ELSEVIER

Computational quantum chemistry and adaptive ligand modeling in mechanistic QSAR

Pier G. De Benedetti[1], and Francesca Fanelli[1,2]

[1] Dipartimento di Chimica, Università di Modena e Reggio Emilia, Via Campi 183, 41125 Modena, Italy
[2] Università di Modena e Reggio Emilia, Via Campi 183, 41125 Modena, Italy

Drugs are adaptive molecules. They realize this peculiarity by generating different ensembles of prototropic forms and conformers that depend on the environment. Among the impressive amount of available computational drug discovery technologies, quantitative structure–activity relationship approaches that rely on computational quantum chemistry descriptors are the most appropriate to model adaptive drugs. Indeed, computational quantum chemistry descriptors are able to account for the variation of the intramolecular interactions of the training compounds, which reflect their adaptive intermolecular interaction propensities. This enables the development of causative, interpretive and reasonably predictive quantitative structure–activity relationship models, and, hence, sound chemical information finalized to drug design and discovery.

Ligand–protein and protein–protein interactions control any cellular process and function by means of complex/dynamic mechanisms, which involve sophisticated adaptive intramolecular and intermolecular communication pathways. In this context, the role of molecular recognition/communication between the interacting partners and their quantitative description constitutes the crucial point. Many computational approaches, at different levels of complexity, have been developed and applied to different ligand–target systems [1–4]. They essentially differ in the accuracy and resolution level of structural description and in the derived descriptors of ligand–target interactions.

Computational quantum chemistry (CQC) and correlation analysis are important tools for the mechanistic understanding and prediction of chemical–biological interactions and drug discovery. Among the available correlative approaches, quantitative structure–activity relationship (QSAR) and Quantitative Structure-Property Relationship (QSPR) modeling is the most relevant in chemistry. Moreover, CQC enables a deeper understanding of the fundamental properties of effective drugs, anticipating potential problems in developing new agents [5–7]. However, this scenario, which usually depicts a static situation, becomes more complicated when we consider the multidimensional dynamic nature of ligands, receptors and their interactions. The well-known conformational equilibria and prototropic equilibria (acid–base and tautomeric) are among the most relevant features in molecular recognition [8]. In fact, ligands and target receptors are adaptive flexible molecules, with acidic and/or basic functions, which can induce tautomeric equilibria. Complex and interdependent prototropic equilibria generate different prototropic forms, characterized by different conformations and stereo-electronic features. Hence, an essential aspect of 3D pharmacophore generation is the search for (i) the different energetically accessible prototropic forms and their corresponding conformations (present in the different biological contexts, in which the affinity and function are measured), and (ii) the quantification, with the appropriate descriptors, of the structure–affinity/function relationships (referring to QSAR studies). In spite of the recognized determinant influence of the different prototropic forms in any step of drug discovery, however, they are often ignored [8,9]. This is probably due to the absence of experimental data [9] and to the inherent complexity of simulating prototropic equilibria. However, there is no substitute for CQC methods to characterize the adaptive propensity of drugs by their computational integration of energetically accessible ensembles of prototropic forms, conformers and stereo-electronic structures [3,8]. Thus, CQC coupled with correlation analysis is an invaluable tool to comparatively describe and

Corresponding author:. De Benedetti, P.G. (deben@unimo.it)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.08.003

predict the drug adaptive pharmacophoric features in terms of simple and intuitive CQC-based causative molecular descriptors and mechanistic QSAR. Moreover, this conceptually simple scheme should overcome possible troubles in interpreting and correctly predicting the physicochemical properties and bioactivity of new molecules [10–14].

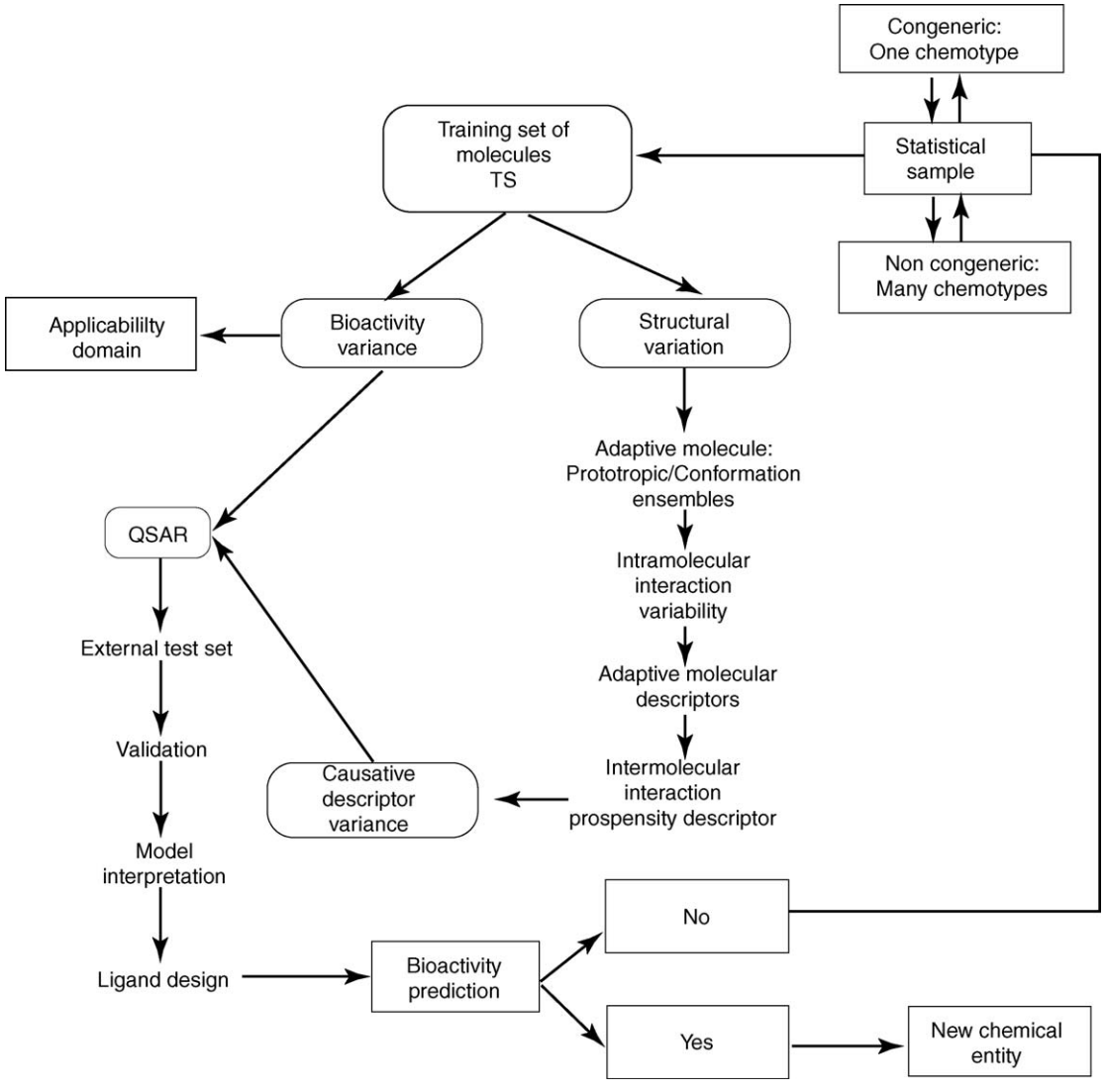
In this article we illustrate, by means of appropriate examples, the concepts of adaptive drugs and the role of CQC-derived intermolecular interaction propensity descriptors of the ligands in interpretative and predictive QSAR models.

Drugs as adaptive recognition and interacting systems

Ligand–receptor specificity relies on highly precise molecular interactions involving the spatial, electronic and dynamic structures of both ligand-binding and target-binding site [15]. Such interactions determine the activation or inhibition of receptors and/or enzymes, by information transfer and/or biochemical reactions [16]. The adaptive character of many drugs is essentially due to their organization in ensembles of prototropic forms and conformations that, in turn, depend on the environment and the stereo-electronic features of the bio-receptor binding site. In this respect, the similarity/diversity classification of ligands should be also based on the quantification of their adaptive molecular recognition and interaction propensities [17–19]. Several modeling strategies are currently available, which describe the adaptive molecular recognition with the goal of better understanding protein function and designing drugs [3,20]. They can be broadly classified into 'direct or structure-based' and 'indirect or ligand-based' [21]. The aims of both approaches are the determination of the adaptive 3D stereo-electronic complementary features of the ligand–target complexes [22].

Ligand–biomolecular receptor binding: QSAR modeling

QSAR analysis correlates information on molecular structure with information on molecular properties. Hence, QSAR is essentially dependent on the information content of the selected training set (TS) of molecules (statistical sample) and on the ways this information is analyzed, extracted, quantified and represented (Scheme 1). QSAR is based on the assumption that the molecular structure must contain the features responsible for its physical, chemical and biological properties. QSAR addresses the problem of generating correlative models between experimentally determined biolo-



SCHEME 1

gical/pharmacological data values for the TS and their experimental and/or computational descriptors. The aims are to explain the causative mechanisms of bioactivity in terms of molecular descriptors of the ligands (and the target) and to predict the bioactivity of new analogs. This implies that the variation of the biological function/affinity of the studied TS needs to be translated into a physicochemical formalism, which confers mechanistic interpretability to the QSAR model. Attention should be paid, however, to the selection of molecular/intermolecular descriptors and, above all, to their variations. This selection should be also driven by sound physical hypotheses on the adaptive interaction propensity of the ligands toward the target receptor, rather than by iteratively searching combinations of molecular descriptors, ultimately leading to the set that forces the best data onto an acceptable response landscape. In this context, the statistical goodness is evaluated by a fitness function that selects, among thousands of descriptors, those descriptors that predict the bioactivity with the best statistics [10]. This implies that if the number of molecular descriptors tends towards infinity, the QSAR model tends towards statistical perfection. Recalling the two founding aspects of QSAR modeling (i.e. the interpretative and the predictive ones), the former turns out to be the most relevant in the context of causative QSAR.

Local, pharmacophore-based and global QSAR models

QSAR modeling can be schematically classified as local, pharmacophore-based and global. Local or substituent-based models, which consider a limited chemical space (the TS is strictly congeneric), are usually easily interpretable and predictive. Local QSAR (2D-QSAR) is an extension of Hammett's equation to biological systems (Hansch's analysis). 2D-QSAR modeling can be efficiently used in the lead optimization step, by modifying type and position of small substituents on a large common chemical scaffold recognized to bear a specific bioactivity. In this framework, bioactivity is modulated by the different effects/properties of the substituents (electronic, steric and lipophilic). Being 2D-QSAR models based on the properties of substituents and their additive mechanistic effects, however, no information can be inferred about the mechanism of intramolecular interactions/electronic perturbations induced by the variable substituents or about the adaptive intermolecular interaction propensities of the ligands and their derived electronic/reactivity descriptors.

By contrast, the pharmacophore-based QSAR models expand the chemical and biological space explored on the basis of a common mechanistic interpretation of drug interaction [23,24]. According to the IUPAC definition, 'A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response...The pharmacophore can be considered as the largest common denominator shared by a set of active molecules [23].' This definition perfectly integrates the above outlined concepts on conformational and prototropic ensembles. Thus, the concept of pharmacophore can be considered as a derivation and extension of the common biospecific molecular scaffold (parent compound) shared by the TS in 2D-QSAR modeling. This extension reflects the adaptive conformational/prototropic propensity and dynamic nature of ligands and the target binding site. The common interacting structural features of the TS that define the pharmacophore are usually considered qualitative constant elements, however, and their possible quantitative variations, in terms of ligand/pharmacophore electronic structure, are frequently ignored.

Finally, the global models are heavily based on statistics and on different similarity search algorithms (including 3D-pharmacophore search) and consider a very large number of compounds characterized by considerably different and (apparently) unrelated chemical structure [4]. In this context, recent QSAR concepts have been developed in a multi-dimensional ligand-property/descriptor space (4D-6D/QSAR) [25]. Although these are useful QSAR approaches, there seems to be a trade-off between statistically rigorous and sophisticated models, which are hard to interpret, and more simple visual models, which are less predictive but more interpretative than the former and, hence, useful for drug design [26].

On the interpretability, predictability, applicability and validity of QSAR models

The dualistic aspect of any QSAR model (i.e. interpretative and predictive [27]) is a sort of 'uncertainty principle' that governs its performance. In fact, if we improve the predictive aspect, we undermine (and complicate) the interpretative one and vice versa. In our view, the employment of descriptors that can be reliably interpreted and translated into chemical concepts and formalism, useful for the design of new chemical entities, is an essential requirement for any QSAR approach. In this respect, CQC methods enable, through the employment of causative molecular descriptors, to obtain mechanistic and self-explanatory QSAR models, which do not require any further complex statistical elaboration for the chemical interpretation of the obtained QSAR models, as very recently proposed [28]. Indeed, CQC-based QSAR models profit from a detailed description of the intermolecular interaction propensity of the TS compounds in terms of quantum chemical reactivity descriptors computed on all (in principle) the energetically accessible prototropic forms and conformers (Scheme 1). This enables a complete and accurate analysis of the similarities and differences among the components of the considered TS. Hence, mechanistic information on drug–target interactions can be inferred from ligand-based CQC–QSAR modeling. In this framework, drug design and discovery are facilitated because the CQC description of drugs and their metabolites can be easily translated into new chemical entities (inverse QSAR problem). The following selected examples of CQC–QSAR models attempt to illustrate the above concepts. The employed CQC descriptors were computed by semi-empirical CQC methods that, for correlative purposes, are accurate enough and relatively fast [5,29,30].

Selected examples: ligand-based QSAR modeling of enzyme–inhibitor interactions

The antibacterial sulfanilamides (SAs) and sulfones (SOs) are inhibitors of dihydropteroate synthase (DHPS) (Fig. 1a), which catalyzes the condensation of dihydropyridine pyrophosphate and 4-aminobenzoate. This class of drugs has been the subject of extensive QSAR modeling [31]. Very recently, Doweyko [11] compared two different statistically valid QSAR models of the bacteriostatic activity of SAs as an illustrative and contrasting example of correlation-inferred causation. One of the two considered models, the first to appear in the literature, correlated the activity of 50 SAs

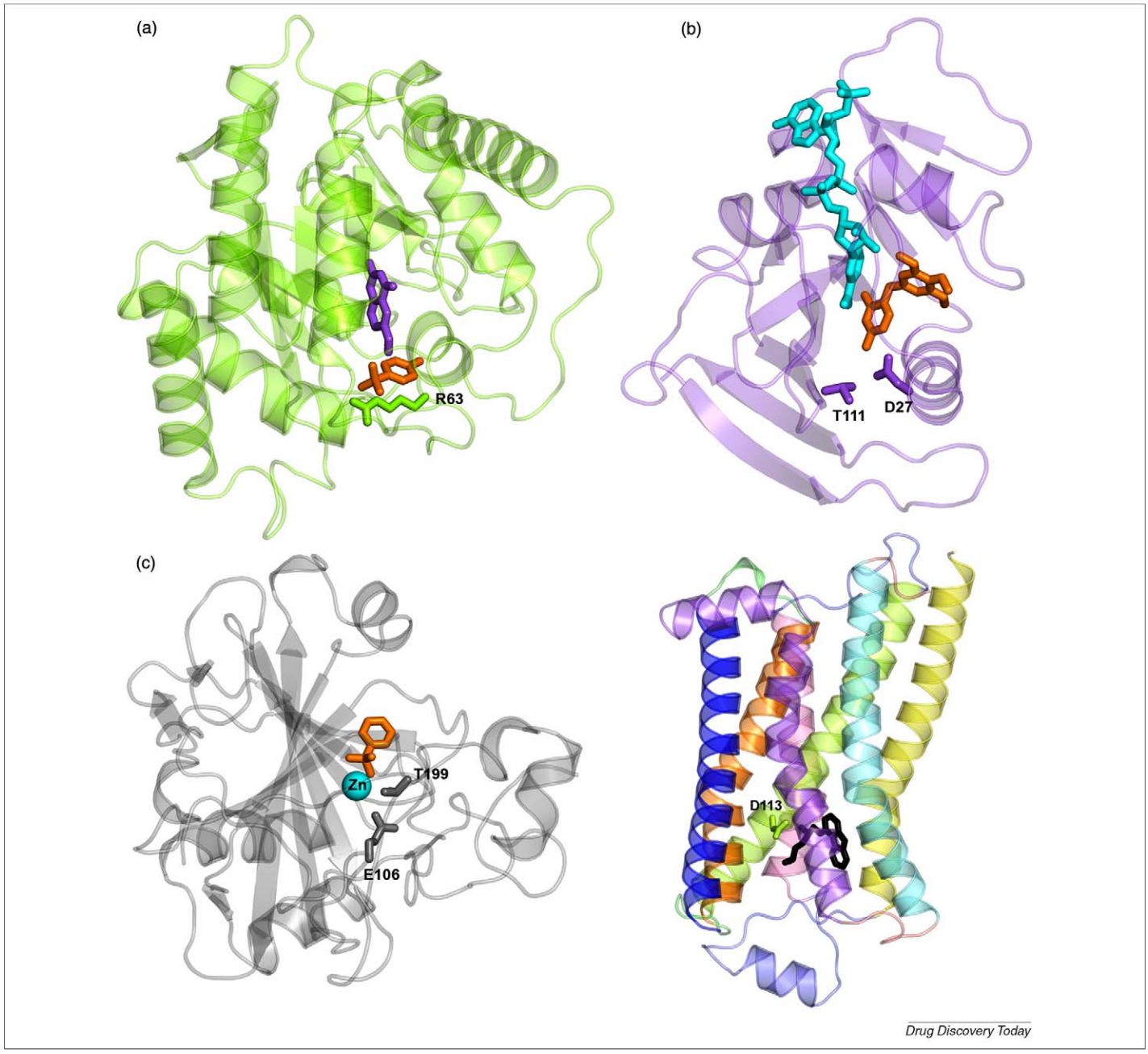


FIGURE 1

Structural models of selected biosystems. Cartoons of the crystal structures are shown of (**a**) dihydropteroate synthase (PDB code: 1AJ0) in complex with sulfanilamide (orange), drawing in sticks the active site R63; (**b**) dihydrofolate reductase (PDB code: 3FRE) in complex with trimethoprim (orange) and nicotinamide adenine dinucleotide phosphate (NADPH, cyan), drawing in sticks the active site D27 and T111; cyan; (**c**) carbonic anhydrase (PDB code: 2WEJ) in complex with benzensulfonamide (orange), drawing in sticks selected active site residues (E106 and T199) and the Zn^{2+} ion as a cyan sphere; and (**d**) β_2-AR (PDB code: 2RH1) in complex with carazolol (black), drawing in sticks the binding site aspartate (D113) conserved in the AMINE GPCRs.

with measured pK_a values, resulting in a parabolic relationship [32]. By contrast, the second model was the result of 3D-QSAR Comparative Molecular Field Analysis (CoMFA) analyses on the same set of compounds. The conclusion drawn by Doweyko, based upon the comparison of the two QSAR models, was that the 'interaction fields explain the variation in the observed bacteriostatic activity without the need to consider p$K_{a'}$. On these bases, the author inferred that two fundamentally different potential explanations for the biological variability of SAs were both statistically valid [11]. The occurrence of statistically valid QSAR models with substantially different mechanistic implications is a frequent situation. In our opinion, this is due to the jungle of computational descriptors, to the automatism of QSAR methods and to the inadequate chemical knowledge of the correlative problem in hand. The fundamentals of this judgment reside in the results of CQC-based QSAR study of SA and SO inhibitors summarized below, which prove that an adequate description of variations in the electronic structure of congeneric series of compounds can lead to unequivocal mechanistic inferences [33–36]. Remarkably, according to the nature of the variable N-substituents (which

TABLE 1

Selected QSAR models.

Eq. no.	Equation	Refs.
(**1**)	$-\log MIC_{50} = -37.12q_{ox} - 22.18$; $n = 26$, $r = 0.94$, $s = 0.50$, $F = 40.5$	[34]
(**2**)	$-\log DHPS_{50} = -0.121(\pm 0.016)\Delta q_{ox} + 13.57(\pm 2.19)$; $n = 23$, $r = 0.94$, $s = 0.169$, $F = 166.13$	[35]
(**3**)	$\log II_{50} = 106.8(\pm 12.7)q_O - 1.2(\pm 0.15)I + 57.9(\pm 6.6)$; $n = 48$, $r = 0.945$, $s = 0.321$, $F = 183$	[49]
(**4**)	$pK_{5HT1A} = 0.036(\pm 0.005)V_{in} - 6.6(\pm 0.61)$; $n = 31$, $r = 0.83$, $s = 0.65$, $F = 66.20$	[57]
(**5**)	$pK_{5HT1A} = 6.60\ (\pm 0.81)V_{dif} - 3.82(\pm 0.30)$; $n = 30$, $r = 0.83$, $s = 0.67$, $F = 62.67$	[57]
(**6**)	$pK_{\alpha 1a} = 6.24(\pm 0.84)V_{in}/V_{mol} + 3.16(\pm 0.76)$; $n = 34$, $r = 0.79$, $s = 0.48$, $F = 55.19$	[57]
(**7**)	$pK_{\alpha 1b} = 5.87(\pm 0.69)V_{in}/V_{mol} + 3.44(\pm 0.53)$; $n = 30$, $r = 0.84$, $s = 0.49$, $F = 71.50$	[57]
(**8**)	$pK_{\alpha 1d} = 7.14(\pm 0.84)V_{in}/V_{mol} + 2.18(\pm 0.67)$; $n = 31$, $r = 0.84$, $s = 0.58$, $F = 71.99$	[57]
(**9**)	$pK_{\alpha 1d} = 9.74(\pm 1.14)V_{dif} + 5.17(\pm 0.33)$; $n = 31$, $r = 0.84$, $s = 0.57$, $F = 72.86$	[57]
(**10**)	$pK_{\alpha 1a} = 5.74(\pm 0.65)V_{dif} - 2.14(\pm 0.44)$f-$P_{\pi-\pi}$(C–O) + $8.61(\pm 0.35)$; $n = 34$, $r = 0.86$, $s = 0.41$; $F = 43.24$	[57]
(**11**)	$pK_{\alpha 1b} = 0.017(\pm 0.0018)V_{in} - 0.0052(\pm 0.00062)$WPSA-1 + $5.59(\pm 0.42)$; $n = 30$, $r = 0.89$, $s = 0.43$, $F = 52.26$	[57]
(**12**)	$pK_{\alpha 1d} = 7.08(\pm 0.68)V_{in}/V_{mol} + 52.18(\pm 14.08)$f-$P_{\sigma-\pi}$(C–X); $n = 31$; $r = 0.90$; $s = 0.48$; $F = 59.84$	[57]
(**13**)	$\log(NMS/OXO\text{-}M) = 0.027(\pm 0.003)S_{2+4}HOMO^* + 0.3(\pm 0.2)$; $n = 29$, $r = 0.842$, $s = 0.614$, $F = 63.14$	[8]

Correlated biological properties include: $-\log MIC_{50}$, SA bacteriostatic activity; $-\log DHPS_{50}$, SO DHPS inhibition index; $\log II_{50}$, sulfonamide CA inhibition index; pK_{5HT1A}, $pK_{\alpha 1a}$, $pK_{\alpha 1b}$, $pK_{\alpha 1d}$ binding affinities for the 5HT1A, as well as the three α1-AR subtypes; log(NMS/OXO-M) cortical muscarinic efficacy. Correlated size/shape and quantum chemical descriptors include: q_{ox}, oxygen total net charge of the $-SO_2^-$ group; Δq_{ox}, oxygen total net charge of the $-SO_2^-$ group; q_O, oxygen total net charge of the $-SO_2NH_2$ group; I, indicator variable; f-$P_{\pi-\pi}$(C–O), bond order; f-$P_{\sigma-\pi}$(C–X), σ–π bond order (X signifies oxygen or nitrogen); WPSA-1, surface weighted charged partial surface area; S_{2+4} HOMO*, electrophilic superdelocalizability summed over the last three highest occupied molecular orbitals (HOMO) of the heteroatoms at positions 2 and 4; V_{in} and V_{out}, inner and outer *vdW* volumes with respect to the *vdW* volume of the supermolecule V_{sup}; $V_{dif} = V_{in} - V_{out}/V_{sup}$. Statistical parameters include *n*, number of compounds; *r*, correlation coefficient; *s* standard deviation; and *F*, Fisher's ratio; the numbers in parentheses are the 95% confidence intervals of regression coefficient and intercept.

modulate the pK_a values), in the test solution SAs can exist in different prototropic ensembles: amidic (4-NH_2–C_6H_4–SO_2–NH–R), imidic (4-NH_2–C_6H_4–SO_2–N=RH) and anionic (4-NH_2–C_6H_4–SO_2–N^-–R). The studies indicated that these different molecular forms show very different electronic features in their common moiety 4-NH_2–C_6H_4–SO_2– (very similar to the substrate para-aminobenzoate 4-NH_2–C_6H_4–CO_2^-) and, consequently, different bioactivity [8]. On these bases, linear QSAR of large series of SAs and SOs led to the conclusion that the electronic structure of their common moiety – modulated by the variable prototropic forms, molecular scaffolds and substituents – is the determinant factor of the inhibitory potency (i.e. both bacteriostatic and enzymatic) of these ligands [33–36]. In particular, the more electron-rich the common moiety is, the more active the compounds are (Table 1, Eqs. (1) and (2)). In this respect, the substrate 4-aminobenzoate shows the best electronic features (i.e. the highest electron density on both the COO^- and 4-NH_2 groups) [35]. Remarkably, the ligand-based QSAR model led the authors to hypothesize a sequence of events characterizing the interaction between SA/SO and the DHPS active site [31]. The hypothesized interaction model was such that initially, the negative oxygen atom of the SO_2 group makes a charge-reinforced hydrogen bond (H-bond) with a positively charged amino acid residue of the enzyme (Fig. 1a) and, successively, an electron flow from the 4-NH_2 group to the SO_2 group becomes operative, thus increasing the H-bonding donor propensity of the amino group for the subsequent formation of the H-bond with an oxygen atom of dihydropyridine pyrophosphate. According to this model, the variable molecular scaffolds of SAs (substituted aryl and heteroaromatics) and SOs (multi-substituted aryl) were assumed not to be directly involved in the interaction with the enzyme's binding site [31]. The interaction model inferred from CQC-ligand-based QSAR was validated by recent X-ray structure determinations of the SA–DHPS and SO–DHPS complexes, thus supporting the interpretative potential of Quantum Chemical (QC) descriptors [37–40].

Another interesting example of electronic/electrostatic interactions is given by the 5-benzyl-2,4-diaminopyrimidines [29], which are highly selective inhibitors for bacterial dihydrofolatereductase (Fig. 1b). The conformational behavior of selected benzylpyrimidines has been extensively investigated in several studies, which, however, neglected the prototropic forms (i.e. protonated ligands) deputed to enzyme recognition [41–44]. The importance of considering the proper prototropic form was demonstrated by a computational study that showed the effects of protonation on the CQC descriptors computed on the minimum energy conformers of benzyl-pyrimidine [29]. Indeed, it was shown that protonation of N1, although necessary for productive electrostatic interaction with an active site aspartate, renders the adjacent 2-amino and 4-amino groups 1588 and 247 times more efficient H-bond donors (as evaluated by comparing the nucleophilic superdelocalizabilities of the N1-protonated form with that of the deprotonated one [29]) towards several active site amino acids, namely the side chains of the conserved aspartate and threonine (for the 2-amino group; Fig. 1b) and the backbone oxygen atom of two hydrophobic amino acids (for the 4-amino) [45,46]. Remarkably, the computed minimum energy conformer of the N1-protonated benzyl-pyrimidine was found to be very similar that in the crystal structure of the trimethoprim *Escherichia coli* dihydrofolatereductase binary and ternary (i.e. including NADPH) complexes [46]. The results of these studies suggested that the N1-protonated diaminopyrimidine framework has a central role in modulating the adaptive conformation and the efficiency of the H-bond network in the inhibitor–enzyme interaction.

Finally, another meaningful example of effectiveness of QC descriptors concerns the carbonic anhydrase (CA) inhibitors (Fig. 1c). The biological relevance of CA has been highlighted in

a recent review article, in which CA is considered as a model system for protein studies [47]. This zinc enzyme catalyzes the reversible hydration of carbon dioxide, recognizing both its regular substrate CO_2 and the product HCO_3^-. Aromatic and heterocyclic sulfonamides are very specific inhibitors of CA when the sulfonamido group ($-SO_2NH_2$) is unsubstituted. Such inhibitors bind as anions (as $-SO_2NH^-$ that is the deprotonated form of $-SO_2NH_2$ and a transition state analog of the substrate HCO_3^-) to the active site of CA, the sulfonamide nitrogen atom being involved in the coordination of the Zn^{2+} ion. Moreover, the NH^- group participates in an H-bond with the oxygen atom of T199, which, in turn, is engaged in another H-bond with the carboxylate group of E106 (Fig. 1c). One of the oxygen atoms of the $-SO_2NH^-$ moiety also participates in an H-bond with the backbone NH group of T199. This interaction model from X-ray crystallography determinations [47] is an exemplar of the complex adaptive interdependencies between intramolecular (e.g. between SO_2NH^- group and its bound substituted aromatic/heteroaromatic ring) and intermolecular (e.g. between inhibitor and CA binding site residues) interactions. QC studies on wide series of aromatic and heterocyclic sulfonamides served to determine the stereo-electronic features of the protonated (neutral) and deprotonated (anionic) forms of sulfonamides [48–50]. The most representative ligand-based QSAR model that summarizes, by a simple linear equation, the correlative behavior of 48 CA inhibitors (i.e. 20 substituted aromatic and 28 heteroaromatic sulfonamides) is represented by Eq. (3) in Table 1. This model suggests that the less electron rich and the less nucleophilic the $-SO_2NH_2$ and $-SO_2NH^-$ groups are, the more active the inhibitors are [48,49]. This conclusion is unusual. In fact, although the interacting species are oppositely charged, there exists an inverse proportionality between the negative charge on the $-SO_2NH^-$ group and the inhibitory activity [48,49]. This apparent contradiction can be explained by recalling the important role of proton exchange in the two-step mechanism of CA inhibition by sulfonamides [47].

The information from ligand-based CQC–QSAR was also integrated with that from intermolecular interaction descriptor-based QSAR (i.e. based on molecular mechanics calculations on ligand–enzyme complexes) providing meaningful interpretations of the functioning mechanism of CA inhibitors [48–51].

Selected examples: ligand-based QSAR modeling of G-protein-coupled receptor ligands

Although integral membrane proteins have essential roles in numerous physiologic functions, structure-based drug design on these systems is hampered by the slenderness of high-resolution structural models of these systems [52]. The most privileged targets of currently used drugs are G-protein-coupled receptors (GPCRs), which constitute the largest superfamily of membrane proteins [3]. GPCRs are dynamic and flexible proteins that exist as conformational ensembles [3]. Thus, the ability to obtain good ligand-based QSAR models by specific ligands showing large chemo-type variability (which reflects the adaptive/dynamic nature of the GPCR binding site) depends primarily on the availability of descriptors able to capture the strict ligand–receptor dynamic complementary criteria, which determine the bio-effect [53,54]. It has been shown that the supermolecule approach and the derived size and shape descriptors defined on the ligand bioactive protonated form were successful in deriving simple QSAR models for molecular series of very heterogeneous ligands of AMINE GPCRs [17,55–57]. In this framework, CQC served to prime ligand conformations for subsequent size/shape-based QSAR modeling. The first crystal structure of an AMINE GPCR, the β_2-adrenergic receptor (β_2-AR) in complex with the partial inverse agonist carazolol, was released at the end of 2007 [58]. The structure confirms the essential role of a conserved aspartate in helix 3 as a fundamental recognition point for the protonated nitrogen atom of the ligands, as predicted by extensive computational modeling experiments both on isolated ligands (using semi-empirical quantum chemical methods) and on ligand–receptor complexes (obtained by molecular mechanics and dynamic calculations) [57] (Fig. 1d). According to the ligand 3D pharmacophore, the supermolecule approach assumes that the *vdW* volume, obtained by superimposing the pharmacophoric elements of the most structurally different ligands characterized by the highest affinities for the same receptor (i.e. supermolecule), might reflect the overall shape and conformational plasticity of the high affinity state of the receptor binding site. With this approach, size and shape descriptors are computed by comparing the *vdW* volume of the CQC-energy minimized structure of each protonated ligand with the *vdW* volume of a supermolecule chosen as a template. For each subset of analogs, the ligand showing the highest affinity for a given receptor is chosen as a component of the reference supermolecule. The CQC-energy minimized ligand(s) chosen for the construction of the supermolecule are superimposed by a topologically rigid body fit procedure based on given pharmacophoric criteria. To compute molecular descriptors relative to the supermolecule, all other CQC-energy minimized compounds constituting the TS are, thus, rigidly superimposed on the analog compound present in the supermolecule or on its structurally closest compound. Size and shape descriptors defined within the supermolecule approach include V_{in} and V_{out}, which are, respectively, the intersection and the outer *vdW* volume of the considered ligand with respect to the volume of the reference supermolecule, and $V_{dif} = (V_{in} - V_{out})/V_{sup}$, where V_{sup} is the molecular volume of the reference supermolecule. According to the definition of these size and shape descriptors, higher affinities are realized by maximizing V_{in} and by minimizing V_{out}. For its formulation, V_{dif} is a normalized size and shape descriptor, which takes into account the information content encoded by both V_{in} and V_{out}.

The good performance of the supermolecule approach, which mimics short-range intermolecular interactions (long-range electrostatic intermolecular interactions being mainly satisfied by the charge-reinforced H-bond between the protonated nitrogen atom of the ligand and the conserved aspartate of the receptor binding site), was shown in several QSAR modeling studies (i.e. Table 1, Eqs. (4)–(12)), in which the supermolecule concept characterizes both the affinity and the selectivity profiles of the considered receptors [17,55–57]. The approach demonstrated was useful as a tool to describe both congeneric and non-congeneric series of compounds in an extended chemical space, amenable of continuous upgrading by new experimental bio-data. Moreover, the approach enables systematic structural permutations of chemically different pharmacophores and scaffolds, which can aid *de novo* design of high affinity compounds.

One of the first examples of CQC–QSAR useful for *in silico* functional screening of GPCR ligands is represented by a study on functionally different ligands (i.e. antagonists, partial and full ago-

nists) of the M1 muscarinic receptor [8,59]. Indeed, it was possible to obtain semi-empirical CQC descriptors, which correlated linearly with the ligand cortical muscarinic efficacy (Table 1, Eq. (13)). The study also emphasized the role of protonation of the tertiary amine function on the H-bonding acceptor propensity (i.e. quantified by the electrophilic superdelocalizability descriptor) of heteroatoms (i.e. nitrogen and oxygen) at positions 2 and 4 [8,59]. Such propensity increased on going from the deprotonated to the protonated forms for agonists, whereas the opposite held for antagonists.

Finally, the most relevant structure–activity relationships on novel potent AT1 angiotensin II receptor antagonists, based on the 4-phenylquinoline structure, were rationalized by considering the electrophilic superdelocalizability computed over the anionic moiety of the antagonists [60]. This reactivity descriptor enabled the classification of ligands with respect to their binding affinities, being 0.43 for subnanomolar ligands, between 0.36 and 0.40 for nanomolar ligands, and approximately 0.32 for micromolar ligands [60].

Concluding remarks

Information/computation technologies provide an astonishing amount of methods/approaches for drug discovery, of which CQC methods are the most integrative and accurate, as well as reasonably fast to calculate (at least at the semi-empirical level of theory). Drug discovery, at the molecular level, is heavily based on the understanding of the adaptive interplay between intramolecular and intermolecular interactions in ligand–drug-target complexes. CQC methods, if focused on the study of the adaptability of ligands/drugs, generate affordable reactivity descriptors (at the atomic and electronic levels) based on the TS variation of intramolecular interactions, thus generating ligand-based adaptive interaction propensity descriptors that produce causative/interpretive QSAR and, hence, sound information for drug design. In conclusion, the enthalpic characterization of ligand interaction propensity for the receptor can be more precisely determined using CQC computations. This approach is even more useful for drug resistance [61], drug-target residence time [62], computational multitarget screening drug discovery [63] and covalent-directed drug discovery [64], in which the correct control of adaptive propensity of ligands/drugs is essential.

Acknowledgement

This study was supported by a Telethon-Italy grant no. S00068TELU (to F.F.).

References

1 Jorgensen, W.L. (2004) The many roles of computation in drug discovery. *Science* 303, 1813–1818
2 Jorgensen, W.L. (2009) Efficient drug lead discovery and optimization. *Acc. Chem. Res.* 42, 724–733
3 Fanelli, F. and De Benedetti, P.G. (2005) Computational modeling approaches to structure-function analysis of G protein-coupled receptors. *Chem. Rev.* 105, 3297–3351
4 Agrafiotis, D.K. *et al.* (2007) Recent advances in chemoinformatics. *J. Chem. Inf. Model.* 47, 1279–1293
5 Karelson, M. *et al.* (1996) Quantum-chemical descriptors in QSAR/QSPR Studies. *Chem. Rev.* 96, 1027–1044
6 Cavalli, A. *et al.* (2006) Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* 106, 3497–3519
7 Raha, K. *et al.* (2007) The role of quantum mechanics in structure-based drug design. *Drug Discov. Today* 12, 725–731
8 De Benedetti, P.G. *et al.* (1995) Prototropic molecular forms and theoretical descriptors in QSAR analysis. *J. Mol. Struct. Theochem* 333, 1–17
9 Martin, Y.C. (2009) Let's not forget tautomers. *J. Comput. Aided Mol. Des.* 23, 693–704
10 Johnson, S.R. (2008) The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* 48, 25–26
11 Doweyko, A.M. (2008) QSAR: dead or alive? *J. Comput. Aided Mol. Des.* 22, 81–89
12 Zvinavashe, E. *et al.* (2008) Promises and pitfalls of quantitative structure–activity relationship approaches for predicting metabolism and toxicity. *Chem. Res. Toxicol.* 21, 2229–2236
13 Scior, T. *et al.* (2009) How to recognize and workaround pitfalls in QSAR studies: a critical review. *Curr. Med. Chem.* 16, 4297–4313
14 Cronin, M.T.D. and Schultz, T.W. (2003) Pitfalls in QSAR. *J. Mol. Struct. Theochem* 622, 39–51
15 Bissantz, C. *et al.* (2010) A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* 10.1021/jm.100112j
16 Bahar, I. *et al.* (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. Rev.* 110, 1463–1497
17 De Benedetti, P.G. *et al.* (2000) The ad hoc supermolecule approach to receptor ligand design. *J. Mol. Struct. Theochem* 503, 1–16
18 Martin, Y.C. (2001) Diverse viewpoints on computational aspects of molecular diversity. *J. Comb. Chem.* 3, 231–250
19 Cozzini, P. *et al.* (2008) Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* 51, 6237–6255
20 Raimondi, F. *et al.* (2008) Mechanisms of inter- and intra-molecular communication in GPCRs and G proteins. *J. Am. Chem. Soc.* 130, 4310–4325
21 Menziani, M.C. and De Benedetti, P.G. (1991) Direct and indirect theoretical QSAR modelling in sulfonamide carbonic anhydrase inhibitors. In *QSAR: Rational Approaches on the Design of Bioactive Compounds* (Silipo, C. and Vittoria, A., eds), p. 331, Elsevier
22 Kortagere, S. and Ekins, S. (2010) Troubleshooting computational methods in drug discovery. *J. Pharmacol. Toxicol. Methods* 61, 67–75
23 Leach, A.R. *et al.* (2010) Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* 53, 539–558
24 Wolber, G. *et al.* (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* 13, 23–29
25 Lill, M.A. (2007) Multi-dimensional QSAR in drug discovery. *Drug Discov. Today* 12, 1013–1017
26 Lewis, R.A. (2005) A general method for exploiting QSAR models in lead optimization. *J. Med. Chem.* 48, 1638–1648
27 Guha, R. (2008) On the interpretation and interpretability of quantitative structure–activity relationship models. *J. Comput. Aided Mol. Des.* 22, 857–871
28 Bajorath, J. *et al.* (2009) Navigating structure–activity landscapes. *Drug Discov. Today* 14, 698–705
29 De Benedetti, P.G. (1992) Electrostatics in quantitative structure–activity relationship analysis. *J. Mol. Struct. Theochem* 256, 231–248
30 Puzyn, T. *et al.* (2008) Calculation of quantum-mechanical descriptors for QSPR at the DFT level: is it necessary? *J. Chem. Inf. Model.* 48, 1174–1180
31 De Benedetti, P.G. (1987) Structure–activity relationships and mechanism of action of antibacterial sulphanilamides and sulphones. In *Advances in Drug Research* (Testa, B., ed.), pp. 227–279, Academic Press
32 Bell, P.H. and Roblin, R.O.J. (1942) Studies in chemotherapy. VII: a theory of the relation of structure to activity of sulfanilamide type compounds. *J. Am. Chem. Soc.* 64, 2905–2917
33 De Benedetti, P.G. *et al.* (1981) Structure–activity relationships in dihydropteroate synthase inhibition by sulfanilamides: comparison with the antibacterial activity. *J. Med. Chem.* 24, 454–457
34 De Benedetti, P.G. and Frassineti, C. (1983) A theoretical study of the structure–activity relationship in diaryl sulphones, comparison with sulpha drugs. *J. Mol. Struct. Theochem* 92, 191–196
35 De Benedetti, P.G. *et al.* (1987) Quantitative structure–activity analysis in dihydropteroate synthase inhibition by sulfones: comparison with sulfanilamides. *J. Med. Chem.* 30, 459–464

36 De Benedetti, P.G. *et al.* (1989) Quantitative structure–activity relationships in dihydropteroate synthase inhibition by multisubstituted sulfones. Design and synthesis of some new derivatives with improved potency. *J. Med. Chem.* 32, 2396–2399
37 Achari, A. *et al.* (1997) Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nat. Struct. Biol.* 4, 490–497
38 Vinnicombe, H.G. and Derrick, J.P. (1999) Dihydropteroate synthase from *Streptococcus pneumoniae*: characterization of substrate binding order and sulfonamide inhibition. *Biochem. Biophys. Res. Commun.* 258, 752–757
39 Baca, A.M. *et al.* (2000) Crystal structure of Mycobacterium tuberculosis 7,8-dihydropteroate synthase in complex with pterin monophosphate: new insight into the enzymatic mechanism and sulfa-drug action. *J. Mol. Biol.* 302, 1193–1212
40 Giordanetto, F. *et al.* (2005) Large scale molecular dynamics simulation of native and mutant dihydropteroate synthase-sulphanilamide complexes suggests the molecular basis for dihydropteroate synthase drug resistance. *Philos. Trans. A: Math. Phys. Eng. Sci.* 363, 2055–2073
41 Koetzle, T.F. and Williams, G.J. (1976) The crystal and molecular structure of the antifolate drug trimethoprim (2,4-diamino-5-(3,4,5-trimethoxybenzyl)pyrimidine). A neutron diffraction study. *J. Am. Chem. Soc.* 98, 2074–2078
42 Hopfinger, A.J. (1981) Inhibition of dihydrofolate reductase: structure–activity correlations of 2,4-diamino-5-benzylpyrimidines based upon molecular shape analysis. *J. Med. Chem.* 24, 818–822
43 Andrews, P.R. *et al.* (1986) Conformational energy calculations and electrostatic potentials of dihydrofolate reductase ligands: relevance to mode of binding and species specificity. *J. Med. Chem.* 29, 698–708
44 Baccanari, D.P. *et al.* (1982) Inhibition of dihydrofolate reductase: effect of reduced nicotinamide adenine dinucleotide phosphate on the selectivity and affinity of diaminobenzylpyrimidines. *Biochemistry* 21, 5068–5075
45 Champness, J.N. *et al.* (1986) Crystallographic investigation of the cooperative interaction between trimethoprim, reduced cofactor and dihydrofolate reductase. *FEBS Lett.* 199, 61–67
46 Matthews, D.A. *et al.* (1985) Refined crystal structures of *Escherichia coli* and chicken liver dihydrofolate reductase containing bound trimethoprim. *J. Biol. Chem.* 260, 381–391
47 Krishnamurthy, V.M. *et al.* (2008) Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein–ligand binding. *Chem. Rev.* 108, 946–1051
48 De Benedetti, P.G. *et al.* (1987) A quantum chemical QSAR study of carbonic anhydrase inhibition by heterocyclic sulfonamides. *Quant. Struct. Act. Relat.* 6, 51–53
49 De Benedetti, P.G. *et al.* (1985) A quantum chemical QSAR study of carbonic anhydrase inhibition by sulfonamides. *Quant. Struct. Act. Relat.* 4, 23–28
50 Menziani, M.C. and De Benedetti, P.G. (1992) Molecular mechanics and quantum chemical QSAR analysis in carbonic anhydrase-heterocyclic sulfonamide interactions. *Struct. Chem.* 3, 215–219
51 Menziani, M.C. *et al.* (1989) The binding of benzenesulfonamides to carbonic anhydrase enzyme. A molecular mechanics study and quantitative structure–activity relationships. *J. Med. Chem.* 32, 951–956
52 Arinaminpathy, Y. *et al.* (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov. Today* 14, 1130–1135
53 Kortagere, S. *et al.* (2009) The importance of discerning shape in molecular pharmacology. *Trends Pharmacol. Sci.* 30, 138–147
54 Nicholls, A. *et al.* (2010) Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.* 53, 3862–3886
55 Cocchi, M. and De Benedetti, P.G. (1998) Use of the supermolecule approach to derive molecular similarity descriptors for QSAR analysis. *J. Mol. Model.* 4, 113–131
56 Barlocco, D. *et al.* (2001) Phenylpiperazinylalkylamino substituted pyridazinones as potent alpha(1) adrenoceptor antagonists. *J. Med. Chem.* 44, 2403–2410
57 Fanelli, F. and De Benedetti, P.G. (2008) Computational modeling of selective pharmacophores at the alpha1-adrenergic receptors. In *Antitargets* (Vaz, R.J. and Klabunde, T., eds), pp. 155–193, Wiley–VCH
58 Cherezov, V. *et al.* (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318, 1258–1265
59 Fanelli, F. *et al.* (1993) Theoretical quantitative structure–activity analysis of quinuclidine-based muscarinic cholinergic receptor ligands. *J. Mol. Struct. Theochem* 283, 63–71
60 Cappelli, A. *et al.* (2004) Design, synthesis, structural studies, biological evaluation, and computational simulations of novel potent AT(1) angiotensin II receptor antagonists based on the 4-phenylquinoline structure. *J. Med. Chem.* 47, 2574–2586
61 Ohtaka, H. and Freire, E. (2005) Adaptive inhibitors of the HIV-1 protease. *Prog. Biophys. Mol. Biol.* 88, 193–208
62 Tummino, P.J. and Copeland, R.A. (2008) Residence time of receptor–ligand complexes and its effect on biological function. *Biochemistry* 47, 5481–5492
63 Jenwitheesuk, E. *et al.* (2008) Novel paradigms for drug discovery: computational multitarget screening. *Trends Pharmacol. Sci.* 29, 62–71
64 Smith, A.J. *et al.* (2009) Beyond picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J. Med. Chem.* 52, 225–233