Part I
Introduction

1
A Fresh Look at Molecular Structure and Properties

Bernard Testa, Giulio Vistoli, and Alessandro Pedretti

Abbreviations

α_1-AR	α_1-adrenoceptors
ADME	absorption, distribution, metabolism and excretion
MC	Monte Carlo
MD	molecular dynamics
MEP	molecular electrostatic potential
MIF	molecular interaction field
PCA	principal component analysis
PSA	polar surface area
QSAR	quantitative structure–activity relationship
SAR	structure–activity relationship
SAS	solvent accessible surface

1.1
Introduction

Molecular structure and properties are key concepts in drug design, but they may not mean the same to all medicinal chemists, not to mention other researchers involved in drug discovery and development such as biochemists, pharmacologists and toxicologists (see Chapter 2). It is therefore the merit of this book to offer a rationalization of these concepts with a view to advocating their value and clarifying their use.

One of the sources of the fuzziness surrounding these concepts may well be the implicit assumption in structure–activity relationship (SAR) studies that molecular structure contains (i.e. encodes) the information on the biological activity of a given compound. Such an assumption cannot be incorrect, since this would imply the fallacy of SAR studies. However, the assumption becomes misleading if not properly qualified to the effect that the molecular structure of a given compound contains only part of the information on its bioactivity. Indeed, what the structure of a compound encodes is information about the molecular features accounting

Molecular Drug Properties. Measurement and Prediction. R. Mannhold (Ed.)

ISBN: 978-3-527-31755-4

for its recognition by a biological system. Such a recognition obviously occurs at the molecular level – the biological components which "recognize" the compound being bio(macro)molecular entities or complexes such as membranes, transporters, enzymes, receptors or polynucleosides. The mutual recognition and interaction of bioactive compound and biochemical entity translates into the formation of a functional complex which triggers the cascade of biochemical events that leads to the observed biological response [1–3].

As far as SARs are concerned, the outcome of processes such as "recognition" and "functional response" need to be formalized for incorporation into mathematical models or simulations. The same is true for "molecular structure", which remains an abstract concept until expressed formally and in quantitative terms. This is what medicinal chemists and their biological colleagues have achieved, as formalized in Table 1.1. Indeed, SAR studies, in general, and quantitative SAR (QSAR) studies, in particular, can be subdivided into four components [4]. First, we find the biological systems themselves, be they functional proteins, molecular machines, membranes, organelles, cells, tissues, organs, organisms, populations or even ecosystems. Second, there are the molecular compounds that interact with these biological systems, be they hits, lead candidates, drug candidates, drugs, agrochemicals, toxins, pollutants and more generally any type of bioactive compounds; in (Q)SAR studies, these compounds are described by their molecular features (i.e. their structure and properties). The third component in (Q)SAR studies are the responses produced by a biological system when interacting with bioactive compounds; here again, a description in the form of pharmacokinetic, pharmacological or toxicological descriptors is necessary. As for the last component, we find mathematical models or simulations which describe how the biological response varies with variations in the molecular structure of bioactive

Tab. 1.1 The four components of SAR and QSAR studies (modified from Ref. [4]).

Component	*Definition*	*Description in SARs*
(A) Biological systems	any biological entity, from a functional protein to an ecosystem	virtual (*in silico*) 3D models; mathematical models
(B) Bioactive compounds	e.g. hits, lead candidates, drug candidates, drugs, toxins, agrochemicals, pollutants	molecular features (i.e. their structure and properties)
(C) Biological responses	the response of A when exposed to B	pharmacological or toxicological descriptors
(D) Mathematical models or simulations	virtual or mathematical models of how variations in C change with variations in the molecular structure of B	variations in C = variations in the values of the descriptors; variations in B = variations in the molecular features of the bioactive compounds

compounds. As is well known to medicinal chemists, the usual statement ". . . how the biological response varies with the structure of bioactive compounds" is a simplifying shortcut.

This book focuses on molecular features and properties, their meaning, measurement, computation, and encoding into parameters and descriptors. The present chapter serves as a general opening, and invites readers to stand back and reflect on the information contained in chemical compounds and on our description of it. We base our approach on a discrimination between the "core features" of a molecule/compound and the physicochemical properties of a compound.

1.2 Core Features: The Molecular "Genotype"

1.2.1 The Argument

In our view, the core features of a molecule are the constant (unchangeable) ones, i.e. those features whose change necessarily implies a transformation into another molecule. This view is somewhat analogous with the genome, since unless they are clones different multicellular organisms necessarily have different genotypes. For this reason we use the term molecular "genotype" to describe the ensemble of the molecular core features.

As shown in Fig. 1.1, the constant features of a molecule/compound are the number and nature of its atoms (its composition), the connectivity of its atoms

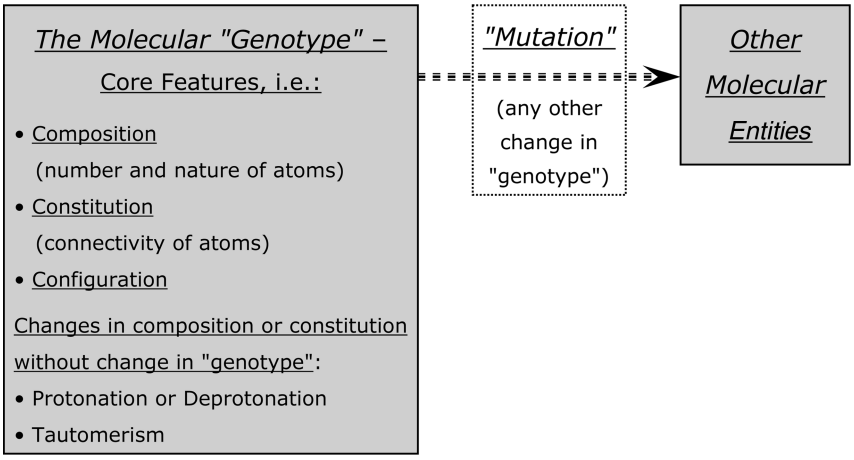

Fig. 1.1 The core features (molecular "genotype") of a molecule/compound are presented here. Attention is drawn to the fact that changes in composition, constitution (connectivity) and configuration (stereochemical features) implies a "mutation" to another molecule/compound. The exceptions are ionization and tautomerism, which are not defined as implying a "mutation" of the "genotype".

(its constitution), and its absolute configuration. Indeed, any change (i.e. "mutation") in composition, constitution or configuration yields another molecule/compound, i.e. a derivative/analog, a constitutional isomer or a stereoisomer.

Note, however, that the above scheme needs further qualification. First and strictly speaking, protonation and deprotonation involve a change in composition and connectivity, but they are reversible processes whose equilibrium is a condition-dependent property. Nevertheless, the low energy barrier and reversibility of the process lead us to view a base and its conjugated acid as two states of the same molecular "genotype". As for tautomerism, it involves a low-energy change in connectivity, again with a condition-dependent equilibrium. Again, two tautomers can be considered as two distinct states of the same compound. A further and more general proviso is the fact that our entire argument is limited to covalent bonds, with the consequence that an ion and its counterion are considered as two separate molecular entities.

1.2.2
Encoding the Molecular "Genotype"

Can various components of the core features be encoded in a form suitable for SAR investigations? Interestingly, the answer is clearly a positive one.

- Composition is partly encoded in molecular weight – a parameter sometimes used.
- Topological indices are used to describe some components of connectivity. A more complete description is afforded by unidimensional codes (linear line notations) such as SMILES. Connectivity plus explicit attention to valence electrons is afforded by the electrotopological indices (see Chapter 4).
- Configuration is described by the *R*- and *S*-descriptors for enantiomerism, and the *E*- and *Z*-descriptors for geometric π-diastereomerism [5].

1.3
Observable and Computable Properties: The Molecular "Phenotype"

1.3.1
Overview

The phenotype of an organism is its huge repertoire of observable properties. This phenotype is the visible expression of the organism' genotype, but is also controlled by the organism's life history and environment. That is to say that a given genotype can translate into a large variety of potential phenotypes – a "phenotype space" [6].

In close analogy with this biological definition, we will designate as molecular "phenotype" the ensemble of observable and computable properties of a chemical entity. These indeed are the observable expression of the core features of the

compound and like a biological phenotype they are influenced by the environment, here the molecular environment. There is a major difference, however, since compounds have no life history, but as we shall see in the last part of this chapter, compounds have a "property space" just like organisms have a phenotype space.

Energy interaction between a probe and a compound is necessary for molecular properties to be observed. As a result, properties can be categorized according to the nature of the probe used to observe them. Properties revealed by low-energy interactions are schematized in Fig. 1.2, which outlines that:

- Spectral properties arise through interactions with electromagnetic radiation.
- Some pharmacologically important properties such as pK_a, tautomeric equilibrium, conformational behavior, solubility and partitioning are temperature and solvent dependent.
- Interactions between a vast number of identical molecules give rise to such solid- or liquid-state properties as melting point and boiling point.
- Interaction with (recognition by) biomolecules triggers the cascade that leads to a biological response (see above).

The approach we follow below in surveying molecular properties is a different one based on their interdependence and the progressive emergence of biologically relevant properties (Fig. 1.3).

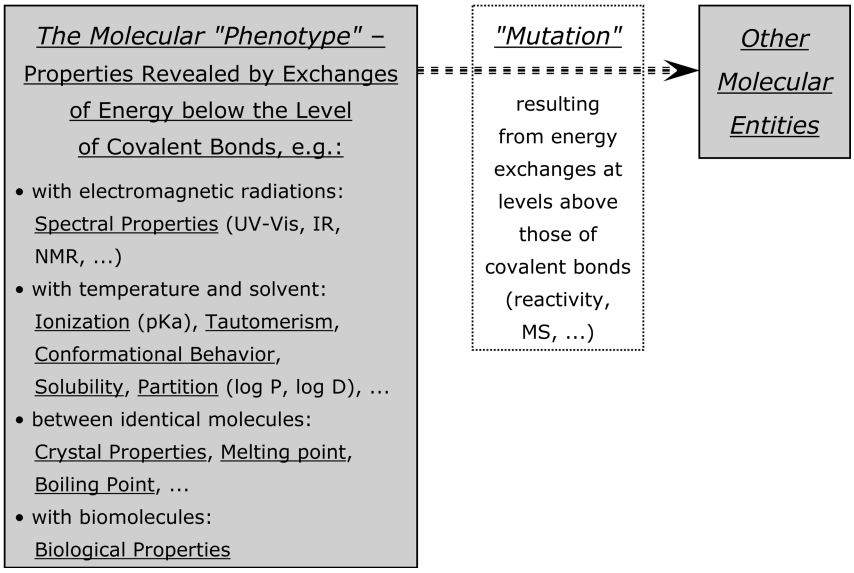

Fig. 1.2 Properties revealed by low-energy exchanges belong to the molecular "phenotype", as exemplified here. This is contrasted with some other chemical properties (e.g. reactivity) which involve the cleavage and/or formation of covalent bonds, and thus imply a "mutation" of the "genotype". UV, ultraviolet; IR, infrared; NMR, nuclear magnetic resonance; MS, mass spectroscopy.

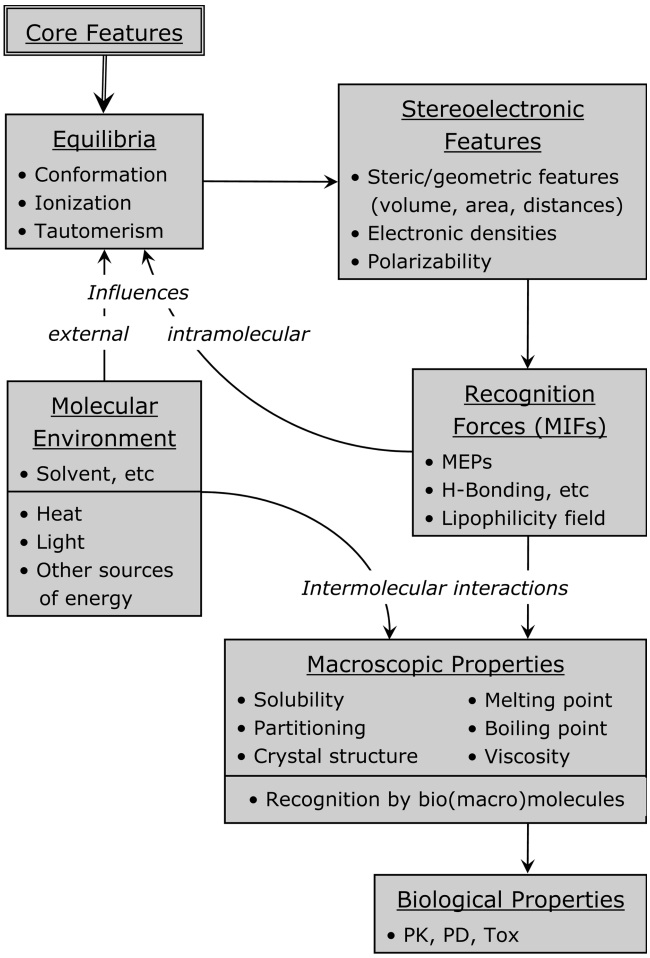

Fig. 1.3 A survey of molecular properties based on their interdependence and the progressive emergence of biologically relevant properties. See text for further details. MIFs, molecular interaction fields; MEPs, molecular electrostatic potentials; PK, pharmacokinetic(s); PD, pharmacodynamic(s).

1.3.2 Equilibria

A two-dimensional (2D) molecule is a simplified abstraction because molecules have a three-dimensional (3D) form and shape. Furthermore, form and shape fluctuate, making them four-dimensional (4D) objects. Some molecular entities may be extremely flexible, others rather rigid, but a totally rigid molecule exists only at 0 K.

A major fluctuation is the conformational behavior of molecular entities, as discussed explicitly in Chapter 9, but also in Chapters 7 and 8. Other equilibria, already mentioned above, are ionization and tautomerism. The former is the most

important as far as drug research is concerned and it is discussed extensively in Chapter 3.

1.3.3 Stereoelectronic Features

The form and shape of a molecule (i.e. its steric and geometric features) derive directly from the molecular "genotype", but they cannot be observed without a probe. Furthermore, they vary with the conformational, ionization and tautomeric state of the compound. Thus, the computed molecular volume can vary by around 10% as a function of conformation. The same is true of the molecular surface area, whereas the key (i.e. pharmacophoric) intramolecular distances can vary much more.

A similar argument can be made for electronic features such as electron density, polarization and polarizability. These are critically dependent on the ionization state of the molecule, but the conformational state is also highly influential. One highly approximate yet useful reflection of electron density is afforded by the polar surface area (PSA), a measure of the extent of polar (hydrophilic) regions on a molecular surface (see Chapter 5).

1.3.4 Recognition Forces and Molecular Interaction Fields (MIFs)

The stereoelectronic features produce actions at a distance by the agency of the recognition forces they create. These forces are the hydrophobic effect, and the capacity to enter ionic bonds, van der Waals interactions and H-bonding interactions. The most convenient and informative assessment of such recognition forces is afforded by computation in the form of MIFs, e.g. lipophilicity fields, hydrophobicity fields, molecular electrostatic potentials (MEPs) and H-bonding fields (see Chapter 6) [7–10].

Like the stereoelectronic features that generate them, the MIFs are highly sensitive to the conformational and ionization state of the molecule. However, they in turn have a marked intramolecular influence on the conformational and ionization equilibria of the compound. It is the agency of the MIFs that closes the circle of influences from molecular states to stereoelectronic features to MIFs (Fig. 1.3).

1.3.5 Macroscopic Properties

As shown in Fig. 1.3, MIFs account not only for intramolecular effects, but also for intermolecular interactions, allowing macroscopic properties to emerge. The interactions of a chemical with a solvent reveal such pharmacologically essential properties as solubility (Chapters 10 and 11) and partitioning/lipophilicity (Chapters 12–16). The interactions between a large number of identical molecules

translate into solid-state properties (including melting point and solubility) or liquid-state properties such as viscosity and boiling point. Note that these macroscopic properties are also influenced by energy influx, both directly and indirectly (via equilibria).

As the same types of intermolecular forces are involved, there is no qualitative difference between solute–solvent interactions and the recognition of a compound by a bio(macro)molecular compound.

Having explained the origin of the adaptable (condition-dependent) character of molecular properties, we now turn to illustrations of this phenomenon. Indeed, stating the variable nature of molecular properties is not sufficient to appreciate its significance in drug design and SAR studies.

1.4 Molecular Properties and their Adaptability: The Property Space of Molecular Entities

1.4.1 Overview

The concept of property space, which was coined to quantitatively describe the phenomena in social sciences [11, 12], has found many applications in computational chemistry to characterize chemical space, i.e. the range in structure and properties covered by a large collection of different compounds [13]. The usual methods to approach a quantitative description of chemical space is first to calculate a number of molecular descriptors for each compound and then to use multivariate analyses such as principal component analysis (PCA) to build a multidimensional hyperspace where each compound is characterized by a single set of coordinates.

Whereas this approach has proven very successful in comparing chemical libraries and designing combichem series, it nevertheless is based on the assumption that the molecular properties being computed are discrete, invariant ones [14]. This assumption derives from the restrictions imposed by the handling of huge databases, but like many assumptions it tends to fade in the background and be taken as fact. Yet as chemistry progresses, so does our understanding of molecular structure taken in its broadest sense, i.e. the mutual interdependence between geometric features and physicochemical properties.

The growing computational power available to researchers proves an invaluable tool to investigate the dynamic profile of molecules. Molecular dynamics (MD) and Monte Carlo (MC) simulations have thus become pivotal techniques to explore the dynamic dimension of physicochemical properties [1]. Furthermore, the powerful computational methods based in particular on MIFs [7–10] allow some physicochemical properties to be computed for each conformer (e.g. virtual log *P*), suggesting that to the conformational space there must correspond a property space covering the ensemble of all possible conformer-dependent property values.

The concept of property space is progressively being used to gain a deeper understanding of the dynamic behavior of a single compound in different media (as we illustrate below with acetylcholine, see Section 1.4.2) or bound to biological targets (the carnosine–carnosinase complex, see Section 1.4.3), but it can be used also with a set of compounds to derive fertile descriptors for dynamic QSAR analyses (4D QSAR, see Section 1.4.4).

In this dynamic vision, a molecular property can be described by either (i) an average value or (ii) descriptors defining its property space. The average value of a property, and especially a weighted average, contains more information than a conformer-specific value (even if it is that of the lowest-energy or hypothetical bioactive conformer). However, this average value does not yield information on the property space itself. To this end, one should use descriptors specifying the property range and distribution in relation to conformational changes and other property profiles.

A property space can be defined using two classes of descriptors. The first class includes descriptors quantifying the variability (spread) of values; their range is probably the most intuitive one. The second class of descriptors relates the dynamic behavior of a given property with other geometric or physicochemical properties. Such correlations can reveal if and how two molecular properties change in a coherent manner.

The relations between physicochemical properties and geometric descriptors describe the ability of a physicochemical property to fluctuate when the 3D geometry fluctuates. These relationships also lead to the concept of molecular sensitivity, since there will be sensitive molecules whose property values are markedly influenced by small geometric changes and insensitive molecules whose properties change little even during major geometric fluctuations. We can assume that molecular sensitivity may affect biological properties, as the latter are in themselves dynamic properties whose emergence will depend on the ability of a molecule to fit into and interact with an active site. Furthermore, molecular sensitivity and adaptability appear as two sides of the same coin, since sensitive molecules will need only small conformational changes to adapt their properties to the environment.

1.4.2
The Versatile Behavior of Acetylcholine

Our first exploration of property space was focused on acetylcholine. This molecule was chosen for its interesting structure, major biological role, and the abundant data available on its conformational properties [15]. The behavior of acetylcholine was analyzed by MD simulations in vacuum, in isotropic media (water and chloroform) [16] and in an anisotropic medium, i.e. a membrane model [17]. Hydrated *n*-octanol (1 mol water/4 mol octanol) was also used to represent a medium structurally intermediate between a membrane and the isotropic solvents [17].

The conformational profile of acetylcholine depended on the τ_2 and τ_3 dihedral angles since τ_1 and τ_4 remained constant during all monitored simulations (Fig. 1.4). It was found that acetylcholine assumes seven low-energy conformations

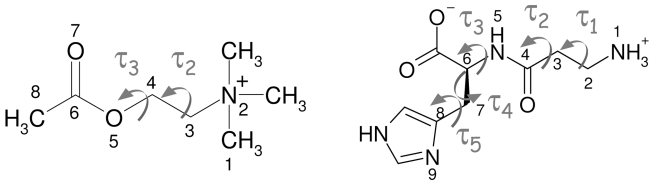

Fig. 1.4 Relevant dihedral angles in acetylcholine (left) and carnosine (right).

(i.e. the full-extended forms, τ_2=**t** and τ_3=**t**, and three pairs of chiral conformational clusters **+g+g/−g−g**; **+gt/−gt** and **t+g/t−g**), which can be clustered in folded (if τ_2 assumes synclinal conformations) and extended forms (if τ_2 is in antiperiplanar geometry). Thus, the conformational profile of acetylcholine strongly depends on τ_2, since τ_3 shows no clear preference in the range 60–300°. Clearly, the extended conformers were poorly populated in a vacuum, presumably due to intramolecular attractions between the cationic head and the electron-rich oxygen atoms. The proportion of extended conformers markedly increased in the isotropic solvents (as seen in Table 1.2) even if their increase seems due mainly to the physical presence of the solvent (i.e. friction and shielding effect) rather than to its specific physicochemical properties (i.e. polarity, H-bonding). In other words, solvent polarity does not appear to significantly affect the conformational profile of acetylcholine.

Notwithstanding this, Table 1.2 clearly shows that the behavior of acetylcholine reflected the physicochemical properties of the simulated media by adapting its property space. This is particularly evident when examining the lipophilicity averages, since the polarity of acetylcholine increased in all media compared to vacuum; although the differences between the mean log *P* values were small, they were significant as assessed by their 99.9% confidence limits.

The adaptability of acetylcholine appears even more evident when comparing the log *P* averages per conformational cluster (Fig. 1.5), which were markedly influenced by the isotropic media. Thus, all averages were lower in water than *in vacuo*, while in chloroform they assumed intermediate values, suggesting that acetylcholine can adjust its lipophilicity behavior by selecting the most suitable conformers within each conformational cluster rather than by modifying its conformational profile. The effects of water and chloroform are easily interpretable in terms of polarity and friction, but in a solvent such as octanol whose size is comparable to its own, the solute minimized steric repulsion by mimicking the shape of the solvent. In octanol, the extended conformers of acetylcholine successfully mimicked the preferred zig-zag conformation of the solvent. It is very intriguing

Tab. 1.2 Limits, ranges and mean values ±99.9% confidence limits of the molecular properties of acetylcholine conformers generated during MD simulations.

Property	*Medium*[1]				
	Vacuum (ε = 1)	*Chloroform*	*Water*	*Octanol*	*Membrane*
SAS (Å^2)	343 to 377	336 to 376	341 to 378	335 to 374	337 to 371
	34	40	37	39	34
	358 ± 0.21	356 ± 0.25	361 ± 0.30	358 ± 0.42	354 ± 0.30
PSA (Å^2)	24.2 to 44.0	28.5 to 50.4	24.4 to 44.8	32.0 to 51.1	30.1 to 49.3
	20.0	21.9	20.4	19.1	19.2
	35.0 ± 0.12	40.1 ± 0.16	37.8 ± 0.11	42.7 ± 0.20	40.7 ± 0.14
Log P_{oct}[3]	−2.53 to −2.15	−2.53 to −2.19	−2.55 to −2.20	−2.52 to −2.24	−2.51 to −2.23
	0.38	0.34	0.35	0.28	0.28
	−2.34 ± 0.0026	−2.36 ± 0.0026	−2.42 ± 0.0026	−2.40 ± 0.0030	−2.39 ± 0.0030
Dipole moment	5.51 to 10.1	7.43 to 9.54	7.80 to 9.71	7.63 to 9.45	7.56 to 9.40
	4.50	2.07	1.91	1.88	1.84
	7.78 ± 0.035	8.40 ± 0.016	8.88 ± 0.014	8.67 ± 0.020	8.66 ± 0.019
Extended geometries (%)	6.4	19.7	16.7	22.8	0

1 In each box, the first line shows the limits (minimum to maximum value), the second line the range and the third line the mean ± 99.9% confidence limits (*t*-test). The compiled results are from [16, 17].
2 Distance (in Å) between (N^+) and (OC)CH_3.
3 "Virtual" log *P* calculated by the molecular lipophilicity potential.

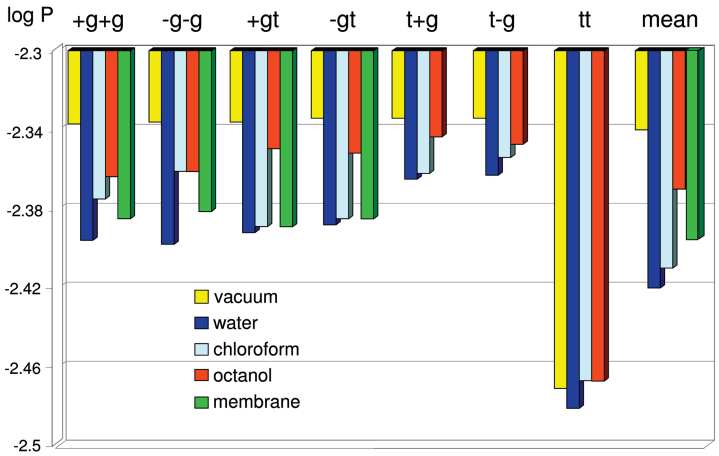

Fig. 1.5 Medium effects on average log *P* values for each conformational cluster.

to note that acetylcholine can modulate the properties of its fully extended conformers in an apparently contrasting way, selecting conformers that are simultaneously the most extended ones to better mimic the shape of the solvent and the most lipophilic ones to preserve an intermediate polarity. This suggests that

conformational space and property spaces are quite independent, and that each cluster of conformers spans most of the property space of acetylcholine.

Conversely, in a membrane model, acetylcholine showed mean log *P* values very similar to those exhibited in water. This was due to the compound remaining in the vicinity of the polar phospholipid heads, but the disappearance of extended forms decreased the average log *P* value somewhat. This suggests that an anisotropic environment can heavily modify the conformational profile of a solute, thus selecting the conformational clusters more suitable for optimal interactions. In other words, isotropic media select the conformers, whereas anisotropic media select the conformational clusters. The difference in conformational behavior in isotropic versus anisotropic environments can be explained considering that the physicochemical effects induced by an isotropic medium are homogeneously uniform around the solute so that all conformers are equally influenced by them. In contrast, the physicochemical effects induced by an anisotropic medium are not homogeneously distributed and only some conformational clusters can adapt to them.

Taken globally, the results show a remarkable adaptability of acetylcholine which can be justified considering both its intrinsic flexibility, and the fact that its intramolecular interactions are not very strong and that almost all media can compete with them. Such adaptability finds a noteworthy implication in significant pairwise correlations between physicochemical properties and geometrical descriptors as well as among physicochemical properties. Thus, Fig. 1.6 shows the revealing 3D

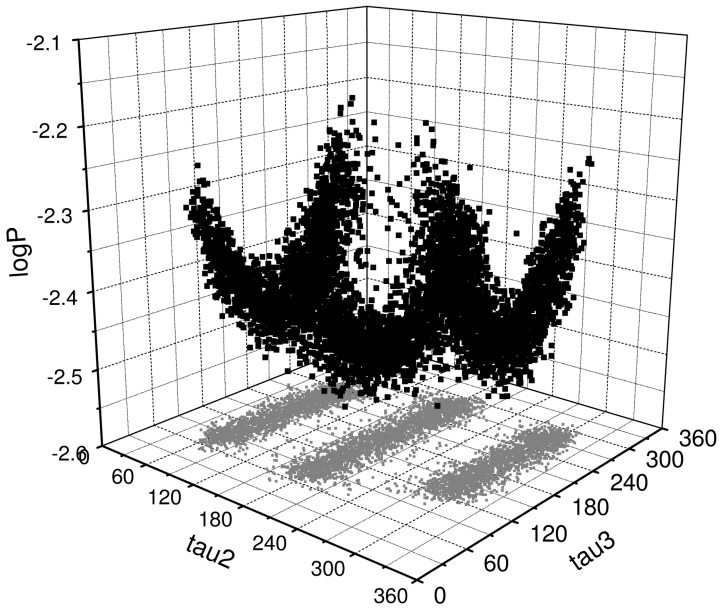

Fig. 1.6 Three-dimensional plot of τ_2 and τ_3 versus virtual log *P* as obtained from MD simulation of acetylcholine in water. Reproduced from Ref. [16] with kind permission of American Chemical Society 2005.

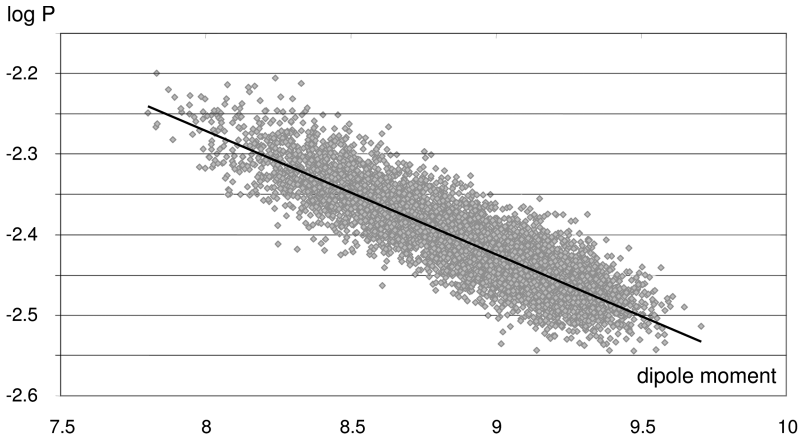

Fig. 1.7 Correlation between virtual log *P* (calculated with the molecular lipophilicity potential) and the dipole moment (r^2=0.76) as obtained from MD simulation of acetylcholine in water. Reproduced from Ref. [16] with kind permission of American Chemical Society 2005.

plots of virtual log *P* versus τ_2 and τ_3 as obtained from MD simulation in water (but all media gave fully comparable plots). Here, lipophilicity was not influenced by variations of τ_2, since the same range was covered for each of the three classes of conformers (i.e. with τ_2=**–g** or **+g** or **t**), while it was highly sensitive to variations in τ_3, with the most lipophilic conformers having τ_3=*gauche* and the most hydrophilic ones having τ_3=*trans*, suggesting that the main variations in log *P* are due to the accessibility of the ester moiety.

Among the correlations between physicochemical properties, the most noteworthy one was between dipole moment and log *P* (e.g. in water, see Fig. 1.7). Clearly, a higher dipole moment implies a greater hydrophilicity, but the fact that the two parameters correlate despite their different nature can be seen as a mutual validation of the respective algorithms used to calculate them.

1.4.3
The Carnosine–Carnosinase Complex

The second example of property space applications concerns the dipeptide carnosine (β-alanine-L-histidine, see Fig. 1.4) which represents the archetype of a series of histidine-containing dipeptides whose full physiological role remains poorly understood despite extensive studies in recent years [18–20]. Carnosine is synthesized by carnosine synthetase and hydrolyzed by dipeptidases (also called carnosinases) which belong to the metalloproteases [21].

The dynamic profile of carnosine was investigated by comparing MD simulations in isotropic solvents (i.e. water and chloroform) with simulation of the compound bound to serum carnosinase (CN1) [22]. This enzyme is characterized by its distribution in plasma and brain, and its ability to hydrolyze also anserine and homocarnosine [23]. The conformational profile of carnosine can be defined by

five torsion angles (i.e. τ_1–τ_5) [24]. The first two angles concern the β-alanine residue, while τ_3, τ_4 and τ_5 involve the L-histidine residue. In fact, τ_1, τ_2 and τ_3 remained constant during the simulations in isotropic solvents (i.e. water and chloroform) due to the strong intramolecular ion-pair which heavily influences the behavior of carnosine in its zwitterionic form. In this case, the variability in conformational and property spaces was almost totally due to the orientation of the imidazole moiety, since the simulated solvents were not able to break the intramolecular salt bridge. Specifically, the β-alanine residue was constantly rigidified in water and chloroform, with τ_1 = **+g**, τ_2 = **–g**, and τ_3 = **–g**. In contrast, the L-histidine residue assumed four different conformers depending on τ_4 and τ_5 (i.e. **t+g**, **t–g**, **–g+g** and **–g–g**).

When comparing the conformational profile of carnosine in isotropic solvents and bound to carnosinase, a contrasting behavior is apparent. Indeed, Fig. 1.8(a) clearly shows that the β-alanine residue is more flexible in the enzyme-bound complex than in isotropic solvents, while the L-histidine residue appears constrained by interactions with carnosinase (Fig. 1.8b). This discrepancy can be explained considering the interaction pattern binding carnosine to the enzyme (Fig. 1.9). Thus, the polar residues lining the catalytic site of carnosinase (including the key zinc ions) can successfully compete with the intramolecular ionic bond in carnosine, while the L-histidine residue must retain an accessible conformation which optimizes the contacts of imidazole with the enzyme. These results confirm that an isotropic solvent is unable to heavily modify the conformational profile of a solute, while an anisotropic medium (including a protein, which is also a structured anisotropic medium) can do it. Interestingly, the membrane model reduced the conformational space of acetylcholine, while its specific recognition interactions with carnosinase partially enlarged that of carnosine.

The marked rigidity of carnosine reflected in its property spaces (Table 1.3) deserves some considerations. The high polarity of carnosine is illustrated by the fact that its PSA in all media is about 50% of its SAS, whereas in acetylcholine PSA is about 10% of the SAS. It is worth observing that, despite the rigidity which characterizes the conformational profile of carnosine, this molecule can modulate its physicochemical properties according to the polarity of the medium, as seen in its lipophilicity space. Indeed, carnosine in water shows the lowest log *P* average, but the highest lipophilicity average when bound to the enzyme, probably due to the marked accessibility of the imidazole ring into the catalytic site. This profile confirms the results observed in isotropic media, that the contribution to lipophilicity of the β-alanine residue is nearly constant – the variability being mainly due to the accessibility of the histidine moiety. Similarly, the most polar group in acetylcholine (i.e. its ammonium head) gave a quite constant contribution, variability in lipophilicity being due to the accessibility of the ester moiety. This suggests that molecules can modulate the physicochemical profile of highly polar groups only with great difficulty. The marked accessibility of the imidazole ring of bound carnosine finds convincing confirmation in the average SAS and PSA values, which are highest in the carnosinase-bound form. Finally, carnosine showed the most narrow property ranges when bound to the enzyme, although

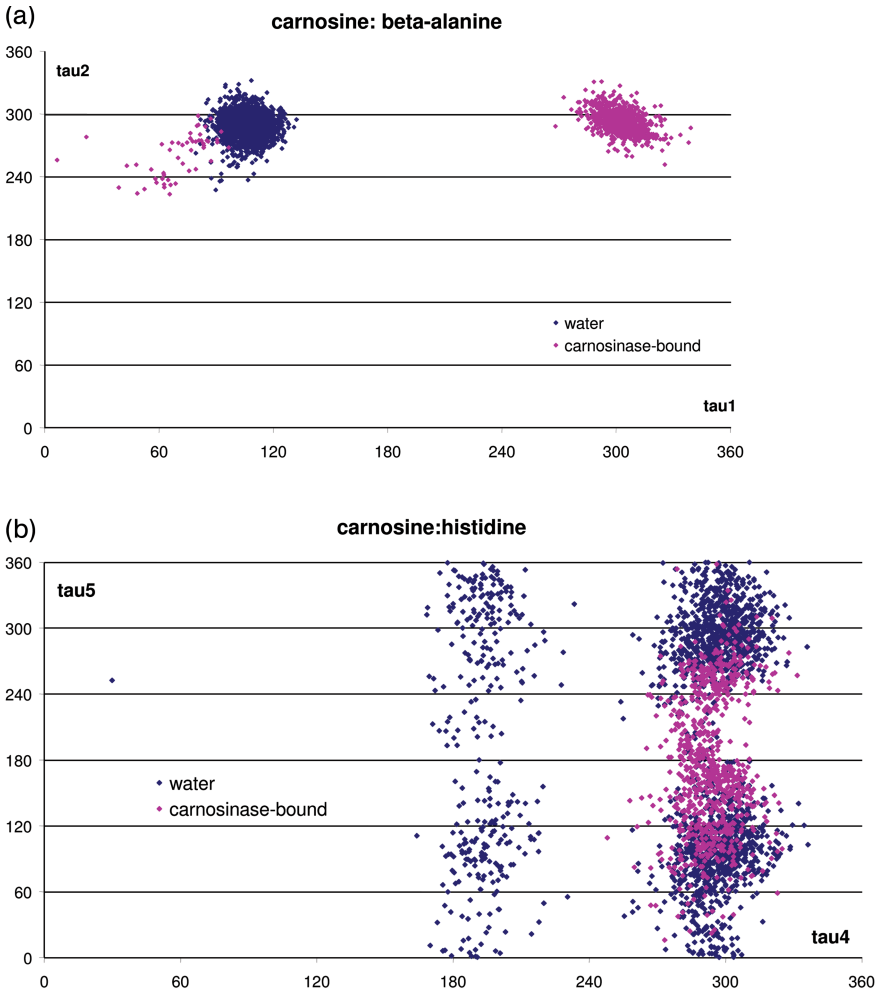

Fig. 1.8 The conformational behavior of the two amino acyl residues of carnosine (τ_2 versus τ_3 plot) as simulated for 5 ns in water (in blue) or when bound to carnosinase (in pink). (a) Conformational behavior of the β-alanine residue (τ_1 versus τ_2 plot). (b) Conformational behavior of the histidine residue (τ_4 versus τ_5 plot).

its conformational space was then markedly enlarged. This suggests that a ligand must assume well-defined property profiles to optimize its recognition by and binding to an enzyme and that this adaptation is only partly explained by a mere conformational fit. In other words, conformational space and property spaces are only partly correlated.

Taken together, our MD simulations of acetylcholine and carnosine emphasize the marked difference between them. Indeed, acetylcholine is representative of a sensitive molecule whose physicochemical and structural properties can vary in a coherent manner, aptly adapting themselves to the simulated media. Conversely,

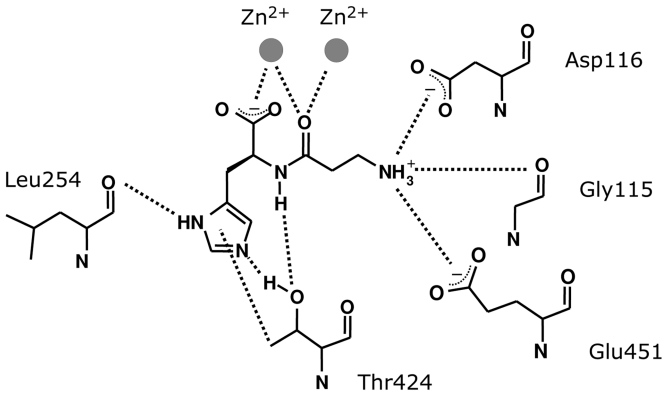

Fig. 1.9 Bidimensional representation of the interaction pattern between carnosinase and its substrate carnosine. The model shows how the enzyme recognizes (binds) the ammonium group, the carboxylate group and the unsubstituted imidazole ring. The amido bond is simultaneously bound for recognition and polarized for catalysis. Reproduced from Ref. [22] with kind permission of American Chemical Society 2005.

Tab. 1.3 Limits, ranges and mean values ± 99.9% confidence limits of the molecular properties of carnosine conformers generated during MD simulations.

Property	***Medium***[1]		
	Water	***Chloroform***	***Carnosinase***
SAS (Å^2)	395 to 434	336 to 376	410 to 446
	39	40	36
	417 ± 0.23	356 ± 0.25	427 ± 0.28
PSA (Å^2)	180 to 226	177 to 221	192 to 234
	47	44	42
	203 ± 0.33	201 ± 0.33	214 ± 0.27
Log P_{oct}[2]	−4.57 to −3.98	−4.49 to −3.85	−4.45 to −3.90
	0.59	0.66	0.55
	−4.28 ± 0.0043	−4.20 ± 0.0044	−4.17 ± 0.0042

1 In each box, the first line shows the limits (minimum to maximum value), the second line the range and the third line the mean ± 99.9% confidence limits (*t*-test).
2 "Virtual" log *P* calculated by the molecular lipophilicity potential.

carnosine appears markedly rigidified by an intramolecular ionic bridge which influences both its conformational space (which is frozen in few conformations) and its property spaces, as evidenced by their narrow ranges and insignificant pairwise correlations. Nevertheless, MD simulations revealed that carnosine could also adjust its physicochemical properties to the simulated medium, suggesting that the conformational space is easier to constrain than the property spaces, which indeed conserve a significant elasticity even in very constrained molecules. In

other words, some physicochemical adaptability to the molecular environment is retained even in rather rigid compounds. Such molecular adaptability can clearly influence biological activity and molecular descriptors accounting for adaptability might find fertile applications in QSAR as described below.

1.4.4 Property Space and Dynamic QSAR Analyses

Our third example illustrates the use in QSAR analyses of parameters describing the property range and distribution in relation to conformational changes and other property profiles. As previously stated, a property space can be defined using two classes of descriptors, i.e. the distribution of property values and the relations between properties. Thus, the relations between geometric descriptors and physicochemical properties describe the ability of a physicochemical property to fluctuate when the 3D geometry fluctuates. These relations lead to the concept of molecular sensitivity, since there will be sensitive molecules whose property values are markedly influenced by small geometric changes and insensitive molecules whose properties change little even during major geometric fluctuations.

From a mathematical point of view, such correlations may be analyzed by considering their regression coefficients. However, using regression coefficients as independent variables may lead to mathematical dead-ends. We thus looked for a descriptor of property space that would be both informative and simple to use. The descriptor we propose and evaluate is the amplitude of variation of a given physicochemical property for a given variation in molecular geometry. If we consider a physicochemical property X for which conformer-specific values can be computed (e.g. dipole moment, polar surface area, virtual log P, etc.), its pairwise sensitivity value ($Pairwise\ Sensitivity_{X,Gij}$) for two given conformers (i, j) and a given geometric descriptor G (e.g. an intramolecular distance, a torsion angle, etc.) can be defined as the ratio between the absolute value of the difference of X and the corresponding absolute value of the difference in G:

$$Pairwise\ Sensitivity_{X,Gij} = \frac{|X_i - X_j|}{|G_i - G_j|} \quad (1)$$

The global sensitivity ($Sensitivity_{X,Gij}$) will be the average of the pairwise sensitivities computed for all possible pairs of N conformers, i.e. for $N(N-1)$ pairs.

For any given physicochemical property of a molecule, one can calculate several sensitivity values according to the geometric descriptors being used. However, when investigating a set of heterogeneous compounds, a geometric descriptor applicable to all molecules must be selected. To this end, the root mean square deviation of atomic coordinates aptly describes geometric differences between pairs of conformers as a function of their atomic positions.

The objective of this example [25] was to examine whether and how range and sensitivity can be successfully used as descriptors of the space of relevant physicochemical properties, and correlated with affinities and receptor subtype selectivities for a heterogeneous set of ligands of α_1-adrenoceptors (α_1-ARs) taken

from the literature [26] and characterized by their large differences in binding affinities. The conformational space was explored using a MC procedure, and the properties considered were dipole moment, lipophilicity, polar area and surface area.

A search for relations between affinity data (pK_i) and descriptors of property spaces (range and sensitivity) failed to uncover any significant correlation (all r values <0.5). This result was expected and understandable, since affinity depends on the ligand ability to assume well-defined property values – a type of information not encoded in range and sensitivity. In contrast, significant correlations were found between some receptor selectivities and some property space descriptors. Indeed, ΔpK_{a-b} and ΔpK_{a-d} yielded significant correlations ($r > 0.7$) with log P, PSA and SAS ranges, whereas ΔpK_{b-d} yielded no correlation whatsoever ($r < 0.1$).

A clear trend was also apparent among the physicochemical properties, since the lipophilicity range yielded the best correlations for both ΔpK_{a-b} and ΔpK_{a-d}, while the dipole space yielded the lowest. Interestingly, all significant correlation coefficients were positive, implying that α_1-AR selectivities are mainly proportional to variations in physicochemical properties, as expressed mainly by range.

The above observations may imply that the ability to selectively interact with the α_{1a}-AR is encoded in property space descriptors and especially in the lipophilicity space, whereas selective interaction with the α_{1b}-AR is only partially encoded in property space descriptors and α_{1d}-AR selectivity not at all. To verify the above hypothesis, we recalculated regressions coefficients between ΔpK_{a-b} selectivity and property space parameters, removing the strongly selective α_{1b}-AR ligands. This indeed produced a slight increase (about 0.05–0.10) in all correlation coefficients between property spaces and ΔpK_{a-b}. The best correlation, i.e. between {*range*_log *P*} and ΔpK_{a-b}, is shown in Eq. 2 and Fig. 1.10:

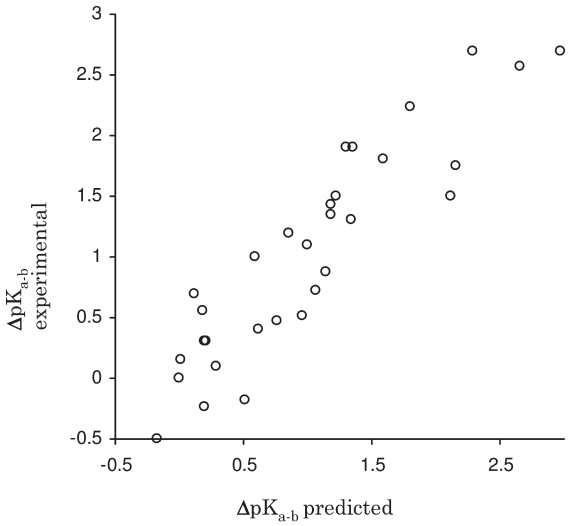

Fig. 1.10 Best one-variable correlation between ΔpK_{a-b} and *range*_log *P* (Eq. 2).

$$\Delta \mathrm{p}K_{\mathrm{a-b}} = 1.49(\pm 0.12)\{range\_\log P\} - 0.12(\pm 0.13) \quad (2)$$
$$n = 32,\ r^2 = 0.79,\ q^2 = 0.78,\ s = 0.41$$

Clearly, this equation cannot take into account α_{1b}-selective ligands (i.e. with $\Delta \mathrm{p}K_{\mathrm{a-b}} < 0$). Indeed, a hypothetical molecule with an impossibly low {*range*_log *P*} of 0 would be predicted to have a $\Delta \mathrm{p}K_{\mathrm{a-b}}$ equal to –0.12. Nevertheless, the goodness-of-fit of this equation is remarkable considering the heterogeneous nature of the ligands and its high q^2 value (i.e. good predictive power) obtained with a single independent variable.

Given the absence of correlation between the sensitivity and range descriptors, we also examined whether a two-variable equation would improve on Eq. (2). As shown by Eq. (3), the inclusion of two independent variables in the same equations improved their predictive capacity:

$$\Delta \mathrm{p}K_{\mathrm{a-b}} = 1.61(\pm 0.13)\{range\_\log P\} + 0.34(\pm 0.04)\{sensitivity\_\log P\} - 0.76(\pm 0.19) \quad (3)$$
$$n = 32,\ r^2 = 0.84,\ q^2 = 0.83,\ s = 0.38\ F = 74.38$$

Compared to Eq. (2), Eq. (3) shows a slight statistical improvement. Also, is has a better predictability for α_{1b} selective ligands, since a hypothetical molecule with very low {*range*_log *P*} and {*sensitivity*_log *P*} values would be predicted to have a $\Delta \mathrm{p}K_{\mathrm{a-b}}$ equal to –0.76.

From a methodological viewpoint, our results suggest that range and sensitivity are useful descriptors of property spaces and can parameterize the capacity of a given molecule to span broad conformational and property spaces. In other words, range and sensitivity appear as promising descriptors of the dynamic behavior of a molecule. Their application to other dynamic QSAR studies [in particular, absorption, distribution, metabolism and excretion (ADME) behavior] is under investigation.

1.5 Conclusions

As the present and other chapters show, the usual way medicinal chemists describe molecular structure and properties is necessarily limited and partial. Indeed, it deals with 4D structures (3D geometry plus conformation) in molecular graphics, 3D and 2D representations on paper, 1D strings in codes, and 0D points in chemical spaces. However, as we have tried to show here, molecules are 3D objects whose shape and the various molecular fields they generate vary in space and time, effectively making them *N*-dimensional objects. The dynamic complexity of molecules arises from their interactions with energy fields and with neighboring molecules, to such extent that a fully isolated molecule is unobservable, the concept of it being a mere abstraction. However, beyond the grasp of this paradigm lies

the challenge of expressing and using our ever-increasing wealth of molecular information in manners suitable for higher-level QSARs yet to be conceived. We hope the present book may contribute to such a progress.

References

1 Testa, B., Kier, L. B., Carrupt, P. A. A systems approach to molecular structure, intermolecular recognition, and emergence–dissolvence in medicinal research. *Med. Res. Rev.* **1997**, *17*, 303–326.

2 Testa, B. Drugs as chemical messages: molecular structure, biological context, and structure–activity relationships. *Med. Chem. Res.* **1997**, *7*, 340–365.

3 Testa, B., Kier, L. B., Bojarski, A. J. Molecules and meaning: how do molecules become biochemical signals? *SEED Electronic J.* **2002**, *2*, 84–101. http://www.library.utoronto.ca/see/SEED/Vol2-2/2-2%20resolved/Testa.pdf.

4 Testa, B., Vistoli, G., Pedretti, A. Musings on ADME predictions and structure–activity relations. *Chem. Biodivers.* **2005**, *2*, 1411–1427.

5 Testa, B. *Principles of Organic Stereochemistry*, Dekker, New York, NY, **1979**.

6 Kirschner, M. W., Gerhart, J. C. *The Plausibility of Life*, Yale University Press, New Haven, CT, **2005**.

7 Testa, B., Carrupt, P. A., Gaillard, P., Billois, F., Weber, P. Lipophilicity in molecular modeling. *Pharm. Res.* **1996**, *13*, 335–343.

8 Carrupt, P. A., Testa, B., Gaillard, P. Computational approaches to lipophilicity: Methods and applications. *Rev. Computat. Chem.* **1997**, *11*, 241–315.

9 Carrupt, P. A., El Tayar, N., Karlén, A., Testa, B. Value and limits of molecular electrostatic potentials for characterizing drug–biosystem interactions. *Methods Enzymol.* **1991**, *203*, 638–677.

10 Rey, S., Caron, G., Ermondi, G., Gaillard, P., Pagliara, A., Carrupt, P. A., Testa, B. Development of molecular hydrogen bonding potentials (MHBPs) and their application to structure–permeation relations. *J. Comput.-Aided Mol. Des.* **2001**, *19*, 521–535.

11 Lazarsfeld, P. F., Barton, H. A. Qualitative measurement in the social sciences: classifications, typologies, and indices. In *The Policy Sciences*, Lerner, D., Lasswell, H. D. (eds.), Stanford University Press, Stanford, CA, **1951**, pp. 151–192.

12 Merschrod, K. A property-space perspective for interaction terms of ordinal variables. *Quality and Quantity*, **1982**, *16*, 549–558

13 Dobson, C. M. Chemical space and biology. *Nature* **2004**, *432*, 824–828.

14 Feher, M., Schmidt, J. M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218–227.

15 Vistoli, G., Pedretti, A., Villa, L., Testa, B. The solute–solvent system: solvent constraints on the conformational dynamics of acetylcholine. *J. Am. Chem. Soc.* **2002**, *124*, 7472–7480.

16 Vistoli, G., Pedretti, A., Villa, L., Testa, B. Solvent constraints on the property space of acetylcholine. I. Isotropic solvents. *J. Med. Chem.* **2005**, *48*, 1759–1767.

17 Vistoli, G., Pedretti, A., Testa, B. Solvent constraints on the property space of acetylcholine. II. Ordered media. *J. Med. Chem.* **2005**, *48*, 6926–6935.

18 Guiotto, A., Calderan, A., Ruzza, P., Borin, G. Carnosine and carnosine-related antioxidants: a review. *Curr. Med. Chem.* **2005**, *12*, 2293–2315.

19 Bauer, K. Carnosine and homocarnosine, the forgotten, enigmatic peptides of the brain. *Neurochem. Res.* **2005**, *30*, 1339–1345.

20 Aldini, G., Facino, R. M., Beretta, G., Carini, M. Carnosine and related dipeptides as quenchers of reactive carbonyl species: from structural studies to therapeutic perspectives. *Biofactors* **2005**, *24*, 77–87.

21 Pegova, A., Abe, H., Boldyrev, A. Hydrolysis of carnosine and related compounds by mammalian carnosinases. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2000**, *127*, 443–446.

22 Vistoli, G., Pedretti, A., Cattaneo, M., Aldini, G., Testa B. Homology modeling of human serum carnosinase, a potential medicinal target, and MD simulations of its allosteric activation by citrate. *J. Med. Chem.* **2006**, *49*, 3269–3277.

23 Teufel, M., Saudek, V., Ledig, J. P., Bernhardt, A., Boularand, S., Carreau, A., Cairns, N. J., Carter, C., Cowley, D. J., Duverger, D., Ganzhorn, A. J., Guenet, C., Heintzelmann, B., Laucher, V., Sauvage, C., Smirnova, T. Sequence identification and characterization of human carnosinase and a closely related non-specific dipeptidase. *J. Biol. Chem.* **2003**, *278*, 6521–6531.

24 Dies, R. P., Baran, E. J. A density functional study of some physical properties of carnosine (*N*-β-alanyl-L-histidine). *J. Mol. Struct. Theochem.* **2003**, *621*, 245–251.

25 Vistoli, G., Pedretti, A., Villa, L., Testa, B. Range and sensitivity as descriptors of molecular property spaces in dynamic QSAR analyses. *J. Med. Chem.* **2005**, *48*, 4947–4952.

26 Bremner, J. B., Coban, B., Griffith, R., Groenewoud, K. M., Yates, B. F. Ligand design for alpha$_1$ adrenoceptor subtype selective antagonists. *Bioorg. Med. Chem.* **2000**, *8*, 201–214.