Chemical Similarity and Biological Activities

Hugo Kubinyi

University of Heidelberg, Germany and BASF AG, Ludwigshafen, Germany

“Similar molecules exert similar biological activities”. Since long, medicinal chemists use this concept to modify the structures of biologically active compounds. Bioisosteric replacements of atoms and groups pave the way from various lead structures to therapeutically useful enzyme inhibitors, receptor agonists and antagonists, and other active principles. However, similarity and diversity of chemical structures cannot be defined in an objective manner. Several surprising structure-activity relationships demonstrate that chemically similar compounds may have significantly different biological actions and activities. Some protein ligands exert unexpected new binding modes, after only minor chemical modification. Of course, even optical enantiomers most often have different biological activities.

Keywords: chemical similarity, similarity vs. activity, structure-activity relationships, agonists and antagonists, hydrogen bonds, optical enantiomers

1. Similarity and Dissimilarity of Molecules

“Similar molecules exert similar biological activities”. Since long, medicinal chemists used this concept to modify the structures of biologically active compounds. Bioisosteric replacements of atoms and groups pave the way from various lead structures to therapeutically useful enzyme inhibitors, receptor agonists and antagonists, and other active principles. Several surprising structure-activity relationships demonstrate that chemically similar compounds may have significantly different biological actions and activities. Of course, even optical enantiomers most often have different biological activities. Correspondingly, a sophisticated consideration of “chemical similarity” and “chemical diversity” is a waste of time. Similarity and diversity of molecules depend on the 3D structure and binding site properties of the biological target, not on any artificial parameters.

The principle of bioisosteric replacement of functional groups serves as a successful optimization strategy. Its systematic application has resulted in a broad variety of therapeutically used drugs, many of them finally having the desired combination of favorable properties. In their attempts to optimize lead structures, medicinal chemists intuitively follow the principles of evolution. The biological activity, in later stages a selectivity index or some other biological property, serve as the “fitness function” for the “survival” of certain structural entities. However, several surprising structure-activity relationships demonstrate how difficult rational drug design can be and to which extent structure-based and computer-aided approaches can support this process.

Similarity and diversity of chemical structures cannot be defined in an objective manner (Figure 1). There is no doubt that for maximum affinity a ligand of a biological macromolecule has to fit the binding pocket geometrically (Figure 2).

Also hydrophobic surfaces of the ligand and the binding site have to be complementary. However, the functional
groups of the ligand need a separate consideration. For lipophilicity, there is no significant difference between -O- and -NH- in an organic molecule; for ionization, there is a big difference whether a nitrogen atom is part of a basic group (an amine) or a neutral group (e.g., an amide); for binding, potency differences of several orders of magnitude may result from the exchange of the hydrogen bond donor function -NH- against an acceptor -O- (Figure 3).26-28

On the other hand, for recognition by the binding site it does not matter whether a certain acceptor atom of the ligand is either nitrogen or oxygen. An illustrative example for this equivalence are two series of scytalone dehydratase inhibitors, salicylamides and quinazolines, which exert the same type of interactions with the protein. Also the important role of water can be seen: several conserved water molecules mediate the contact between the ligand and the enzyme. In many cases, the replacement of such water molecules reduces binding affinity but a significant enhancement of affinity is observed for the scytalone dehydratase inhibitors, if one water molecule is replaced by a cyano substituent - its nitrogen atom mimics the oxygen atom of the water molecule (Figure 4).29

Figure 1. Bioisosteric replacement of atoms or groups can a) retain biological activity, b) significantly enhance biological activity, c) destroy biological activity, d) turn a substrate into an inhibitor, or e) modify selectivity. a) Thyroxin, \( T_4 \) (\( X = Y = Z = I \)), is the less active depot form of the natural hormone, \( T_3 \) (\( X = Y = Z = I \)); a monoisopropyl analog (\( X = I, Y = \text{isopropyl} \)) of \( T_3 \) is slightly more active than the natural hormone; even an analog without any halogen atoms (\( X = E, Y = \text{isopropyl} \)) has some thyromimetic activity.22 b) Trifluoromethanesulfonamide is 50,000 times more active as a carboanhydrase inhibitor than the methanesulfonamide.23 c) An exchange of the ester oxygen atom in acetylsalicylic acids to -CH₂- or -NH- destroys biological activity because these analogs cannot any longer transfer an acetyl residue to cyclooxygenase. d) Sulfanilamide is an antimetabolite of p-aminobenzoic acid; it acts as a false substrate and therefore inhibits bacterial dihydrofolate biosynthesis. e) GABA (\( \gamma \)-aminobutyric acid) binds with equal affinities to \( \Gamma \) and \( \Gamma \) receptors; its analog \( \gamma \)-aminopropylphosphonous acid is a highly selective \( \Gamma \) receptor ligand.24

Figure 2. Valine tRNA synthase selectively recognizes valine because the isoleucine side chain is too large for the binding pocket. The situation is more complex with isoleucine tRNA synthase: a "suspicious", energy-consuming check by isoleucine tRNA synthase rejects valines (1:200,000) but also 80% of all isoleucines; correspondingly, the error rate is about 1:40,000.25

2. Different Mechanisms of Action of Similar Molecules

Several well-known examples of different modes of action of closely related analogs can be found in medicinal
Chemical Similarity and Biological Activities

Norepinephrine, epinephrine and isoproterenol are adrenergic agonists. However, in going from R = H to R = CH₃ and R = isopropyl, the mechanism of action gradually changes from a more or less specific \(\alpha\)-adrenergic agonism to a pure \(\beta\)-adrenergic agonism. If the two hydroxyl groups of isoproterenol are exchanged against two chlorine atoms, the \(\beta\)-adrenergic antagonist dichloroisoproterenol (DCI) results; in fact, this compound was the first \(\beta\)-blocker. Some other receptor antagonists, e.g., the histamine, dopamine and morphine receptor antagonists, are lipophilic analogs of the corresponding agonists (Figure 5). Several peptide receptor antagonists differ from the corresponding agonists only by minor changes in their alkyl groups (Figure 6).

A well-known example of different modes of action of closely related analogs are the anti-allergic agent promethazine, the neuroleptic drug chlorpromazine, and the antidepressants imipramine and desipramine. Despite their very similar chemical structures, promethazine acts mainly as a histamine \(H₁\) antagonist, chlorpromazine is a dopamine antagonist, imipramine is an unspecific norepinephrine and serotonin uptake inhibitor, and desipramine is a norepinephrine-specific uptake inhibitor (Figure 7). Steroid hormones, like the estrogens, the gestagens, the male sex hormones and anabolics, provide another striking example of different biological effects of chemically closely related analogs (Figure 8).

Several therapeutically used drugs bind to more than one receptor and are correspondingly termed “promiscuous” ligands. However, whether a certain (balanced) unspecific mode of action is advantageous for therapy or not remains still uncertain.
3. The Biological Activity of a Ligand Depends on its Flexibility

Other common structural modifications in the optimization of a lead structure are the dissection of rings or the rigidification of flexible molecules. Molecules with several rotatable bonds may adopt many different geometries, some of them being favorable because of low internal energies, others being less favorable because of van der Waals or electrostatic repulsion between nonbonded atoms or groups. If different conformations of such molecules are “frozen” by closing rings between certain atoms, either one of two very different consequences results. If the frozen conformation differs from the bioactive conformation of the flexible lead or if the added atoms interfere with the binding, biological activity will be more or less destroyed. If the ring closure stabilizes the bioactive conformation, usually a significant increase in biological activity results.

The computer program CAVeAT was developed for the design of rigid analogs which bear a pharmacophore in a certain geometry. CAVeAT starts from a structural hypothesis or from the known 3D structure of a ligand and extracts vectors of residues that participate in binding. In a peptide, these vectors are e.g. the Cα-Cβ bonds of the interacting amino acid side chains. Then the program identifies ring systems that are suited to accommodate these residues in exactly the same relative geometry (Figure 9). An example of rigid analogs with remarkable selectivity differences are chemically closely related integrin ligands (Figure 10). For compounds with comparable threedimensional structures, most often similar analogs have also similar biological activities. That this needs not always be the case can be demonstrated by a comparison of three series of thermolysin inhibitors. Analogs with X = -NH- and -CH2- are about 1,000-times more potent than the X = -O- analogs. The explanation for this effect can be easily derived from the 3D structure of thermolysin. If X is an -NH- group, a hydrogen bond is formed between this group and the oxygen atom of an alanine carbonyl group. In the -O- analog, this hydrogen bond cannot be formed; in addition, an electrostatic repulsion between the two oxygen atoms results. The biological activity of the -CH2- analog has been predicted to be comparable to the -NH- analog and to be much higher than for the -O- analog. This was later confirmed by the synthesis of these inhibitors (Figure 11; cf. Figure 3).

4. Biological Potencies of Similar Molecules

Figure 7. Promethazine, chlorpromazine, imipramine and desipramine are chemically closely related tricyclic drugs. Despite this analogy, promethazine is an H1 antagonist, acting as an antiallergic drug, chlorpromazine is a dopamine antagonist for the treatment of schizophrenia, and imipramine and desipramine are neurotransmitter uptake inhibitors that are used in the therapy of depression.

Figure 8. Steroid hormones are another example of chemically related compounds with strikingly different actions. These orally active, synthetic analogs of natural sex hormones are estrogens, gestagens (the female sex hormones), androgens (male sex hormone) and anabolics.

Figure 9. The concept of the program CAVeAT is based on the observation that only three distinct amino acids of serine proteinase inhibitors interact with the binding site of the enzyme. Correspondingly, the rest of the large molecule serves only for the stabilization of a certain conformation of the interacting loop. The Cα-Cβ bonds of the critical amino acids are stored as vectors in space to search ring systems which bear these vectors in exactly the same geometry. Then a peptidomimetic is constructed by attaching the side chains to this system. In further steps, the side chains can be modified in an appropriate manner.
A similar but less pronounced effect is observed for thrombin inhibitors; in this case the nonbonded contact between the -X- group of the ligand and the carbonyl group of Gly 216 in the binding site of thrombin is responsible for the structure-activity relationship. If, on the other hand, a carbonyl group of the inhibitors, which forms a hydrogen bond with the -X- group of Gly 216, is replaced by -CH₂-, affinity is reduced by some orders of magnitude; only in one case a small reduction of affinity is observed (Figure 12). 48-51

There are many examples in literature where the introduction of an -OH group into a ligand either causes an increase or a decrease of biological affinities. From a theoretical point of view, this is not surprising. If the new -OH group forms hydrogen bonds with polar groups at the binding site (either as a hydrogen bond donor or as an acceptor), the net free energy depends on the balance between the desolvation energies of the water shells at the surfaces of the ligand and the binding site, as compared to the energy of the formed hydrogen bond (or bonds) and the entropy gain by the release of some water molecules. Certain tightly bound water molecules in the binding cavity of a protein (usually seen in the X-ray structures), e.g., those which form more than two hydrogen bonds to the protein, are not easily removed. The attempt to introduce an -OH group into the ligand, to replace such a water molecule, must necessarily fail.

An example where this is not the case and where significantly enhanced binding affinities result after the introduction of such a hydroxyl group, are cytidine and adenosine deaminase inhibitors, which are capable to add a water molecule in the binding site, to mimic a transition state (Figure 13). 52,53 In this special case the resulting affinity differences are 7 to 8 orders of magnitude!

5. Biological Potencies vs. Similar Biological Targets

Functionally corresponding proteins from different species have identical or very similar 3D structures, but they normally differ in their amino acid composition. Although they always show, dependent on the evolutionary relationship between the two species, a certain degree of homology, structure-activity relationships may significantly differ, even after the replacement of just one
single amino acid by another one (Figures 14 and 15). In the past, potential drug candidates have always been tested in animals before they could be applied to humans. Human proteins were not available, except in rare cases. Only for those proteins that could be extracted from human material, e.g. hemoglobin or thrombin from blood, could one be sure about the therapeutic potential of a new drug from in vitro studies, prior to studies in man. With the

**Figure 12.** Typical thrombin inhibitors, a) X = NH, form β-sheet type hydrogen bonds with Gly-216. If X = -NH- is exchanged against -O- or -CH2-, a structure-activity relationship is observed that is similar but less pronounced as in the thermolysin inhibitors (Figure 11). b) An even smaller effect is observed, if the amino group is removed. Reduction of the interacting carbonyl group to a -CH2- group reduces affinities by factors of about 400 (c), 2,000 (d), 10,000 (e), and 4 (f). The small activity difference in the latter compound pair can be explained by the fact that, even after reduction, the attached nitrogen atom remains neutral whereas it becomes a basic nitrogen in the other cases.48-51

**Figure 13.** The natural product zebularin is a highly potent inhibitor of cytidine deaminase because it perfectly mimics the transition state of the enzymatic reaction after addition of a conserved water molecule within the binding site. 3,4-Dihydrozebularine can neither add nor replace this water molecule. Correspondingly, this compound is much less active; two hydrogen atoms make an activity difference of more than seven orders of magnitude.52,53

**Figure 14.** Remikiren, a renin inhibitor, shows different potency against the renins of rats and dogs, as compared to monkeys and humans. Such differences are caused by minor amino acid sequence differences and the resulting different binding site geometries of these homologous renins.54,55

**Figure 15.** Binding affinities of various ligands to rat, human wild type (filled circles) and human Thr355Asn mutant 5-HT1B receptors (open circles). If the amino acid Thr-355 of the human 5-HT1B receptor is replaced by Asn, the corresponding amino acid of the rat receptor, the observed binding affinities are significantly different. One amino acid makes the difference - it converts the binding characteristics of the human wild-type receptor into the binding characteristics of the rat receptor. Whereas there is no correlation in binding affinities between the rat and human wildtype receptors, the Thr355Asn mutant yields a correlation coefficient r = 0.98.56
progress in gene technology it is now possible to produce human proteins in sufficient quantities to establish test models. Thus, their biological activity in humans can be forecasted from investigations at the molecular level.

Thiorphan and its retro-inverso peptide, retro-thiorphan, are inhibitors of the structurally related zinc proteases thermolysin and NEP 24.11. Although the affinities of the ligand pair differ, from enzyme to enzyme, by three orders of magnitude, there are no significant activity differences between them. Thus, they may be considered to be “similar”, which was also confirmed by the X-ray structure analyses of their thermolysin complexes. On the other hand, their activities against yet another related zinc protease, angiotensin converting enzyme (ACE), are significantly different; with respect to this enzyme the analogs are “dissimilar” (Figure 16).

Corresponding problems are also observed in predictions for toxicity in humans. 2,3,7,8-Tetrachlorodioxin is highly toxic to several species, e.g. guinea pigs and mink. It is much less toxic for mice, rats, hamsters, rabbits, dogs, and monkeys. If one extrapolates from the monkey, dioxin should be relatively “harmless” for humans, at least if only acute toxicity is considered. However, the significantly different toxicity vs. the closely related species guinea pig and hamster puts a caveat on too simple and straightforward extrapolations (Figure 17).

6. Chirality and Biological Activities

Due to the chiral nature of amino acids (except glycine), drug binding sites of proteins are asymmetric. In the past, the different actions of enantiomers of chiral molecules on enzymes and receptors were often neglected. For economic reasons, racemates of synthetic drugs were used in therapy. Today, researchers and drug companies are more aware of the different effects of enantiomers and diastereomers, in their biological activities (Figure 18) as well in their pharmacokinetics.

Figure 16. Thiorphan and retro-thiorphan differ in the direction of the amide group of both molecules. Nevertheless, they are equally potent against the zinc protease thermolysin. X-ray structure determination of the complexes confirms that both molecules display equivalent interactions. Although being much more potent, they also show identical activities against another zinc protease, NEP 24.11 (originally called enkephalinase). However, their activities differ significantly against a third zinc protease, the angiotensin-converting enzyme (ACE), which gives evidence that chemical and biological similarities cannot be defined in an objective manner - they depend on the structure of the biological target.

Figure 17. 2,3,7,8-Tetrachlorodioxin (the so-called “Seveso dioxin”) shows significantly different acute toxicities in different species. The different toxicity against evolutionary related species (e.g. hamster vs. guinea pig) makes it impossible to extrapolate to humans without further investigations, e.g. of different receptor structures or metabolic pathways.

Figure 18. Butaclamol is just one example that the eudismic ratio, the ratio of affinities of (+)- and (-)-enantiomers, differs from receptor to receptor, for the same compound. The (S)-(−) form of the calcium channel ligand Bay K 8644 is an agonist (stabilizing the open calcium channel), whereas the (R)-(+) form is a weak antagonist, a calcium channel blocker (stabilizing the closed channel). Corresponding differences are observed for a CCK, ligand, where one diastereomer is an agonist, whereas its enantiomer is an antagonist.
Sensory receptors are G-protein coupled receptors that differentiate between enantiomers, like all other receptors. Correspondingly, enantiomers can even be recognized by their characteristic odor, e.g. the monoterpenes (R)- and (S)-limonene and (R)- and (S)-carvone or by their odor intensity, e.g. some diastereomeric wine lactones (Figure 19).

7. Conclusions

An overreliance in target-independent similarity indices has to be questioned, because of the dependence of “similarity” on the biological macromolecule to which the analogs bind. Sophisticated investigations on the “dissimilarity” of chemical databases are most often futile. Similar compounds may have very different actions and different molecules can be very similar in their biological activities. Considering the examples presented in this paper and the many more cases in literature, one has to admit that we are far from a deeper understanding of the details which underlie the observed structure-activity relationships. Applying the results from one series of analogs to another, one may arrive at completely wrong conclusions.

Acknowledgements

This publication reviews the content of two lectures, presented at the 3rd Workshop on Chemical Structure and Biological Activity: Perspectives in QSAR, Instituto de Química (IQ) and Faculdade de Ciências Farmacêuticas (FCF), Universidade de São Paulo (USP), SP, Brazil, November 08-10, 2001, and at the 1st Simpósio Brasileiro em Química Medicinal, QSAR e Modelagem Molecular, Caxambú, MG, Brazil, November 11-16, 2001. I am grateful to Prof. Dr. Antonia Tavares do Amaral, IQ, USP, São Paulo, Brazil, and Prof. Dr. Carlos Alberto Montanari, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil, for their invitation and support to participate in these symposia. The financial contributions of Deutscher Akademischer Austauschdienst (DAAD), Bonn, Germany, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), São Paulo, Brazil, and the congress organisations in São Paulo and Caxambu, Brazil, are gratefully acknowledged.

References

17. Borchardt, R. T.; Freidinger, R. M.; Sawyer, T. K.; Smith, P. L.