binding sites for each data set plotted onto the first three components. Projections of these plots on the F1–F2 and F1–F3 planes can be found in the Supporting Information, Figures S2–S6. Table 3 shows the averages and standard deviations of each data set in the three components. While the average values localize the centers of the data sets in the chemical space, the standard deviations roughly quantify the spreading of each data set.

The inspections of Table 3 and Figure 2 reveal that primary amine binding sites appear more clustered at low values of the second and third components while spreading from negative to very positive values of the first component. Conversely, guanidine binding sites are inclined to spread at positive values of the first and second components with a low mean value and spreading on the third component. Protein sites binding quaternary amines tend to cluster at positive values of the third component and low negative values of the first and second components. In contrast, the centers of secondary and tertiary amine data set are localized at very low values of the third component and negative values of the first and second components.

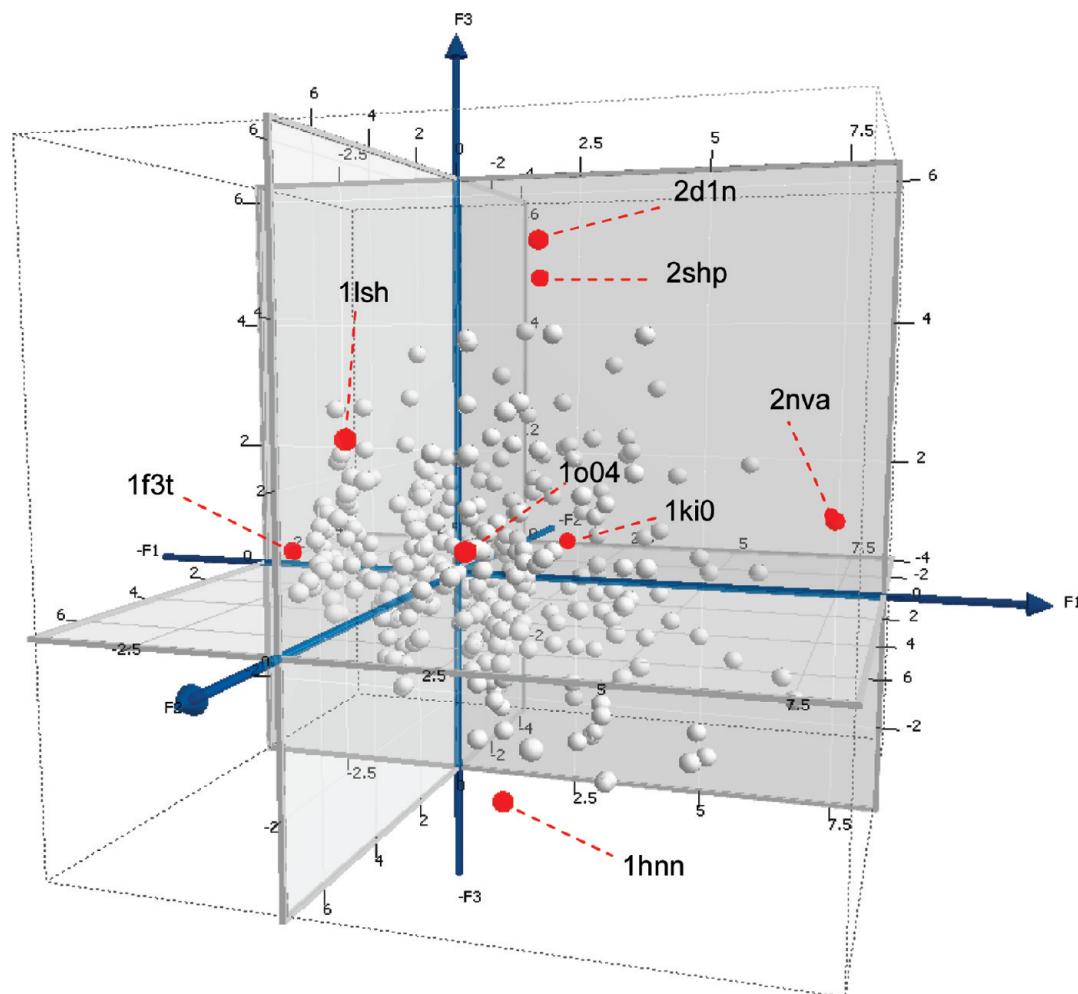
**Binding Sites at the Edges of the Chemical Space.** In order to have a close view at the interactions occurring at binding sites located at the edges of the chemical space, we selected seven binding sites laying at the border of each of the three components (Table 4, Figure 3). Selected binding sites lying at high positive values of the third component comprised tyrosine phosphatase (SHP-2) in complex with dodecane-trimethylamine (pdb code: 2shp)<sup>36</sup> and collagenase-3 (MMP-13) in complex with a hydroxamic acid inhibitor (pdb code: 2d1n).<sup>37</sup>

**Table 4.** Binding Sites Selected at the Edges of the Chemical Space of Target Sites

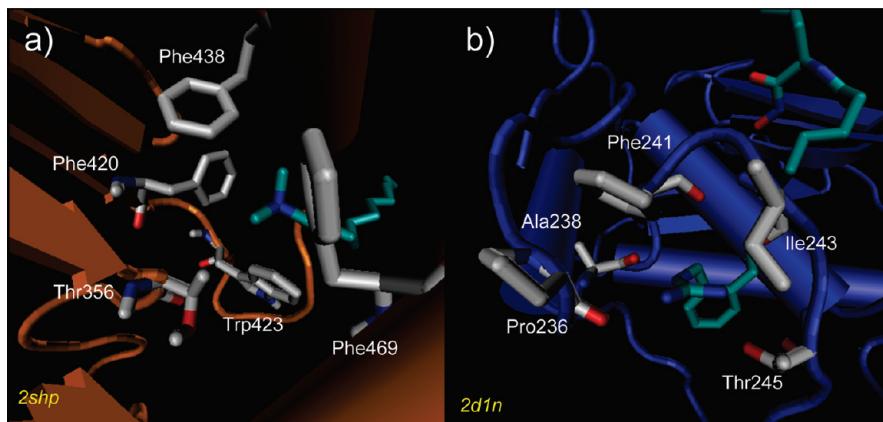
PDB Code	F1	F2	F3	Ligand
2shp	1.331	-1.221	4.768	 <b>15</b>
2d1n	2.253	1.820	5.179	 <b>16</b>
1hnn	2.237	3.657	-2.617	 <b>17</b>
2nva	7.357	-1.111	0.967	 <b>18</b>
1f3t	-3.482	0.424	0.285	 <b>19</b>
1o04	1.983	4.801	1.035	 <b>21</b>
1ish	-0.030	4.944	2.461	 <b>21</b>
1ki0	1.294	-3.475	0.160	 <b>22</b>

In this region of the space it is expected to have sites with high specificity of recognition for ligands bearing quaternary amines that exploit mostly  $\pi$ -cation interactions to bind the target proteins. Accordingly, the structure of SHP-2 is crystallized with a quaternary amine derivative, namely dodecane-trimethylamine (**15**), that interacts with the binding site of the enzyme using  $\pi$ -cation interactions with an aromatic cage composed of Phe420, Trp423, Phe438, and Phe469 (Figure 4a). Conversely to what was expected, MMP-13 binds a hydroxamic acid inhibitor (SM-25453, **16**) bearing a guanidine moiety. SM-25453 is a potent MMP-3 and MMP-13 inhibitor whose structure is composed of three

parts: (i) a hydroxamic acid, (ii) D-lysine, and (iii) a long aliphatic chain with a terminal hydrophilic guanidinomethyl moiety (position P1'). This latter group binds to the hydrophobic S1' pocket of the enzyme composed of the aromatic residue Phe241, the hydrophobic residues Pro236, Ala238, and Ile243, and the polar residue Thr245. The crystal structure of MMP-13 reveals four hydrogen bonding interactions that anchor the guanidine group to the carbonyl moieties of the hydrophobic residues and to the hydroxyl group of Thr245 (Figure 4b). Despite the potency of compound **16**, it should be mentioned that several potent and selective inhibitors of MMP-13 have been reported in literature.<sup>38–45</sup>



**Figure 3.** Selected binding sites laying at the edges of the three components.

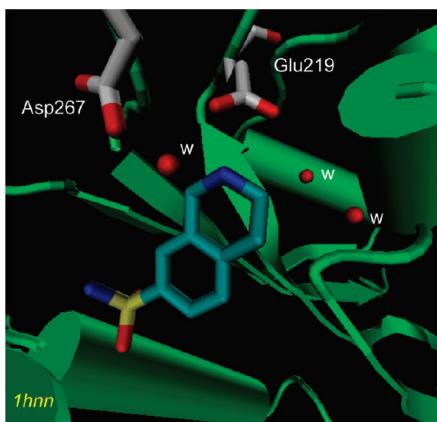


**Figure 4.** High PC3 score values. a) Binding site of tyrosine phosphatase (SHP-2) in complex with dodecane-trimethylamine (pdb code: 2shp). b) Binding site of collagenase-3 (MMP-13) in complex with a hydroxamic acid inhibitor (pdb code: 2d1n).

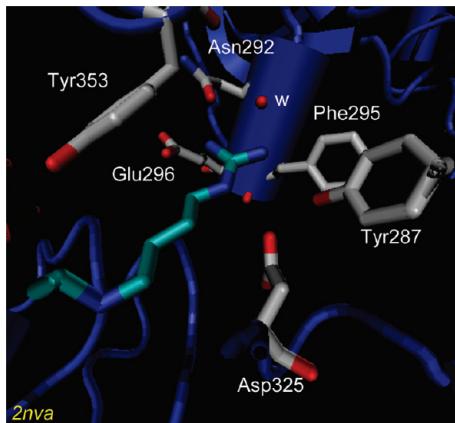
These derivatives generally bear aromatic side chains at position P1' that tightly bind the S1' pocket of the enzyme through hydrophobic and  $\pi$ - $\pi$  interactions. In view of that, the interaction of SM-25453 (**16**) at the binding site of MMP-13 may represent an outlier not only in the property space of binding sites but also in the chemical space covered by the inhibitors of MMP-13. Unfortunately, no MMP-13 inhibitors have been hitherto reported bearing a quaternary amine chain at position P1'. Indeed, our study gives rise to the interesting hypothesis that this type of replacement might

produce a tight binding of the ligand to the enzyme by mostly exploiting  $\pi$ -cation interaction with Phe241 in the S1' pocket.

Among the binding sites falling at very low values of the third component, we selected the crystal structure of the adrenaline-synthesizing enzyme PNMT (pdb: 1hnn) and a submicromolar inhibitor bearing a secondary amine group, namely 1,2,3,4-tetrahydroisoquinoline-7-sulfonic acid amide (**17**).<sup>46</sup> According to the loadings of the descriptors mapped on the third component that point out a significative weight of the electrostatic interactions, a different scenario occurs



**Figure 5.** Low PC3 score values. Binding site of the adrenaline-synthesizing enzyme PNMT (pdb: 1hnn) in complex with 1,2,3,4-tetraidroisoquinoline-7-sulfonic acid amide (**17**).

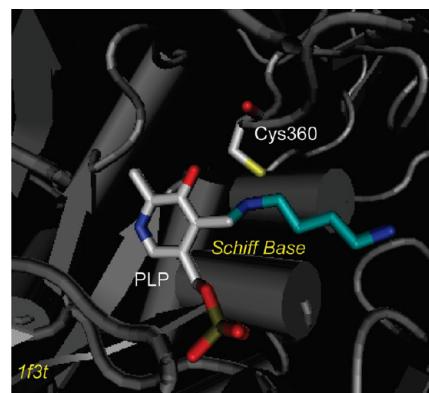


**Figure 6.** High PC1 score values. Binding site of arginine decarboxylase (pdb code: 2nva) bound to agmatine (**18**).

in the binding site of PNMT compared to that of the enzymes SHP-2 and MMP-13. Here, a protonated secondary amine group is anchored to the binding site of PNMT through electrostatic interactions with Glu219 and Asp267 and hydrogen bonds that also involve bridge interactions with three water molecules (Figure 5a).

At the edges of the first component, we selected the binding sites of arginine decarboxylase (pdb code: 2nva) bound to agmatine (**18**)<sup>47</sup> and ornithine decarboxylase (pdb code: 1f3t) bound to putrescine (**19**)<sup>48</sup>, respectively, located at very positive and negative values of the component. In the case of arginine decarboxylase lying at the positive end of the component, the binding site of the guanidine group is large and very polar for the presence of several polar and negatively charged residues (Figure 5b). In contrast to the previous selected binding sites at the edges of the third component, this guanidine binding site shows strong electrostatic interactions (occurring with Glu296 and Asp325) as well as  $\pi$ -cation interactions (involving Tyr287, Tyr353, and Phe295) and hydrogen bonds (water molecule and Asn292).

At the other end of the first component, the binding site of ornithine decarboxylase is a much more simple pocket consisting of Cys360 and the pyridoxal 5'-phosphate (PLP) covalently bound to one of the primary amine groups of putrescine (**19**) through a Schiff base (Figure 6a). Despite its simplicity, this primary amine binding site is in possession of all the requisites to achieve a subtle specificity of



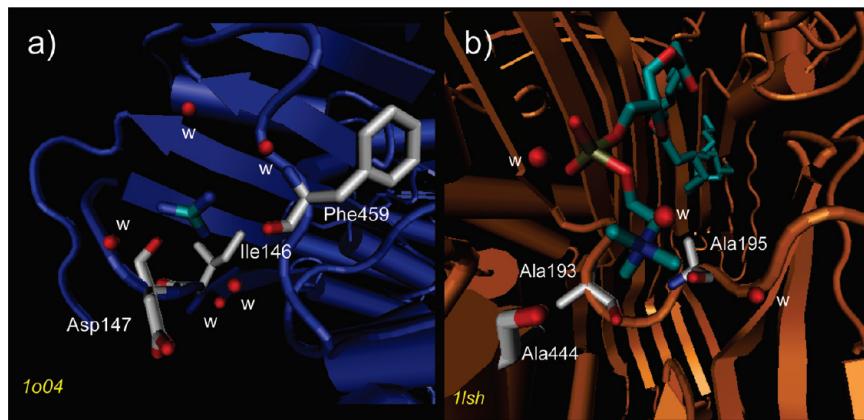
**Figure 7.** Low PC1 score values. Binding site of ornithine decarboxylase (pdb code: 1f3t) bound to putrescine (**19**).

recognition. Indeed, while the covalent interaction provides a very strong anchoring of the ligand to the enzyme, Cys360 controls the protonation step of the  $\text{C}\alpha$  of ornithine that, as evidenced by biochemical studies, catalyzes the formation of putrescine (**19**).<sup>48</sup>

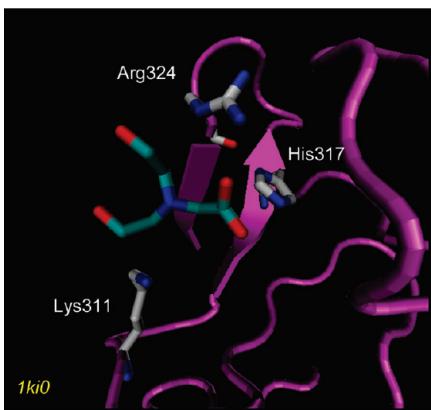
A guanidine and a quaternary amine binding site were selected at the positive end of the second component: a mutant of the mitochondrial aldehyde dehydrogenase (pdb code: 1o04)<sup>49</sup> and lipovitellin (pdb code: 1lsh) bound to diundecyl phosphatidyl choline (**21**).<sup>50</sup> It should be mentioned that in the first case the guanidine moiety is not present as a functional group in the ligand, but it is part of the crystallization buffer of the protein.<sup>49</sup> In view of that, nine guanidine binding sites were identified in the mitochondrial aldehyde dehydrogenase (Figure 2f). Although it is debatable regarding the inclusion of such binding sites in this study, our aim was to chart the widest number of different kinds of amine and guanidine sites regardless of the binding affinities shown by these moieties. Thus, the selected binding site displays a guanidine ion (**20**) that is surrounded by seven hydrogen bond acceptors that are provided by five water molecules and two carbonyl groups of the backbone of Asp147 and Phe459 (Figure 6b). The presence of a large number of water molecules is ascribed to the fact that the pocket is fully exposed to the solvent. A further observation is that, even if there are a negatively charged residue (Asp157) and an aromatic residue (Phe459), no electrostatic enforced hydrogen bonding and  $\pi$ -cation interactions are clearly evidenced.

A similar situation occurs in the binding site of lipovitellin bound to diundecyl phosphatidyl choline (**21**) (see Figure 8). Here, the quaternary amine moiety of the choline fragment is trapped in a partially solvent exposed pocket by six hydrogen bond acceptor groups supplied by three water molecules and the carbonyl groups of Ala193, Ala195, and Ala444 (Figure 7a). Opposed to the former case, this pocket may only marginally contribute to the binding affinity of the ligand since the quaternary amine group will be entailed in charge dipole interactions with the electronegative oxygen atoms and van der Waals interactions with the methyl groups of the alanine residues.

At the bottom end of the second component we have selected the binding site of angiostatin bound to 2-(bis-(2-hydroxyethyl)amino)acetic acid (**22**, bicine) (see Figure 9).<sup>51</sup> In the chemical structure of bicine, the tertiary amine nitrogen acts more as a scaffold atom than a functional



**Figure 8.** High PC2 score values. a) One of the nine guanidine ion binding sites of the mitochondrial aldehyde dehydrogenase (pdb code: 1o04). b) Binding site of lipovitellin (pdb code: 1lsh) bound to diundecyl phosphatidyl choline (**21**).



**Figure 9.** Low PC2 score values. Binding site of angiotensin (pdb code: 1ki0) bound to 2-(bis(2-hydroxyethyl)amino)acetic acid (**22**, bicine).

group. Indeed, it ensures a roughly planar pose of the hydroxyl and carboxyl functional groups. It ensues that this pocket is not a “real” amine binding site, as evidenced by the presence of the positively charged residues Lys311, Arg324, and His317 that interact with the hydroxyl and carboxyl groups of bicine (Figure 7b). Nonetheless, the presence of such a site in the chemical space unveils the double functional aspect of substituted amines: (i) they may behave as true interacting groups of ligands contributing to the free energy of binding or (ii) they may act as scaffold atoms providing the correct binding shape of the ligand to the target site.

## DISCUSSION

Very recently, we introduced the concept of the “other side” of chemical space as composed of constellations of binding sites found in proteins. Then, we navigated this space to study the binding mode and bioisosteric relationships of three acidic functional groups, namely carboxylic, sulfonate, and phosphonate groups.<sup>4</sup>

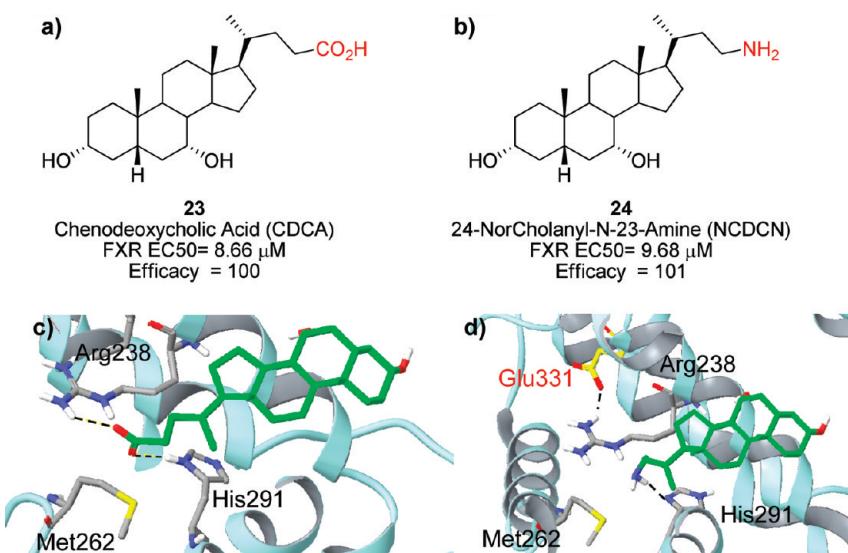
In this work we have extended the previous analysis to binding sites that interact with different amine and guanidine moieties. First, we have investigated the distribution of some properties of these binding sites using a number of descriptors as reported in Table 1.

Our results reveal the existence of specific motifs in the binding sites that may constitute the molecular basis of the specificity of the recognition of amine and guanidine groups.

In particular, highly polar and large binding sites endowed with broad conformational flexibility have the propensity to bind guanidine and primary amine moieties. Noteworthy, these sites have respectively the triple and double number of hydrogen bond acceptor groups on median values than the secondary, tertiary, and quaternary amine binding sites. A further observation is that guanidine and part of primary amine binding sites have a net negative formal charge compared to the majority of secondary, tertiary, and quaternary amine binding sites. This feature nicely fits with the high number of hydrogen bond acceptors of these pockets and underlines their strong propensity to make salt bridges with the positively charged guanidine and primary amine groups. On the other side, many polyalkylated amine moieties display a nearly neutral binding site that is apparently at odds with their positive charge at physiological pH. This observation can be rationally explained with the presence of aromatic residues in the pocket that reinforce the binding of N-alkyl substituted amines to the protein with  $\pi$ -cation interactions. Hence, while primary amines and guanidines exploit mostly directional electrostatic enforced hydrogen bond interactions to anchor the biological target, part of the secondary, tertiary, and quaternary amines rely on  $\pi$ -cation and van der Waals interactions to tightly bind the protein. Since these latter interactions display only a marginal directionality, it ensues that small molecules bearing N-alkyl substituted groups can potentially have promiscuous binding at several biological targets. This fact is indeed in agreement with the results of Azzaoui and co-workers reporting the preferential promiscuity of small molecules bearing bulky and hydrophobic amines at diverse biological targets.<sup>21</sup>

In the last part of the work, we have addressed the issue of bioisosteric relationships between amine and guanidine groups in the context of the chemical space. The construction of the chemical space of binding sites, in particular, has been instrumental to provide a qualitative visualization of the diversity of amine and guanidine group recognition on the part of biological targets. Thus, by using a principal component analysis, we have generated a three-dimensional space (F1, F2, and F3) on the basis of the autoscaled variables obtained from the descriptors of the binding sites.

While the first dimension (F1, Table 2) encodes diverse properties of the recognition sites comprising polarity, size, and conformational flexibility, the second dimension (F2, Table 2) discriminates the binding sites according to their



**Figure 10.** Chemical structures and activities of CDCA (**23**, a) and its primary amine derivative (**24**, b) at FXR. Binding mode of CDCA (c) and compound **24** (d) in the binding site of FXR (see ref 52).

desolvation energy and the tendency to make hydrogen bonding interactions. The third dimension (F3, Table 2) weights the van der Waals and  $\pi$ -cation interactions occurring in the binding sites of amine and guanidine groups.

In the principal component graphs (Figure 2) we can identify a large central region where binding sites of different amine moieties overlap and only a few regions occupied by discrete portions of binding sites that recognize specific amine groups. Accordingly, while replacements between different amine and guanidine moieties will be likely bioisosteric in ligands interacting at binding sites that occupy the central region of the chemical space, these will be likely unfruitful as much as the ligand binding sites fall at highly positive values of the three components.

In particular, the analysis of Table 3 suggests that small molecules bearing guanidine groups may selectively bind sites that lie almost exclusively at high positive values of the first and second components. Likewise, ligands bearing quaternary amines may specifically bind pockets that are located at high positive values of the third component. Although the above bioisosteric replacements between amine and guanidine groups may be considered trivial, the analysis of specific motifs that are present in a subset of binding sites of primary amines reveals patterns of interactions that may allow the explanation of unexpected bioisosteric relationships. Thus, according to Table 1, a portion of binding sites of primary amines are endowed with the presence of one hydrogen bond donor group (mean =  $0.60 \pm 0.63$ , median = 1) and a neutral formal charge (mean =  $-0.53 \pm 1$ ; median = 0). This observation implies that such binding sites may also interact with ligands bearing acidic groups. This is the case of the binding site of the bile acid nuclear receptor FXR. The endogenous ligand for this receptor is CDCA (**23**), and its carboxylic group interacts with Arg238 through an apparently electrostatic enforced hydrogen bond. This site, however, has a neutral formal charge with the side chain of Arg238 being salt bridged with Glu331 (Figure 10). Remarkably, this site is also endowed with the presence of both hydrogen bond acceptor (His291) and donor (Arg238) groups. Supporting the observation done in this study, we have shown that the carboxylic group of CDCA (**23**) can be

replaced by a primary amine (**24**) without losing the activity.<sup>52</sup> This case study proves that atypical bioisosteric relationships may exist even between acidic and basic groups depending upon the interaction pattern present in the target site.

## CONCLUSIONS

Although we are aware that the major caveat of constructing a chemical space of binding sites is the lack of accurate structures and annotation of activity data of small molecules and molecular fragments, charting this chemical space provides insights into the principles that rule the interactions at the target sites and the bioisosteric relationships of functional groups.

The present study represents the extension of our previous work on binding sites recognizing acidic fragments and includes sites that interact with small molecules bearing different amine and guanidine groups. The analysis of the chemical space covered by these later points out that there is a large central region where these pockets are likely to host different bioisosteric replacements among different amine and guanidine groups. Nevertheless, while binding sites lying at extreme positive values of the first and second components may specifically interact with guanidine moieties, protein pockets falling at high positive values of the third component may better interact with quaternary amine groups.

The inspection of selected binding sites at the edges of the chemical space unveils different type of interactions that protein pockets exploit to anchor amine and guanidine groups. Furthermore, this analysis pinpoints the double role of substituted amine moieties in drug discovery, where they can be used both as a functional group or a scaffold moiety.

Finally, these results may prove useful in improving our knowledge on the binding mode of functional groups and aid the selection of bioisosteric replacements in the processes of ligand design and optimization.

## ACKNOWLEDGMENT

This work was supported by the Italian Ministry of Research (MIUR-PRIN 2006).

**Supporting Information Available:** Cartoon depicting the definition of the binding site for different amine and guanidine groups (Figure S1), projections of the binding sites on the F1–F2 (a) and F1–F3 (b) planes (Figures S2–S6), and the file “Amine Binding Sites.xls” of the data set used for this study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (1) Booth, B.; Zemmel, R. Prospects for productivity. *Nat. Rev. Drug Discovery* **2004**, *3*, 451–456.
- (2) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **2003**, *22*, 151–185.
- (3) Ertl, P. In silico identification of bioisosteric functional groups. *Curr. Opin. Drug Discovery Devel.* **2007**, *10*, 281–288.
- (4) Macchiarulo, A.; Pellicciari, R. Exploring the other side of biologically relevant chemical space: insights into carboxylic, sulfonic and phosphonic acid bioisosteric relationships. *J. Mol. Graphics Modell.* **2007**, *26*, 728–739.
- (5) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (6) Laskowski, R. A.; Thornton, J. M.; Humblet, C.; Singh, J. X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins. *J. Mol. Biol.* **1996**, *259*, 175–201.
- (7) von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418–423.
- (8) Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. Protein clefts in molecular recognition and function. *Protein Sci.* **1996**, *5*, 2438–2452.
- (9) Nobeli, I.; Laskowski, R. A.; Valdar, W. S.; Thornton, J. M. On the molecular discrimination between adenine and guanine by proteins. *Nucleic Acids Res.* **2001**, *29*, 4294–4309.
- (10) Laskowski, R. A.; Watson, J. D.; Thornton, J. M. Protein function prediction using local 3D templates. *J. Mol. Biol.* **2005**, *351*, 614–626.
- (11) Glaser, F.; Morris, R. J.; Najmanovich, R. J.; Laskowski, R. A.; Thornton, J. M. A method for localizing ligand binding pockets in protein structures. *Proteins* **2006**, *62*, 479–488.
- (12) Torrance, J. W.; Holliday, G. L.; Mitchell, J. B.; Thornton, J. M. the geometry of interactions between catalytic residues and their substrates. *J. Mol. Biol.* **2007**, *369*, 1140–1152.
- (13) Kahraman, A.; Morris, R. J.; Laskowski, R. A.; Thornton, J. M. Shape variation in protein binding pockets and their ligands. *J. Mol. Biol.* **2007**, *368*, 283–301.
- (14) Najmanovich, R. J.; Allali-Hassani, A.; Morris, R. J.; Dombrovsky, L.; Pan, P. W.; Vedadi, M.; Plotnikov, A. N.; Edwards, A.; Arrowsmith, C.; Thornton, J. M. Analysis of binding site similarity, small-molecule similarity and experimental binding profiles in the human cytosolic sulfotransferase family. *Bioinformatics* **2007**, *23*, e104–109.
- (15) Bashton, M.; Nobeli, I.; Thornton, J. M. Cognate ligand domain mapping for enzymes. *J. Mol. Biol.* **2006**, *364*, 836–852.
- (16) Fastier, F. N. Structure-activity relationships of amidine derivatives. *Pharmacol. Rev.* **1962**, *14*, 37–90.
- (17) Durant, G. J.; Parsons, M. E.; Black, J. W. Potential histamine H<sub>2</sub>-receptor antagonists. 2. N-alpha-Guanylhistamine. *J. Med. Chem.* **1975**, *18*, 830–833.
- (18) Ganellin, R. 1980 Award in Medicinal Chemistry: Medicinal chemistry and dynamic structure-activity analysis in the discovery of drugs acting at histamine H<sub>2</sub> receptors. *J. Med. Chem.* **1981**, *24*, 913–920.
- (19) Donetti, A.; Cereda, E.; Bellora, E.; Gallazzi, A.; Bazzano, C.; Vanoni, P.; Del Soldato, P.; Micheletti, R.; Pagani, F.; Giachetti, A. (Imidazolylphenyl)formamidines. A structurally novel class of potent histamine H<sub>2</sub> receptor antagonists. *J. Med. Chem.* **1984**, *27*, 380–386.
- (20) Cereda, E.; Ezhaya, A.; Gil Quintero, M.; Bellora, E.; Dubini, E.; Micheletti, R.; Schiavone, A.; Brambilla, A.; Schiavi, G. B.; Donetti, A. Synthesis and biological evaluation of new antimuscimilic compounds with amidine basic centers. A useful bioisosteric replacement of classical cationic heads. *J. Med. Chem.* **1990**, *33*, 2108–2113.
- (21) Azzaoui, K.; Hamon, J.; Faller, B.; Whitebread, S.; Jacoby, E.; Bender, A.; Jenkins, J. L.; Urban, L. Modeling Promiscuity Based on in vitro Safety Pharmacology Profiling Data. *ChemMedChem* **2007**, *2*, 874–880.
- (22) Fabian, L.; Chisholm, J. A.; Galek, P. T.; Motherwell, W. D.; Feeder, N. Hydrogen-bond motifs in the crystals of hydrophobic amino acids. *Acta Crystallogr., Sect. B: Struct. Sci.* **2008**, *64*, 504–514.
- (23) Allen, F. H.; Bird, C. M.; Rowland, R. S.; Harris, S. E.; Schwalbe, C. H. Correlation of hydrogen-bond acceptor properties of nitrogen with the geometry of the N(sp<sub>2</sub>)-N(sp<sub>3</sub>) transition: reaction pathway for the protonation of nitrogen. *Acta Crystallogr., Sect. B: Struct. Sci.* **1995**, *51*, 1068–1081.
- (24) Taylor, R.; Kennard, O.; Versichel, W. Geometry of the imino-carbonyl (N-H···O=C) hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **1983**, *105*, 5761–5766.
- (25) Taylor, R.; Kennard, O.; Versichel, W. Geometry of the nitrogen-hydrogen···oxygen-carbon (N-H···O=C) hydrogen bond. 2. Three-center (bifurcated) and four-center (trifurcated) bonds. *J. Am. Chem. Soc.* **1984**, *106*, 244–248.
- (26) Taylor, R.; Kennard, O.; Versichel, W. The geometry of the N-H···O=C hydrogen bond. 3. Hydrogen-bond distances and angles. *Acta Crystallogr., Sect. B: Struct. Sci.* **1984**, *40*, 280–288.
- (27) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375–383.
- (28) Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins* **1996**, *24*, 433–438.
- (29) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The proteomics protocols handbook*; Walker, J. M., Ed.; Humana Press: 2005; pp 571–607.
- (30) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (31) Bhaskaran, R.; PonnuSwamy, P. K. Positional flexibilities of amino acid residues in globular proteins. *Int. J. Pept. Protein Res.* **1988**, *32*, 242–255.
- (32) Zimmerman, J. M.; Eliezer, N.; Simha, R. The characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.* **1968**, *21*, 170–201.
- (33) Sokal, R. F.; Rohlf, F. J. In *Biometry*, 2nd ed.; Friedman, H. W., Ed.; New York, 1981; pp 43–46.
- (34) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (33–38), 27–38.
- (35) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75.
- (36) Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M. J.; Shoelson, S. E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **1998**, *92*, 441–450.
- (37) Kohno, T.; Hochigai, H.; Yamashita, E.; Tsukihara, T.; Kanaoka, M. Crystal structures of the catalytic domain of human stromelysin-1 (MMP-3) and collagenase-3 (MMP-13) with a hydroxamic acid inhibitor SM-25453. *Biochem. Biophys. Res. Commun.* **2006**, *344*, 315–322.
- (38) Levin, J. I. The design and synthesis of aryl hydroxamic acid inhibitors of MMPs and TACE. *Curr. Top. Med. Chem.* **2004**, *4*, 1289–1310.
- (39) Reiter, L. A.; Freeman-Cook, K. D.; Jones, C. S.; Martinelli, G. J.; Antipas, A. S.; Berliner, M. A.; Datta, K.; Downs, J. T.; Eskra, J. D.; Forman, M. D.; Greer, E. M.; Guzman, R.; Hardink, J. R.; Janat, F.; Keene, N. F.; Laird, E. R.; Liras, J. L.; Lopresti-Morrow, L. L.; Mitchell, P. G.; Pandit, J.; Robertson, D.; Sperger, D.; Vaughn-Bowser, M. L.; Waller, D. M.; Yocom, S. A. Potent, selective pyrimidinetrione-based inhibitors of MMP-13. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5822–5826.
- (40) Freeman-Cook, K. D.; Reiter, L. A.; Noe, M. C.; Antipas, A. S.; Danley, D. E.; Datta, K.; Downs, J. T.; Eisenbeis, S.; Eskra, J. D.; Garmene, D. J.; Greer, E. M.; Griffiths, R. J.; Guzman, R.; Hardink, J. R.; Janat, F.; Jones, C. S.; Martinelli, G. J.; Mitchell, P. G.; Laird, E. R.; Liras, J. L.; Lopresti-Morrow, L. L.; Pandit, J.; Reilly, U. D.; Robertson, D.; Vaughn-Bowser, M. L.; Wolf-Gouveia, L. A.; Yocom, S. A. Potent, selective spiropyrrolidone pyrimidinetrione inhibitors of MMP-13. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6529–6534.
- (41) Hu, Y.; Xiang, J. S.; DiGrandi, M. J.; Du, X.; Ipek, M.; Laakso, L. M.; Li, J.; Li, W.; Rush, T. S.; Schmid, J.; Skotnicki, J. S.; Tam, S.; Thomason, J. R.; Wang, Q.; Levin, J. I. Potent, selective, and orally bioavailable matrix metalloproteinase-13 inhibitors for the treatment of osteoarthritis. *Bioorg. Med. Chem.* **2005**, *13*, 6629–6644.
- (42) Wu, J.; Rush, T. S., III; Hotchandani, R.; Du, X.; Geck, M.; Collins, E.; Xu, Z. B.; Skotnicki, J.; Levin, J. I.; Lovering, F. E. Identification of potent and selective MMP-13 inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4105–4109.
- (43) Noe, M. C.; Natarajan, V.; Snow, S. L.; Wolf-Gouveia, L. A.; Mitchell, P. G.; Lopresti-Morrow, L.; Reeves, L. M.; Yocom, S. A.; Otterness,

- I.; Bliven, M. A.; Carty, T. J.; Barberia, J. T.; Sweeney, F. J.; Liras, J. L.; Vaughn, M. Discovery of 3-OH-3-methylpipelicolic hydroxamates: potent orally active inhibitors of aggrecanase and MMP-13. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3385–3388.
- (44) Blagg, J. A.; Noe, M. C.; Wolf-Gouveia, L. A.; Reiter, L. A.; Laird, E. R.; Chang, S. P.; Danley, D. E.; Downs, J. T.; Elliott, N. C.; Eskra, J. D.; Griffiths, R. J.; Hardink, J. R.; Haugero, A. I.; Jones, C. S.; Liras, J. L.; Lopresti-Morrow, L. L.; Mitchell, P. G.; Pandit, J.; Robinson, R. P.; Subramanyam, C.; Vaughn-Bowser, M. L.; Yocom, S. A. Potent pyrimidinetrione-based inhibitors of MMP-13 with enhanced selectivity over MMP-14. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1807–1810.
- (45) Kim, S. H.; Pudzianowski, A. T.; Leavitt, K. J.; Barbosa, J.; McDonnell, P. A.; Metzler, W. J.; Rankin, B. M.; Liu, R.; Vaccaro, W.; Pitts, W. Structure-based design of potent and selective inhibitors of collagenase-3 (MMP-13). *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1101–1106.
- (46) Martin, J. L.; Begun, J.; McLeish, M. J.; Caine, J. M.; Grunewald, G. L. Getting the adrenaline going: crystal structure of the adrenaline-synthesizing enzyme PNMT. *Structure* **2001**, *9*, 977–985.
- (47) Shah, R.; Akella, R.; Goldsmith, E. J.; Phillips, M. A. X-ray structure of Paramecium bursaria Chlorella virus arginine decarboxylase: insight into the structural basis for substrate specificity. *Biochemistry* **2007**, *46*, 2831–2841.
- (48) Jackson, L. K.; Brooks, H. B.; Myers, D. P.; Phillips, M. A. Ornithine decarboxylase promotes catalysis by binding the carboxylate in a buried pocket containing phenylalanine 397. *Biochemistry* **2003**, *42*, 2933–2940.
- (49) Perez-Miller, S. J.; Hurley, T. D. Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. *Biochemistry* **2003**, *42*, 7100–7109.
- (50) Thompson, J. R.; Banaszak, L. J. Lipid-protein interactions in lipovitellin. *Biochemistry* **2002**, *41*, 9398–9409.
- (51) Abad, M. C.; Arni, R. K.; Grella, D. K.; Castellino, F. J.; Tulinsky, A.; Geiger, J. H. The X-ray crystallographic structure of the angiogenesis inhibitor Angiostatin. *J. Mol. Biol.* **2002**, *318*, 1009–1017.
- (52) Pellicciari, R.; Costantino, G.; Camaiori, E.; Sadeghpour, B. M.; Entrena, A.; Willson, T. M.; Fiorucci, S.; Clerici, C.; Gioiello, A. Bile acid derivatives as ligands of the farnesoid X receptor. Synthesis, evaluation, and structure-activity relationship of a series of body and side chain modified analogues of chenodeoxycholic acid. *J. Med. Chem.* **2004**, *47*, 4559–4569.

CI800414V