

Drug design for ever, from hype to hope

G. Seddon · V. Lounnas · R. McGuire ·
T. van den Bergh · R. P. Bywater · L. Oliveira ·
G. Vriend

Received: 22 November 2011 / Accepted: 5 December 2011 / Published online: 18 January 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract In its first 25 years JCAMD has been disseminating a large number of techniques aimed at finding better medicines faster. These include genetic algorithms, COM-FA, QSAR, structure based techniques, homology modelling, high throughput screening, combichem, and dozens more that were a hype in their time and that now are just a useful addition to the drug-designers toolbox. Despite massive efforts throughout academic and industrial drug design research departments, the number of FDA-approved new molecular entities per year stagnates, and the pharmaceutical industry is reorganising accordingly. The recent spate of industrial consolidations and the concomitant move towards outsourcing of research activities requires better integration

of all activities along the chain from bench to bedside. The next 25 years will undoubtedly show a series of translational science activities that are aimed at a better communication between all parties involved, from quantum chemistry to bedside and from academia to industry. This will above all include understanding the underlying biological problem and optimal use of all available data.

Keywords Drug design · Protein modeling · QSAR · G-protein coupled receptors · Translational research · Review

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9519-9) contains supplementary material, which is available to authorized users.

G. Seddon
Adelard Institute, Manchester, UK

V. Lounnas · G. Vriend (✉)
CMBI, Radboud University Nijmegen Medical Centre, Geert
Grooteplein 26–28, 6525 GA Nijmegen, The Netherlands
e-mail: vriend@cmbi.ru.nl

R. McGuire
BioAxis Research, Bergse Heihoek 56, Berghem 5351 SL,
The Netherlands

T. van den Bergh
Bio-Product, Dreijenplein 10, 6703 HB Wageningen,
The Netherlands

R. P. Bywater
Magdalen College, Oxford, UK

L. Oliveira
Sao Paulo Federal University (UNIFESP), Sao Paulo, Brazil

Introduction

Life expectancy of man, and especially man in the western world, increased by more than 2 days per week for the whole previous century [1]. Much of this dramatic increase is to the credit of hygiene, but medicines, and especially antibiotics and vaccines, have contributed significantly too. In the first world war, for example, almost as many soldiers died of disease as of bullets [2]. During the second world war this unfortunate situation got 'remedied' by the introduction of sulphonamides and penicillin [3].

At this moment medical doctors around the world can write prescriptions for tens of thousands of medicines [4], and an even larger number is available of herbal medicines, homeopathic wonder-cures, and other preparations for which the medicinal value has not been proven [5]. Most medicines function by interacting with proteins in the body. Of the more than twenty thousand protein types in our body less than five hundred are targeted by all these medicines [6]. This, of course, gives hope for the future of drug design because most proteins are still available as a target for which a blockbuster drug can be designed.

Despite massive, world-wide efforts the number of new molecular entities (NMEs) that the FDA approves per year for use as medicines certainly isn't growing [7], while the amount of money involved goes up much faster than inflation [7] even when we include Obama's Troubled Asset Relief Program [8].

This journal (JCAMD) has published many, many articles on techniques that according to the authors of these articles were the holy grail for drug design, and that in today's reality are just good tools used in this process. Following a path familiar in science, someone has a good idea, gives it a name and publishes it. Others follow suit and publish improvement after improvement, after which yet others start testing all similar methods. An example is the use of support vector machines for ligand selection. This was introduced in 2000 [10] and only 3 years of improvements were needed before the first comparison methods were published [11]. Figure 1 illustrates the desperation of pharmaceutical industries. The ever increasing costs mainly result from development and marketing [12] and, unfortunately for us, not from research. This might explain why each time a new drug design research tool gets published pharmaceutical industries immediately jump on it and give it a hype status.

A brief history of tools

The first hype in drug design was born out of the famous article by Hol [13] in which he coined the name 'rational drug design' for all protein-structure based techniques, thereby implicitly calling all methods that actually worked, such as screening or luck, irrational; see Fig. 2.

It is not by eye that we can determine either the fitness of a ligand for a pocket, or the safety or efficacy of a drug. It does not seem illogical to assume that the founders of JCAMD were at least subconsciously dealing with the

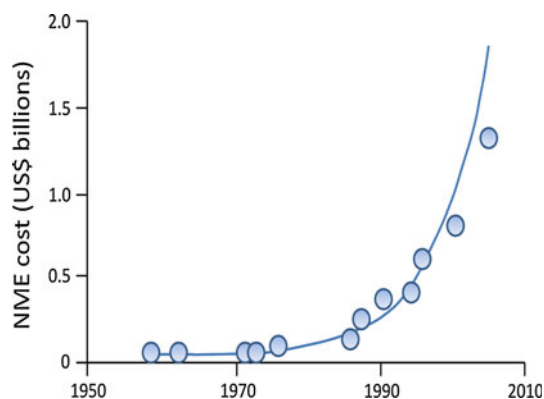


Fig. 1 Amount of money spent in billion US\$/NME (after Munos [7]). Munos summarises these numbers eloquently in his 2009 review, but you are also encouraged to read the commentary by Firestone [9]

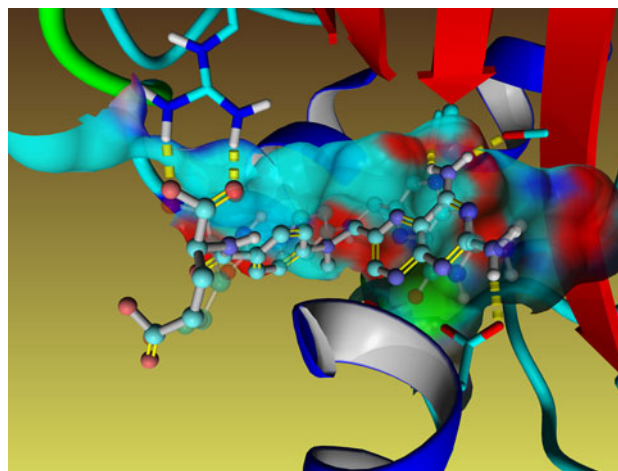


Fig. 2 Methotrexate in the active site pocket of dihydrofolate reductase (PDBid:4DFR [14]). Next to a monochrome picture showing this same fit, Hol wrote: "As to whether a drug can actually reach its target, e.g. the active center of an enzyme, is primarily a spatial problem. Assuming that the structures of both components are known, computer graphics can help in checking the suitability of a potentially active substance. As example, the structure of the complex formed between a bacterial dihydrofolate-reductase, NADP and the anticancer drug methotrexate (gray dots) is shown on the right. As one can see, it fits" [13]

oversimplification implied by Fig. 2 when they started this journal. And we believe that most articles published in JCAMD have dealt with aspects of drug design 'left out' of Fig. 2. The advent of faster computers like first the VAX/VMS, then supercomputers such as the CRAY, and finally the PC, have allowed scientist to numerically solve chemical problems of ever increasing size and complexity. Semi-empirical quantum calculation methods have been devised to calculate the chemically relevant aspects of the electronic wave-functions associated with small organic molecules and thus compute their 3D dimensional structures as well as the energy of their conformers [15–19]. All the techniques derived in this domain are referred to as ligand-based drug design. In parallel, the development of molecular mechanics force fields combined with the fact that Newton's equations of motion could be solved for entire proteins in their aqueous environments were true innovations in the investigation of the structure function relationships [20–28]. Thus, not only the geometry and the potential energy surface of macromolecular assemblies could be calculated but also their dynamic and thermodynamic properties [29, 30]. For the early computational chemists this opened the perspective of testing at will the energy of interactions between protein targets and large collections of small molecule ligands [29, 31, 32]. The original thoughts that this would replace experimental validation processes, though, have long been shown to be a nice dream at best. The perception that the underlying

mechanism of protein–ligand recognition would be unravelled and would thus allow what ever since has been called structure-based drug design has never looked so clear and promising as at that particular moment in 1986.

With the exception of a very small fraction of ligands that are purely rigid, most bioactive ligands have a number of rotatable bonds that make them flexible. The values of the torsion angles in ligands are determined by the valence electrons of the atoms. The development of empirical molecular mechanics force field in the late 1970s [33, 34] have allowed for the *in silico* determination of the geometries (low energy conformers) of ligands *in vacuo*. Application of these methods relies on two underlying assumptions: (1) that the conformation of the dissolved ligand corresponds closely to its gas-phase conformation [35]; and (2) that the biologically active conformation of the ligand is likely to be found among the set of low energy conformers of the isolated ligand [36, 37]. The combined knowledge of the ligand structure (determined by NMR or X-ray), the measured binding affinities, and the spatial overlay of the low energy conformations should then be sufficient to establish a structure activity relationship [38] and pinpoint the spatial organization of the recurrent chemical features correlated with activity (pharmacophore). This paved the way for a series of successes for ligand-based drug design [e.g. 39, 40, 16]. However, although it seems fairly reasonable at first sight, both assumptions in practice proved to be incomplete and/or insufficient [41–50].

The computational process by which the complementary aspects between a ligand and a receptor binding site can be ascertained has been explored with the design of specifically dedicated docking programmes. Early docking methods were based uniquely on assessing the shape complementarity [51] between a pocket in the 3D structure of a protein and low energy conformers of a ligand. The approach was computationally cumbersome due to the need to systematically search all possible ligand orientations within the pocket and scoring each of these poses by its steric hindrance. Subsequent developments have taken place in several directions: improved scoring functions [52–63] different ways to deal with ligand flexibility [60, 64–72], and most recently also ways to deal with receptor flexibility [73–78]. Fundamental research has been performed into directions such as desolvation energies [79–83], or other aspects of the force fields used for scoring docking poses [66, 78, 84–96].

The idea to calculate from first principles all atomic motions occurring in an active enzyme in its aqueous environment has attracted many scientists to computer aided molecular design. Starting with the atomic loci obtained from the X-ray structure of an enzyme it can be envisaged to integrate Newton's equations [29, 31]. A

series of snapshots describing the trajectory of the enzyme over time could thus be produced and ensemble average properties calculated based on Boltzmann's ergodic hypothesis. The near infinite computer time needed for such experiments muted this field till concepts from alchemy could be embraced. *In silico*, one is not bound by the sequential order of events that govern paths between states, and hence so-called thermodynamic cycle methods could be developed that replaced chemical steps with alchemical steps that in principle should lead to the same outcome [29, 30, 97, 98].

Comparative Molecular Field Analysis (CoMFA) is based on the overlay of active ligands [99, 100–102]. Initially, the technique was more a concept than an effective tool as computer power was very limited and molecular descriptors as well as dedicated algorithms needed to be developed [103]. The underlying idea that the 3D dimensional steric/non-bonded (Van der Waals) and electrostatic potential fields generated by the spatial organization of the chemical features around the scaffold of a ligand (Fig. 3) play a fundamental role in the biomolecular receptor recognition was so intuitively right and the technique made a break-through in 1988 [99]. Examples of the application of the method are plentiful [100]. About 15% of all articles in JCAMD refer to the use of this technique, refined and applied in all sorts of ways to produce the overly famous quantitative structure activity relationship (QSAR) equations. However, CoMFA suffers from three drawbacks: (1) the alignment of the ligands in the pocket must be either known or gambled correctly; (2) the method has been established for rigid or quasi-rigid classes of molecules (e.g. steroids); and (3) the detailed influence of the protein pocket is not known which means that any feature that is not implicitly present in the training set will be missed [104–109]. These nearly fatal drawbacks prevented the generalization of the method as a standalone solution to rational drug design. Certainly, the best way to apply CoMFA is to combine it with a pharmacophore model and a carefully conducted conformational study of the ligands [110].

Many drug design projects include at some stage knowledge of the 3D structure of the target protein, and homology modelling is normally used when neither X-ray nor NMR derived coordinates are available [111–118]. Many computer programs were written for this purpose [119–123] and the CASP competition [124] illustrates every 2 years where the field stands. Presently, YASARA seems to be performing very well [122], but many labs are working hard so this situation might change again in the future. For example, methods are under development that use PLIM [125] to provide a first fix on the ligand docking site where-after steered Molecular Dynamics is used to continue the trajectory to convergence.

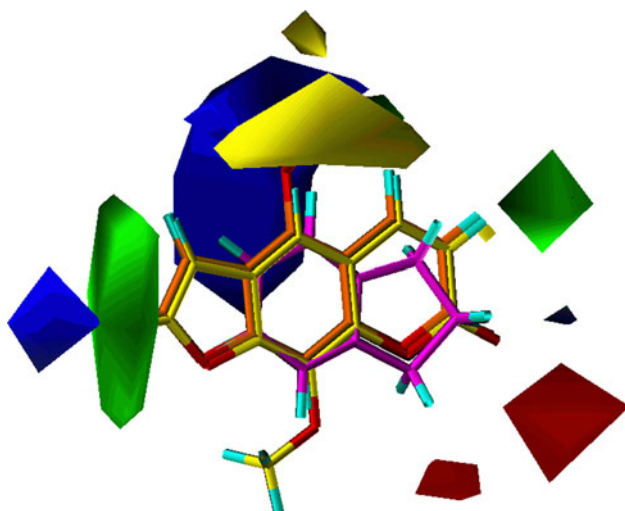


Fig. 3 Contour representation of key features from a CoMFA analysis of a series of coumarin substrates and inhibitors of cytochrome P4502A5 [Poso et al, adapted from the publicly available UCLA Chemistry 125 course]. The red and blue regions indicate positions where it would be favourable, respectively unfavourable to place a negative charge and the green/yellow regions where it would be favourable/unfavourable to locate bulky groups

Similar to the CASP competitions, the GPCR-DOCK [126, 127] competitions have evaluated the quality of docking software, but with the additional complexity that the target structures needed to be modelled before docking could be attempted. In recent years a whole series of studies have been published in which homology modelling, combined with other tools, proved a viable replacement for the cumbersome experimental determination of target structures [111, 114, 128, 129]. The good performance of two Dutch teams [130] in the recent GPCR Dock competition [127] beautifully illustrates the often mentioned fact that even the best tools only perform well in the hands of good scientists [131]. In this latter article we find the interesting quote “Interestingly enough, it is the model built with most human intervention which proves to be the best”.

In the early 1990s the radical new idea emerged that instead of the virtual and/or real screening of large libraries of already existing molecules to identify new bioactive hits, one could rather attempt to construct entirely new synthesizable molecular entities solely based on the knowledge of the active site of the pharmaceutical target enzyme. [132–134]. To do so, small organic fragments composed of few atoms only must be assembled *in silico* inside the binding sites of enzymes in such a way that optimal protein–ligand, steric, and electronic complementarity is achieved [84, 125, 135–141]. The major problem of this approach arises from the complexity of the active site landscape and the combinatorial vastness of all

possible arrangements of fragments in the volume delineated by an enzyme active site [66, 142–144]. How to choose the first fragment and where should it be positioned and oriented with respect to the inner surface of the binding pocket or cleft [143, 145, 146]? Which next fragment should be attached to it? [147]? The genetic algorithms [66, 142, 148–150] have been invented which allow this concept to be realized within a tractable amount of computer time by performing random transformations on a ligand collection. These transformations are selection, mutation, and crossover, and are reminiscent of the corresponding evolutionary processes in biology underlying the optimisation of genes, hence their name ‘genetic algorithms’. Experience shows that these algorithms provide solutions that nicely fit the objective function, although it often is difficult to understand exactly why [66].

Randomly screening very large libraries containing up to 10^5 or even 10^6 chemical entities in *in vitro* enzymatic assays to produce leads has been the central paradigm of the pharmaceutical industry across the 1990s. However, after years of operating very expensive screening facilities, it has been realized that the hits produced were not of the expected quality. For example, often a bias is observed toward too lipophilic compounds that are impossible to optimize. Compared to the actual number of chemically entities (\sim infinite) the any amount of compounds that can be screened via this process is essentially zero [151, 152]. In parallel, computational chemists had inferred that screening could be successfully operated virtually throughout computers at all stages in the drug design process from hit identification via hit optimization to lead optimization [153–162]. In each of these three stages virtual libraries can be created and filtered either using chemometrics to exclude molecules that obviously aren’t drug-like because of their predicted solubility or ADME/Tox properties, or using 3D chemical molecular descriptors (pharmacophores), or using docking results. Thus, libraries of compound that do not actually exist can be screened and a much smaller, manageable number of compounds selected. This is of particular advantage at the stage of lead optimization, when only few compounds are left. Scaffolds of lead compounds usually carry a number of branching points where chemical variation is allowed. The *in silico* creation of combinatorial libraries of all the variant compounds is a dramatically faster process than its *in vitro* counterpart [163–165].

One of the main difficulties in establishing reliable and/or transferrable QSAR equations is that, even within a class of chemical analogues, ligand affinities may not respond linearly to the variation of one or several of the molecular descriptors that have been identified as related to activity. For instance, across a series of chemical substituents sorted by increasing polarity the measured affinity may respond

linearly only for a restricted number of them because steric hindrance or global effect such as desolvation may penalize the binding of slightly larger groups. The modification of a branched group at another point around the scaffold may however allow some of the previously excluded ligands to become highly active. Indeed the mere addition of one methyl group may result in a sudden tenfold leap in potency, dramatically increasing ligand efficiency [166, 167]. It was demonstrated that these problems could be circumvented using artificial intelligence methods (neural network, support vector machine, etc.) that are insensitive to the spatial alignment of the ligand scaffolds and that are able to recognize particular combinations of properties distributed around the scaffold of a set of active ligands [168–170]. Artificial intelligence can be ubiquitously implemented at various stages in the rational drug design process to improve results that can be otherwise be more uncertainly obtained with classical methods, especially when assessing general properties that are the result of the subtle combination of many different factors in relation to others such as drug-likeness [171]. Various examples of artificial intelligence applications and their limitations have been published in JCAMD [172–175]. Notwithstanding the utility of artificial intelligence, normal intelligence remains useful in avoiding some of the all too common pitfalls in the derivation and application of QSAR models [176].

We apologise to the many authors of methods that didn't make it into the above list (see ESM Table 1). Much good work has been done that the editors certainly wouldn't allow us to include because citing all 1,200 articles published in JCAMD in the first 25 years would perhaps be a bit excessive. We could have mentioned the work by Che on privilege structures [177], or by Lotta et al. [178] on multivariate PLS modelling of data sets. The recent work by Zhou et al. [179] on the use of DFT calculation to accurately assess the existence of intermolecular H-bonds in docking instances. Sarmah et al.'s [180] work on solvent effects also added significantly to the drug-designers toolbox, but the methods described in these articles didn't achieve hype status.

Where do we stand today?

The rapid increase in costs of developing and marketing new medicines is not leaving the pharmaceutical industry untouched. Recent years have seen a strong concentration of activities in terms of mergers, buy-outs, and closures [7]. It may simply be, that a research-intensive industry like the pharmaceutical industry does not lend itself to the type of management that is common in consumer goods, fashion and footwear. It seems a paradox, though, that the high costs associated with drug design are caused by development, marketing, and legal fees, but when it comes to cost-

reduction research departments are, euphemistically called, consolidated. The past years have also seen a consolidation of methods. JCAMD has published a large series of articles in which multiple methods have been combined. [22, 128, 129, 181–185]. All these pipelines and otherwise combined methods speed up the use of the existing tools, and allow them to be applied to ever larger numbers of small molecules in ever shorter times.

Actually, there is a new hype raging at the moment, and it is called 'translational science'. In the Wikipedia we find under translational research: "In the field of medicine, for example, it is used to translate the findings in basic research more quickly and efficiently into medical practice and, thus, meaningful health outcomes, whether those are physical, mental, or social outcomes". In a sense, the recent spate of articles on combining existing techniques into more easily applicable super-tools fit nicely to this translational paradigm. It must be stated, though, that the translation science hype is feeling stiff competition from systems biology [186] and modelling pharmacokinetics and pharmacodynamics [187]. Between the lines we read in translational science that the pharmaceutical industry has finally realized that our deep lack of understanding of all aspects of the interaction of a medicine with a human being is the main cause for luck still being the most determining factor in the drug design process. Consequently, we see the out-sourcing budgets of the large pharmaceutical industries go up [188], and more and more fundamental research performed in academia is finding its way to small and medium size enterprises (SMEs) where it can be incorporated in their lean and mean research machines [9]. Big pharma will at some time buy either their products or the whole SMEs and convert validated targets and leads or even Phase I products into new medicines.

This new paradigm will probably also be proven a hype soon; only time can tell if translational research will rescue the pharmaceutical industry, or that it will only better illustrate what it all is that we don't know yet. It remains a fact that better understanding the underlying biology, better treatment of all available data, and more intelligent combinations of data, information, and knowledge must be beneficial for the drug design process and thus, on the long run, for all of us.

If the pharmaceutical industry wants academia more involved in the drug design process they could themselves make a giant first step by making available all (or at least very many) X-ray structures of protein–ligand complexes. We estimate that the number of PDB [189, 190] files collecting computer dust in the pharmaceutical industry is considerably bigger than the 75,000 structures now in the PDB. We have discussed this possibility with industrial crystallographers who realized that they were sometimes sitting on thousands of structure files for which secrecy was

no longer an issue. They remained nevertheless hesitant to even consider discussing with their management the release of these data in fear of paranoia based rejection. Another often heard rejection criterion is that they are a bit ashamed for these data because often these files have not been refined any further than was needed to answer the biological or pharmaceutical question at hand. We offer to set up a database for these files, and we offer to re-refine all industrial structures of protein–ligand complexes. We will then only release those coordinates to the wider public that pass certain minimal validation criteria [191]. Obviously, the files in this system will remain the property of the depositors. If one day deposition of coordinate files into the PDB becomes significantly easier, we can consider depositing all files in the PDB on behalf of the original depositors. It might seem a bold promise to re-refine perhaps even 100,000 structures, but the PDB_REDO experiment [192–196] shows that today this can be done. In PDB_REDO we significantly improved 85% of all presently available PDB files that were solved by X-ray. It seems likely that structures that often have been minimally refined can be improved even easier. One can even envisage that industries would like to look back at their own coordinates after we went through the elaborate and time consuming refinement process for them; in management speak that would be the ultimate win–win situation.

Dealing with data, information, and knowledge: from hype to hope?

Despite massive efforts in the design of tools, databases, robotic techniques, and management innovations, luck seems to be at the basis of the discovery of most new medicines [197]. The blockbuster Viagra is probably the best illustration of the opportunism that we tend to call serendipity [198].

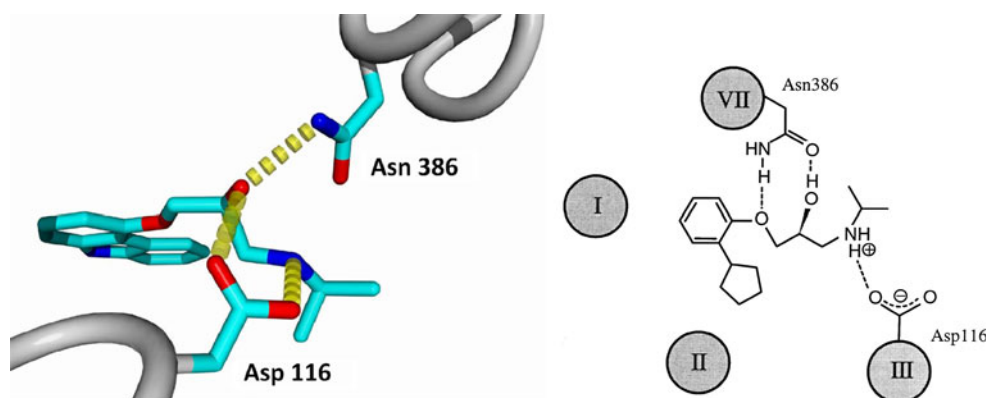
In 1997, i.e., long before the first GPCR structure became available, Kuipers et al. [199] performed a massive

literature search for aryloxypropanolamines and similar compounds binding to the serotonin 5HT-1a receptor and a series of sequence similar amine receptors. A correlation analysis [200] revealed that only one residue's presence/absence showed a perfect correlation with binding/non-binding of a series of compounds. A mutational study validated the hypothesis that this correlation indicated a direct hydrogen bond between an alcohol group in the aminergic ligand and asparagine 719 [201]. When the structure of the human β 2 adrenoceptor bound to carazolol was solved by X-ray [PDBid 2RH1; 202], it showed indeed two hydrogen bonds between Asn-719 and this similar ligand (see Fig 4). By the way, in none of the GPCR homology models available in 199 \times , did Asn-719 interact with a ligand.

In another GPCR related project aimed at using as much heterogeneous data as can possibly be combined, Oliveira et al. [203] predicted the role of all 'active site' residues in GPCRs, the pivotal role of Arg-340 [204], and even a series of residue interactions involved in the activation process, and the presence and location of helix VIII [205]. The recent flurry of articles on GPCR Xray structures [206–209], and especially the structure with a covalently agonist-bound G protein [210] showed all these predictions to be conceptually right.

These two GPCR-related examples make clear that there is a lot to be gained from using experimental data. But these examples also taught us how hard it is to actually get access to those data. With the GPCRDB [211–213] we have started a trend to make Molecular Class Specific Information Systems (MCSIS). And a small company, Bio-Product (www.bio-product.nl) recently caught on and is now making MCSISes for a wide variety of commercially interesting molecules [214–218]. Their systems (some of which are freely accessible from their website) revolve around a structure based, and thus very accurate multiple sequence alignment (MSA) for a whole protein super-family. This MSA then functions as the anchor on which to position all kinds of data that can range from 3D structures to genome

Fig. 4 Ligand binding by Asn-386. *Left:* (part of) the X-ray structure of the β 2 adrenoceptor bound to an inverse agonist that is 'somewhat similar' to (S)-penbutolol. *Right:* (S)-penbutolol binding of Asn-386 in serotonin 5HT-1A predicted long before the first GPCR structure data became available



related data, from mutation studies to ligand binding constants, or from sequence correlation patterns to the prediction of mutations that enhance the protein's stability. As the most powerful information tends to be carefully hidden in the literature, an extensive set of literature-mining scripts aids with the extraction of, for example, mutation information. In fact, it was shown that the suite of mutation data extracting scripts reaches a much better coverage than can be obtained by human experts [214–218].

A recent development that will aid the drug hunters of the future is the Utopia PDF reader [213, 219]. Vroling et al. [213] showed how this programmable PDF reader could be used to directly couple data in articles on GPCRs to the GPCRDB. This intelligent hyperlinking has a series of benefits. First, the residue numbering problem gets solved because the reader can ask the GPCRDB for the position in the GPCR MSA of any residue mentioned in the article, and it can even modify or correct the sequence numbers in the article if needed. Much good GPCR mutation data was published in the pre-GPCR-structure era that ended with the opsin structure article [220], and often these data were misinterpreted because of the poor quality of the available homology models [221]. The Utopia-GPCR PDF reader can correct those interpretations thereby salvaging old, high quality experimental data for future use. Figure 5 shows an image from an old mutation study [222] in which the authors describe several ground-breaking mutations in the guinea pig histamine H1 receptor, building and validating a homology model using these data, and arguing, for example, that residue Trp161 plays an important role in receptor-ligand binding. This assumption was based on the effect of the mutation on receptor function, leading to a model in which Trp161 was modelled in the ligand-binding site. By

contrast, the GPCRDB generated annotation listed in the sidebar of the reader indicates that this residue, located in TM IV, points towards the membrane and possibly interacts with cholesterol. This is a completely different situation from that proposed by the authors. Looking at the model provided by the GPCRDB, based on the latest crystal structures, it can be seen that a direct role of Trp161 in receptor-ligand binding is highly unlikely.

Folkerstma et al. (2005) analyzed nearly 100 nuclear receptor (NR) ligand binding domains. Combined with manually curated multiple sequence alignments, key positions in the ligand binding pocket were identified that had specific interactions with functionally diverse compounds. For example, residues at position 26 in Fig. 6 were shown to only have interactions with antagonists. This analysis required a substantial amount of work: categorizing structures and compounds, creating multiple sequence alignments, analyzing ligand contacts, and transferring the results into a homogeneous residue numbering scheme (the so-called 3D numbers). With the 3DM information system [223; see the help movie], these analyses can today be performed in a matter of minutes [215, 217, 224].

More than 100 articles were found that discuss the effects of mutating this residue on the ligand binding of the receptor. In all these articles this same residue has 14 different residue numbers ranging from 52 to 709. The use of a common 3D numbering scheme enables transfer of heterogeneous information between protein family members. Figure 7 shows 40 antagonists in red and 70 agonists in blue. In this example, a hundred articles had to be 'read' to extract all available mutation information for this single position mutated in 22 different receptor—species combinations. That these 100 articles had to be found among

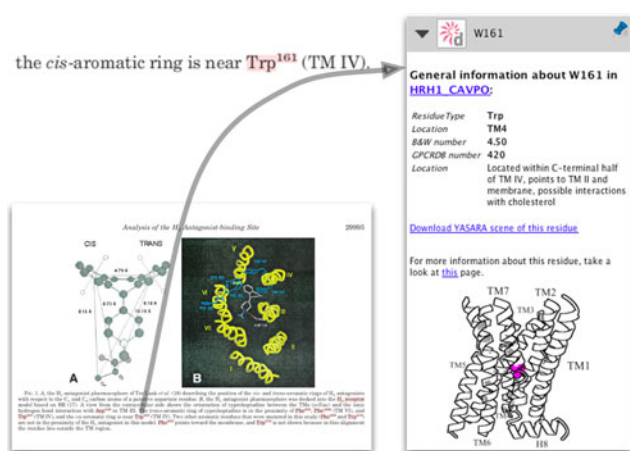
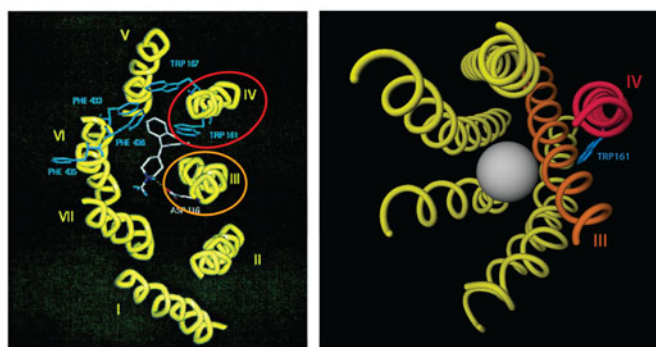


Fig. 5 *Left*, one page from the Histamine H1 article by Wieland et al. [222] in which Trp161 is suggested to interact with the ligand while the PDF reader sidebar shows today's interpretation that this tryptophan is facing outwards towards the lipid or a dimer partner. The original picture of the modelled active site is shown enlarged in



the middle panel while the right hand side figure is a plot of the GPCRDB-derived model of this receptor. The GPCRDB does not (yet) dock ligands, so the ligand is represented by a hand-added gray ball

Fig. 6 Bargraph showing the number of ligand contacts per residue extracted from 776 nuclear receptor ligand binding domain structures plotted as function of their 3D numbers. The *blue bars* represent the number of contacts with agonistic compounds. The *red bars* indicate the number of contacts with antagonistic compounds. The residue with 3D number 26 is only bound to antagonistic compounds

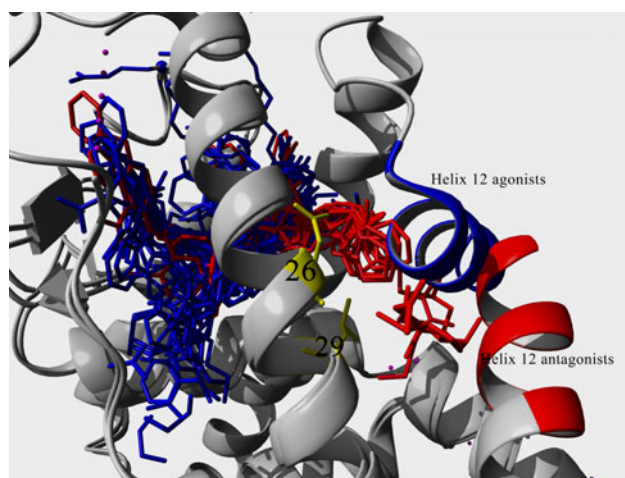
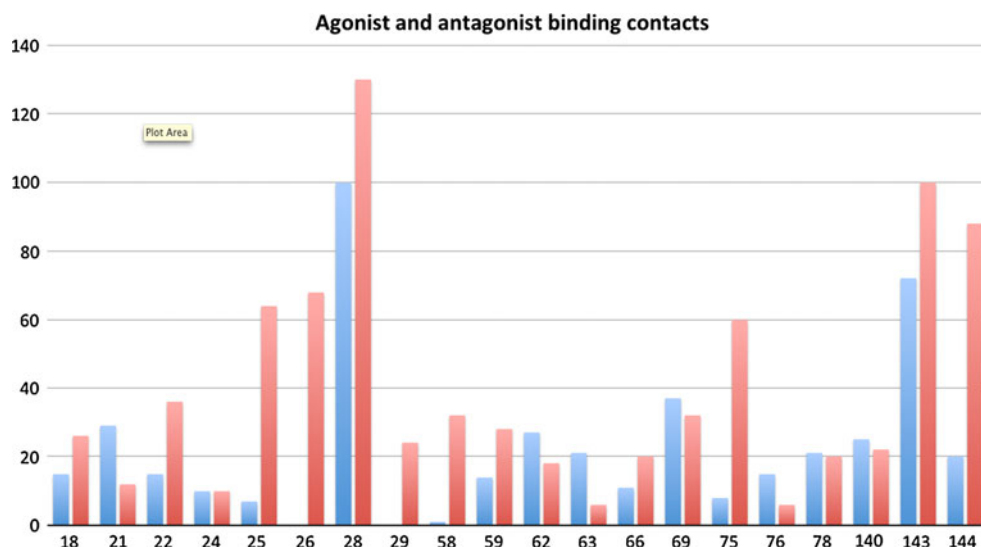


Fig. 7 Cartoon representation of two superposed representative NR structures (one bound to an agonist; one bound to an antagonist). These two structures, obviously, differ most in the location of Helix 12. The *blue* ligands are agonists; the *red* ones are antagonists. The ligands were placed in the same orientation as found in their native PDB file. All PDB files were superposed on the representative NR structures. Residue 26, for which the antagonist interaction had been mentioned in the literature, is shown in *yellow*, as is residue for which Fig. 6 also indicates antagonist interactions, albeit with less antagonists than residue 26. This role of residue 29 might represent a novel finding. Figure made with the 3DM-plugin for the YASARA—WHAT IF suite

100,000 PubMed entries that contain NR information is a whole different story in itself.

If, one day, all structures of NR-ligand complexes that now are scattered over inaccessible industrial hard disks could be concentrated in one system, then we could consider asking much more elaborate questions. We could consider correlating aspects of ligands with protein atom characteristics, or we could analyse if residues not

contacting the ligand have an influence on binding or activation, etcetera.

It is not only important to get as much information as possible stored in systems amenable to scrutiny, but it is also important to realize that for every one bioinformatician or drug hunter there are one hundred scientists who do not use molecular software regularly. Project Hope aims to predict the molecular phenotype of point mutations that were shown causally related to human disease states [225]. This system attempts in all stages of user interaction to cater for human geneticists who typically do not use molecular software at all. Hope only asks the user to cut-n-paste the sequence, and click the residue mutated and the mutation residue type. It then builds a homology model if needed, calls dozens of servers and services in seven countries, combines all possible information and writes a final report that can be directly used in publications, but, more importantly, that is written without using any bioinformatics jargon and even has a build-in dictionary that explains terms such as ‘active site’, ‘salt-bridge’, or ‘torsion angle’ in human genetics understandable terms. Hope thus is the ultimate translation machine because in doing translational research it even translates between the researchers.

We believe that the recent spate of consolidations in the pharmaceutical industry is not a problem but an opportunity. Mankind needs medicines, and now that pushing ones luck is slowly becoming a less successful technique, only research can find them. This research can progress rapidly if the thousands and thousands of X-ray structures of protein–ligand complexes would find their way from hard-disks behind pharmaceutical industry firewalls to the public domain. Drug design research in the next 25 years will revolve around ever broader collaborations, ever more holistic understanding of the drug–human interactions,

and ever better use of the available data, information, and knowledge.

Acknowledgments VL and GV acknowledge financial support from NBIC, and TIPharma, TvdB appreciate the support from Bio-Product (www.bio-product.com). The authors thank Jan Kelder for critically reviewing the manuscript. Elmar Krieger helped with YA-SARA, Maarten Heekelman, Coos Baakman, Bas Vroling, Wilmar Teunissen, Barbara van Kampen, provided technical support. The authors mention with pleasure the many stimulating discussions with Sander Nabuurs, Daniel Gironés, Gijs Schaftenaar, Friedrich Rippmann, Ad IJzerman, Margot Beukers, Isabel Duarte, Christof Francke, Henk-Jan Joosten, Jacob de Vlieg.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Life expectancy (2011) http://en.wikipedia.org/wiki/Life_expectancy. Accessed 6 Dec 2011
- World War I casualties (2011) http://en.wikipedia.org/wiki/World_War_I_casualties. Accessed 6 Dec 2011
- Medecine and World War II (2011) http://www.historylearning.site.co.uk/medicine_and_world_war_two.htm. Accessed 16 Feb 2011
- Snell ES, Griffin JP (1985) How many medicines are there? *Br Med J* 290:773–774
- Lewington A (1993) Medicinal plants and plant extracts: a review of the importation into Europe. Traffic network report. Traffic International, Cambridge, UK
- Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov*. doi:10.1038/nrd2199
- Munos B (2009) Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov*. doi:10.1038/nrd2961
- Troubled Asset Relief Program (2011) http://en.wikipedia.org/wiki/Troubled_Asset_Relief_Program. Accessed 6 Dec 2011
- Firestone RA (2011) Lessons from 54 years of pharmaceutical research *Nat Rev Drug Discov*. doi:10.1038/nrd2961-c1
- Robert B, Matthew T, Sean H, Bernard B (2000) Drug design by machine learning: support vector machine for pharmaceutical data analysis. Proceedings of the AISB'00 symposium on artificial intelligence in bioinformatics. pp 1–4
- Byvatov E, Fechner U, Sadowski J, Schneider G (2003) Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci* 43:1882–1889
- Gagnon MA, Lexchin J (2008) The cost of pushing pills: a new estimate of pharmaceutical promotion expenditures in the United States. *PLoS Med*. doi:10.1371/journal.pmed.0050001
- Hol WGJ (1986) Protein crystallography and computer-graphics toward rational drug design. *Angew Chem Int Ed Engl* 25:767–778
- Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J (1982) Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. *J Biol Chem* 257:13663–13672
- Thompson PE, Manallack DT, Blaney FE, Gallagher T (1992) Conformational studies on (+)-anatoxin-a and derivatives. *J Comput Aided Mol Des* 6:287–298
- Ruiz J, López M, Milà J, Lozoya E, Lozano JJ, Pouplana R (1993) QSAR and conformational analysis of the antiinflammatory agent amfenac and analogues. *J Comput Aided Mol Des* 7:183–198
- Aleman C, Perez JJ (1993) SCF-MO study of the polyglycine II structure. *J Comput Aided Mol Des* 7(2):241–250
- Oyasu H, Nakanishi I, Tanaka A, Murano K, Matsuo M (1995) Conformational studies on the four stereoisomers of the novel anticholinergic 4-(dimethylamino)-2-phenyl-2-(2-pyridyl)pentanamide. *J Comput Aided Mol Des* 9:171–180
- Schaftenaar G, Noordik JH (2000) Molden: a pre- and post-processing program for molecular and electronic structures. *J Comput Aided Mol Des* 14:123–134
- Manzetti S, McCulloch DR, Herington AC, van der Spoel D (2003) Modeling of enzyme-substrate complexes for the metalloproteases MMP-3, ADAM-9 and ADAM-10. *J Comput Aided Mol Des* 17:551–565
- Duran D, Aviyente V, Baysa C (2004) Solvent effect on the synthesis of clarithromycin: a molecular dynamics study. *J Comput Aided Mol Des* 18:145–154
- Curioni A, Mordasini T, Andreoni W (2004) Enhancing the accuracy of virtual screening: molecular dynamics with quantum-refined force fields. *J Comput Aided Mol Des* 18:773–784
- Hammond PS, Wu Y, Harris R, Minehardt TJ, Car R, Schmitt JD (2005) Protonation-induced stereoisomerism in nicotine: conformational studies using classical (AMBER) and ab initio (Car-Parinello) molecular dynamics. *J Comput Aided Mol Des* 19:1–15
- Roccatano D, Sbardella G, Aschi M, Amicosante G, Bossa C, Nola AD, Mazza F (2005) Dynamical aspects of TEM-1 beta-lactamase probed by molecular dynamics. *J Comput Aided Mol Des* 19:329–340
- Chipot C, Rozanska X, Dixit SB (2005) Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. *J Comput Aided Mol Des* 19:765–770
- Fanelli F, De Benedetti PG (2006) Inactive and active states and supramolecular organization of GPCRs: insights from computational modeling. *J Comput Aided Mol Des* 20:449–461
- Bharatham K, Bharatham N, Kwon YJ, Lee KW (2008) Molecular dynamics simulation study of PTP1B with allosteric inhibitor and its application in receptor based pharmacophore modeling. *J Comput Aided Mol Des* 22:925–933
- Eyrich S, Helms V (2009) What induces pocket openings on protein surface patches involved in protein-protein interactions? *J Comput Aided Mol Des* 23:73–86
- van Gunsteren WF, Berendsen HJ (1987) Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J Comput Aided Mol Des* 1:171–176
- Hansson T, Marelius J, Aqvist J (1998) Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aided Mol Des* 12:27–35
- Wilcox GL, Quioco FA, Levinthal C, Harvey SC, Maggiora GM, McCammon JA (1988) Symposium overview. Minnesota conference on supercomputing in biology: proteins, nucleic acids, and water. *J Comput Aided Mol Des* 1:271–281
- Wimmer E (1988) Future in biomolecular computation. *J Comput Aided Mol Des* 1:283–290
- Allinger NL (1977) Conformational-analysis. 130. Mm2—hydrocarbon force-field utilizing V1 and V2 torsional terms. *J Am Chem Soc* 99:8127–8134
- Stewart JJP (1990) Special issue—Mopac—a semiempirical molecular-orbital program. *J Comput Aided Mol Des* 4:1–45

35. Allen FH, Harris SE, Taylor R (1996) Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J Comput Aided Mol Des* 10:247–254
36. Klebe G, Mietzner T (1994) A fast and efficient method to generate biologically relevant conformations. *J Comput Aided Mol Des* 8:583–606
37. Bostrom J, Norrby PO, Liljefors T (1998) Conformational energy penalties of protein-bound ligands. *J Comput Aided Mol Des* 12:383–396
38. Mayer D, Naylor CB, Motoc I, Marshall GR (1987) A unique geometry of the active site of angiotensin-converting enzyme consistent with structure-activity studies. *J Comput Aided Mol Des* 1:3–16
39. Martin J, Andrews P (1987) Conformation-activity relationships of opiate analgesics. *J Comput Aided Mol Des* 1:53–72
40. Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA (1993) A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J Comput Aided Mol Des* 7:83–102
41. Lakdawala A, Wang M, Nevins N, Liotta D, Rusinska-Roszak D, Lozynski M, Snyder JP (2001) Calculated conformer energies for organic molecules with multiple polar functionalities are method dependent: taxol (case study). *BMC Chem Biol*. doi:10.1186/1472-6769-1-2
42. Vieth M, Hirst JD, Brooks CL (1998) Do active site conformations of small ligands correspond to low free-energy solution structures? *J Comput Aided Mol Des* 12:563–572
43. Klebe G (1995) Toward a more efficient handling of conformational flexibility in computer-assisted modelling of drug molecules. *Perspect Drug Discov Des* 3:85–105
44. Ota N, Agard DA (2001) Binding mode prediction for a flexible ligand in a flexible pocket using multi-conformation simulated annealing pseudo crystallographic refinement. *J Mol Biol* 314:607–617
45. Diller DJ, Merz KM Jr (2002) Can we separate active from inactive conformations? *J Comput Aided Mol Des* 16:105–112
46. Huse M, Kuriyan J (2003) The conformational plasticity of protein kinases. *Cell* 109:275–282
47. Teague SJ (2003) Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2:527–541
48. Chouard T (2005) Structural biology: breaking the protein rules. *Nature* 471:151–153
49. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208
50. Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* 1804:1231–1264
51. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 161:269–288
52. Bohm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 8:243–256
53. Meng EC, Kuntz ID, Abraham DJ, Kellogg GE (1994) Evaluating docked complexes with the HINT exponential function and empirical atomic hydrophobicities. *J Comput Aided Mol Des* 8:299–306
54. Jain AN (1996) Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des* 10:427–440
55. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* 16:11–26
56. Kelly MD, Mancera RL (2003) A new method for estimating the importance of hydrogen-bonding groups in the binding site of a protein. *J Comput Aided Mol Des* 17:401–414
57. Muryshev AE, Tarasov DN, Butygin AV, Butygina OY, Aleksandrov AB, Nikitin SM (2003) A novel scoring function for molecular docking. *J Comput Aided Mol Des* 17:597–605
58. Morley SD, Afshar M (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *J Comput Aided Mol Des* 18:189–208
59. Tame JR (2005) Scoring functions—the first 100 years. *J Comput Aided Mol Des* 19:445–451
60. Jain AN (2009) Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aided Mol Des* 21:281–306
61. Cincilla G, Vidal D, Pons M (2009) An improved scoring function for suboptimal polar ligand complexes. *J Comput Aided Mol Des* 23:143–152
62. Dobeš P, Fanfrlík J, Rezáč J, Otyepka M, Hobza P (2011) Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. *J Comput Aided Mol Des* 25(3):223–235
63. Tondel K, Anderssen E, Drablos F (2006) Protein alpha shape (PAS) dock: a new gaussian-based score function suitable for docking in homology modelled protein structures. *J Comput Aided Mol Des* 20:131–144
64. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP (1994) FLOG: a system to select ‘quasi-flexible’ ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* 8:153–174
65. Kearsley SK, Underwood DJ, Sheridan RP, Miller MD Flexi-bases: a way to enhance the use of molecular docking methods. *J Comput Aided Mol Des* 8:565–582
66. Oshiro CM, Kuntz ID, Dixon JS (1995) Flexible ligand docking using a genetic algorithm. *J Comput Aided Mol Des* 9:113–130
67. Knegt RM, Bayada DM, Engh RA, von der Saal W, van Gerestein VJ, Grootenhuys PD Comparison of two implementations of the incremental construction algorithm in flexible docking of thrombin inhibitors. *J Comput Aided Mol Des* 13:167–183
68. Makino S, Ewing TJ, Kuntz ID (1999) DREAM++: flexible docking program for virtual combinatorial libraries. *J Comput Aided Mol Des* 13:513–532
69. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 15:411–428
70. Hindle SA, Rarey M, Buning C, Lengae T (2002) Flexible docking under pharmacophore type constraints. *J Comput Aided Mol Des* 16:129–149
71. Grasselli M, Cascone O, Birger Anspach F, Delfino JM (2002) On the molecular interaction between lactoferrin and the dye Red HE-3b. A novel approach for docking a charged and highly flexible molecule to protein surfaces. *J Comput Aided Mol Des* 16:917–934
72. Bursulaya BD, Totrov M, Abagyan R, Brooks CL 3rd (2003) Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des* 17:755–763
73. Bottegoni G, Kufareva I, Totrov M, Abagyan R (2008) A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). *J Comput Aided Mol Des* 22:311–325
74. Zhao Y, Sanner MF (2008) Protein-ligand docking with multiple flexible side chains. *J Comput Aided Mol Des* 22:673–679
75. Kang L, Li H, Jiang H, Wang X (2009) An improved adaptive genetic algorithm for protein-ligand docking. *J Comput Aided Mol Des* 23:1–12

76. Jain AN (2009) Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *J Comput Aided Mol Des* 23:355–374
77. Garden DP, Zhorov BS (2010) Docking flexible ligands in proteins with a solvent exposure- and distance-dependent dielectric function. *J Comput Aided Mol Des* 24:91–105
78. Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 18:178–184
79. Bohm HJ, Klebe G (1996) What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? *Angew Chem Int Ed Engl* 35:2589–2614
80. Shoichet BK, Leach AR, Kuntz ID (1999) Ligand solvation in molecular docking. *Protein Struct Funct Genet* 34:4–16
81. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed* 41:2644–2676
82. Demchuk E, Wade RC (1996) Improving the continuum dielectric approach to calculating pK(a)s of ionizable groups in proteins. *J Phys Chem* 100:17373–17387
83. Nielsen JE, Vriend G (2001) Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. *Protein Struct Funct Genet* 43:403–412
84. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 261:470–489
85. Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* 9:1–5
86. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749
87. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748
88. Hartmann C, Antes I, Lengauer T (2009) Docking and scoring with alternative side-chain conformations. *Proteins* 74:712–726
89. Taylor RD, Jewsbury PJ, Essex JW (2003) FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J Comput Chem* 24:1637–1656
90. Muegge I (2006) PMF scoring revisited. *J Med Chem* 49:5895–5902
91. Englebienne P, Moitessier N (2009) Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *J Chem Inf Model* 49:1568–1580
92. Oda A, Tsuchida K, Takakura T, Yamaotsu N, Hirono S (2006) Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J Chem Inf Model* 46:380–391
93. Foloppe N, Hubbard R (2006) Towards predictive ligand design with free-energy based computational methods? *Curr Med Chem* 13:3583–3608
94. Jain AN (2006) Scoring functions for protein-ligand docking. *Curr Protein Pept Sci* 7:407–420
95. Robertson TA, Varani G (2007) An all-atom, distance-dependent scoring function for the prediction of protein-DNA interactions from structure. *Proteins* 66:359–374
96. Rajamani R, Good AC (2007) Ranking poses in structure-based lead discovery and optimization: current trends in scoring function development. *Curr Opin Drug Discov Devel* 10:308–315
97. Tembre BL, McCammon JA (1984) Ligand-receptor interactions. *Comput Chem* 8:281–283
98. Ferguson DM, Radmer RJ, Kollman PA (1991) Determination of the relative binding free-energies of peptide inhibitors to the Hiv-1 protease. *J Med Chem* 34:2654–2659
99. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular-field analysis (Comfa).1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110:5959–5967
100. Norinder U (1990) Experimental design based 3-D QSAR analysis of steroid-protein interactions: application to human CBG complexes. *J Comput Aided Mol Des* 4:381–389
101. Bursi R, Grootenhuis PD (1999) Comparative molecular field analysis and energy interaction studies of thrombin-inhibitor complexes. *J Comput Aided Mol Des* 13:221–232
102. Zhang Z, An L, Hu W, Xiang Y (2007) 3D-QSAR study of hallucinogenic phenylalkylamines by using CoMFA approach. *J Comput Aided Mol Des* 21:145–153
103. Cramer RD, Milne M (1979) Lattice model—general paradigm for shape-related structure-activity correlation. Abstracts of papers of the American chemical society, 19th ACS Meeting COMP 44
104. Nicklaus MC, Milne GW, Burke TR Jr (1992) QSAR of conformationally flexible molecules: comparative molecular field analysis of protein-tyrosine kinase inhibitors. *J Comput Aided Mol Des* 6:487–504
105. Rault S, Bureau R, Pilo JC, Robba M (1992) Comparative molecular field analysis of CCK-A antagonists using field-fit as an alignment technique. A convenient guide to design new CCK-A ligands. *J Comput Aided Mol Des* 6:553–568
106. Calder JA, Wyatt JA, Frenkel DA, Casida JE (1993) CoMFA validation of the superposition of six classes of compounds which block GABA receptors non-competitively. *J Comput Aided Mol Des* 7:45–60
107. Kroemer RT, Hecht P (1995) A new procedure for improving the predictiveness of CoMFA models and its application to a set of dihydrofolate reductase inhibitors. *J Comput Aided Mol Des* 9:396–406
108. Gohda K, Mori I, Ohta D, Kikuchi T (2000) A CoMFA analysis with conformational propensity: an attempt to analyze the SAR of a set of molecules with different conformational flexibility using a 3D-QSAR method. *J Comput Aided Mol Des* 14:265–275
109. Manchester J, Czermiński R (2009) CAUTION: popular “Benchmark” data sets do not distinguish the merits of 3D QSAR methods. *J Chem Inf Model* 49:1449–1454
110. Kharkar PS, Reith ME, Dutta AK (2008) Three-dimensional quantitative structure-activity relationship (3D QSAR) and pharmacophore elucidation of tetrahydropyran derivatives as serotonin and norepinephrine transporter inhibitors. *J Comput Aided Mol Des* 22:1–17
111. Park H, Lee S (2004) Homology modeling, force field design, and free energy simulation studies to optimize the activities of histone deacetylase inhibitors. *J Comput Aided Mol Des* 18:375–388
112. Tomich CH, da Silva P, Carvalho I, Taft CA (2005) Homology modeling and molecular interaction field studies of alpha-glucosidases as a guide to structure-based design of novel proposed anti-HIV inhibitors. *J Comput Aided Mol Des* 19:83–92
113. Rossi KA, Markwalder JA, Seitz SP, Chang CH, Cox S, Boisclair MD, Brizuela L, Brenner SL, Stouten PF (2005) Understanding and modulating cyclin-dependent kinase inhibitor specificity: molecular modeling and biochemical evaluation of pyrazolopyrimidinones as CDK2/cyclin A and CDK4/cyclin D1 inhibitors. *J Comput Aided Mol Des* 19:111–122
114. Schlegel B, Lagner C, Meier R, Langer T, Schnell D, Seifert R, Stark H, Höltje HD, Sippl W (2007) Generation of a homology model of the human histamine H(3) receptor for ligand docking and pharmacophore-based screening. *J Comput Aided Mol Des* 21:437–453

115. Katritch V, Byrd CM, Tseitin V, Dai D, Raush E, Totrov M, Abagyan R, Jordan R, Hraby DE (2007) Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J Comput Aided Mol Des* 21:549–558
116. Neves MA, Simoes S, Sa e Melo ML (2010) Ligand-guided optimization of CXCR4 homology models for virtual screening using a multiple chemotype approach. *J Comput Aided Mol Des* 24:1023–1033
117. Knehans T, Schüller A, Doan DN, Nacro K, Hill J, Güntert P, Madhusudhan MS, Weil T, Vasudevan SG (2011) Structure-guided fragment-based in silico drug design of dengue protease inhibitors. *J Comput Aided Mol Des* 25:263–274
118. Eberini I, Daniele S, Parravicini C, Sensi C, Trincavelli ML, Martini C, Abbracchio MP (2011) In silico identification of new ligands for GPR17: a promising therapeutic target for neurodegenerative diseases. *J Comput Aided Mol Des* 25:743–752
119. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
120. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:52–56
121. Reichelt J, Dieterich G, Kvesic M, Schomburg D, Heinz DW (2005) BRAGI: linking and visualization of database information in a 3D viewer and modeling tool. *Bioinformatics* 21:1291–1293
122. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* 9:114–122
123. Thompson J, Baker D (2011) Incorporation of evolutionary information into Rosetta comparative modeling. *Protein Struct Funct Bioinfo* 79:2380–2388
124. CASP1 proceedings (1995) *Protein Struct Funct Genet* 23: 295–460
125. Harris MR, Kihlen M, Bywater RP (1993) PLIM: a protein-ligand interaction modeller. *J Mol Recognit* 6:111–115
126. Michino M, Abola E; GPCR Dock 2008 participants, Brooks CL 3rd, Dixon JS, Moulton J, Stevens RC (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008 *Nat Rev Drug Discov* 8:455–463
127. Kufareva I, Rueda M, Vsevolod K, Stevens RC, Abagyan R, GPCR Dock 2010 participants Status of GPCR modeling and docking as reflected by community wide GPCR Dock 2010 assessment. *Structure* 19:1108–1126
128. Broer BM, Gurrath M, Holtje HD (2003) Molecular modelling studies on the ORL1-receptor and ORL1-agonists. *J Comput Aided Mol Des* 17:739–754
129. Miguet L, Zhang Z, Barbier M, Grigorov MG (2006) Comparison of a homology model and the crystallographic structure of human 11beta-hydroxysteroid dehydrogenase type 1 (11betaHSD1) in a structure-based identification of inhibitors. *J Comput Aided Mol Des* 20:67–81
130. Roumen L, Sanders MP, Vrolijk B, De Esch IJ, De Vlieg J, Leurs R, Klomp JP, Nabuurs SB, De Graaf C (2011) The pitfalls and challenges of predicting GPCR-ligand interactions. *Pharmaceuticals* 4:1196–1215
131. Henriques ES, Floriano WB, Reuter N, Melo A, Brown D, Gomes JA, Maigret B, Nascimento MA, Ramos MJ (2001) The search for a new model structure of beta-factor XIIa. *J Comput Aided Mol Des* 15:309–322
132. Brown N, McKay B, Gasteiger J (2004) The de novo design of median molecules within a property range of interest. *J Comput Aided Mol Des* 18:761–771
133. Belda I, Madurga S, Llorà X, Martinell M, Tarragó T, Piqueras MG, Nicolás E, Giralt E (2005) ENPDA: an evolutionary structure-based de novo peptide design algorithm. *J Comput Aided Mol Des* 19:585–601
134. Zaliani A et al (2009) Second-generation de novo design: a view from a medicinal chemist perspective. *J Comput Aided Mol Des*
135. Goodford PJ (1985) A computational-procedure for determining energetically favorable binding-sites on biologically important macromolecules. *J Med Chem* 28:849–857
136. Wade RC, Clark KJ, Goodford PJ (1993) Further development of hydrogen-bond functions for use in determining energetically favorable binding-sites on molecules of known structure. I. Ligand probe groups with the ability to form 2 hydrogen-bonds. *J Med Chem* 36:140–147
137. Rotstein SH, Murcko MA (1993) GroupBuild: a fragment-based method for de novo drug design. *J Med Chem* 36:1700–1710
138. Willett P (1995) Genetic algorithms in molecular recognition and design. *Trends Biotechnol* 13:516–521
139. Bohm HJ (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J Comput Aided Mol Des* 6:61–78
140. Bohm HJ (1992) LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 6:593–606
141. Clark DE, Frenkel D, Levy SA, Li J, Murray CW, Robson B, Waszkowycz B, Westhead DR (1995) PRO-LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. *J Comput Aided Mol Des* 9:13–32
142. Westhead DR, Clark DE, Frenkel D, Li J, Murray CW, Robson B, Waszkowycz B (1995) PRO-LIGAND: an approach to de novo molecular design. 3. A genetic algorithm for structure refinement. *J Comput Aided Mol Des* 9:139–148
143. Rarey M, Wefing S, Lengauer T (1996) Placement of medium-sized molecular fragments into active sites of proteins. *J Comput Aided Mol Des* 10:41–54
144. Barakat MT, Dean PM (1995) The atom assignment problem in automated de novo drug design. 1. Transferability of molecular fragment properties. *J Comput Aided Mol Des* 9:341–350
145. Rotstein SH, Murcko MA (1993) GenStar: a method for de novo drug design. *J Comput Aided Mol Des* 7:23–43
146. Roe DC, Kuntz ID (1995) BUILDER v.2: improving the chemistry of a de novo design strategy. *J Comput Aided Mol Des* 9:269–282
147. Leach AR, Kilvington SR (1994) Automated molecular design: a new fragment-joining algorithm. *J Comput Aided Mol Des* 8:283–298
148. Golberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, New York
149. Todorov NP, Dean PM (1997) Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J Comput Aided Mol Des* 11:175–192
150. Schneider G, Lee ML, Stahl M, Schneider P (2000) De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J Comput Aided Mol Des* 14:487–494
151. Blum LC, Reymond JL (2009) 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J Am Chem Soc* 131:8732–8733
152. Kubinyi H (1992) HTS technologies—IBC inform conference. *IDrugs* 4:168–173
153. Sudarsanam S, Virca GD, March CJ, Srinivasan S (1992) An approach to computer-aided inhibitor design: application to cathepsin L. *J Comput Aided Mol Des* 6:223–233
154. Bohm HJ, Banner DW, Weber L (1999) Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J Comput Aided Mol Des* 13:51–56
155. Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, James TL (2000) Identification of ligands for

- RNA targets via structure-based virtual screening: HIV-1 TAR. *J Comput Aided Mol Des* 14:593–610
156. Furet P, Meyer T, Mittl P, Fretz H (2001) Identification of cylin-
dependent kinase 1 inhibitors of a new chemical type by
structure-based design and database searching. *J Comput Aided
Mol Des* 15:489–495
157. Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Marko-
poulos J, Igglessi-Markopoulou O (2006) Investigation of
substituent effect of 1-(3, 3-diphenylpropyl)-piperidiny phenyl-
acetamides on CCR5 binding affinity using QSAR and virtual
screening techniques. *J Comput Aided Mol Des* 20:83–95
158. Kortagere S, Welsh WJ (2006) Development and application of
hybrid structure based method for efficient screening of ligands
binding to G-protein coupled receptors. *J Comput Aided Mol
Des* 20:789–802
159. Talevi A, Bellera CL, Castro EA, Bruno-Blanch LE (2007) A
successful virtual screening application: prediction of anticon-
vulsant activity in MES test of widely used pharmaceutical and
food preservatives methylparaben and propylparaben. *J Comput
Aided Mol Des* 21:527–538
160. Irwin JJ (2008) Community benchmarks for virtual screening.
J Comput Aided Mol Des 22:193–199
161. Krueger BA, Weil T, Schneider G (2009) Comparative virtual
screening and novelty detection for NMDA-GlycineB antago-
nists. *J Comput Aided Mol Des* 23:869–881
162. Didziapetris R, Dapkunas J, Sazonovas A, Japertas P (2010)
Trainable structure-activity relationship model for virtual
screening of CYP3A4 inhibition. *J Comput Aided Mol Des*
24:891–906
163. Joseph-McCarthy D (2002) An overview of in silico design and
screening: toward efficient drug discovery. *Curr Drug Discov*
20–23
164. Bajorath J (2002) Virtual screening: methods, expectations, and
reality. *Curr Opin Drug Discov Dev* 2:24–28
165. Rarey M, Stahl M (2001) Similarity searching in large combi-
natorial chemistry spaces. *J Comput Aided Mol Des* 15:
497–520
166. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a
useful metric for lead selection. *Drug Discov Today* 9:430–431
167. Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as
guideposts for drug discovery. *Drug Discov Today* 10:464–469
168. Jain AN, Koile K, Chapman D (1994) Compass: predicting
biological activities from molecular surface properties. Perfor-
mance comparisons on a steroid benchmark. *J Med Chem*
37:2315–2327
169. Handschuh S, Goldfuss B, Chen J, Gasteiger J, Houk KN (2000)
Steroid binding by antibodies and artificial receptors: explora-
tion of theoretical methods to determine the origins of binding
affinities and specificities. *J Comput Aided Mol Des* 14:611–629
170. Jain AN, Dietterich TG, Lathrop RH, Chapman D, Critchlow RE
Jr, Bauer BE, Webster TA, Lozano-Perez T (1994) A shape-
based machine learning tool for drug design. *J Comput Aided
Mol Des* 8:635–652
171. Walters WP, Murcko MA (2002) Prediction of ‘drug-likeness’.
Adv Drug Deliv Rev 54:255–271
172. Jung E, Choi SH, Lee NK, Kang SK, Choi YJ, Shin JM, Choi K,
Jung DH (2011) Machine learning study for the prediction of
transdermal peptide. *J Comput Aided Mol Des* 25:339–347
173. Sapre NS, Gupta S, Pancholi N, Sapre N (2008) Molecular
docking studies on tetrahydroimidazo-[4, 5, 1-jk][1, 4]-benzo-
diazepinone (TIBO) derivatives as HIV-1 NNRT inhibitors.
J Comput Aided Mol Des 22:69–80
174. Hirst JD, King RD, Sternberg MJ (1994) Quantitative structure-
activity relationships by neural networks and inductive logic
programming. II. The inhibition of dihydrofolate reductase by
triazines. *J Comput Aided Mol Des* 8:421–432
175. Livingstone DJ, Manallack DT, Tetko IV (1997) Data modelling
with neural networks: advantages and limitations. *J Comput
Aided Mol Des* 11:135–142
176. Doweiko AM (2008) QSAR: dead or alive? *J Comput Aided
Mol Des* 22:81–89
177. Che Y, Brooks BR, Marshall GR (2006) Development of small
molecules designed to modulate protein-protein interactions.
J Comput Aided Mol Des 20:109–130
178. Lotta T, Taskinen J, Bäckström R, Nissinen E (1992) PLS mod-
elling of structure-activity relationships of catechol O-methyl-
transferase inhibitors. *J Comput Aided Mol Des* 6:253–272
179. Zhou H, Lai WP, Zhang Z, Li WK, Cheung HY (2009) Com-
putational study on the molecular inclusion of andrographolide
by cyclodextrin. *J Comput Aided Mol Des* 23:153–162
180. Sarmah P, Deka RC (2009) DFT-based QSAR and QSPR
models of several cis-platinum complexes: solvent effect.
J Comput Aided Mol Des 23:343–354
181. Singh N, Chev  G, Ferguson DM, McCurdy CR (2006) A
combined ligand-based and target-based drug design approach
for G-protein coupled receptors: application to salvinorin A, a
selective kappa opioid receptor agonist. *J Comput Aided Mol
Des* 20:471–493
182. Zheng M, Yu K, Liu H, Luo X, Chen K, Zhu W, Jiang H (2006)
QSAR analyses on avian influenza virus neuraminidase inhibi-
tors using CoMFA, CoMSIA, and HQSAR. *J Comput Aided
Mol Des* 20:549–566
183. Clark RD (2008) A ligand’s-eye view of protein binding.
J Comput Aided Mol Des 22:507–521
184. Caballero J, Quiliano M, Alzate-Morales JH, Zimic M, Deharo E
(2011) Docking and quantitative structure-activity relationship
studies for 3-fluoro-4-(pyrrolo[2, 1-f][1, 2, 4]triazin-4-yloxy)
aniline, 3-fluoro-4-(1H-pyrrolo[2, 3-b]pyridin-4-yloxy)aniline,
and 4-(4-amino-2-fluorophenoxy)-2-pyridinylamine derivatives
as c-Met kinase inhibitors. *J Comput Aided Mol Des* 25:349–369
185. Yoo J, Medina-Franco JL (2011) Homology modeling, docking
and structure-based pharmacophore of inhibitors of DNA
methyltransferase. *J Comput Aided Mol Des* 25:555–567
186. Papp B, Notebaart RA, P l C (2011) Systems-biology approaches
for predicting genomic evolution. *Nat Rev Genet* 12:591–602
187. Breimer DD (2008) PK/PD modelling and beyond: impact on
drug development. *Pharm Res* 25:2720–2722
188. Swati Chaturvedi Outsourcing in Pharmaceutical Industry
(2011) [http://www.bionity.com/en/whitepapers/49803/outsour-
cing-in-pharmaceutical-industry.html](http://www.bionity.com/en/whitepapers/49803/outsourcing-in-pharmaceutical-industry.html). Accessed 6 Dec 2011
189. Protein Data Bank (2011) <http://www.pdb.org>. Accessed 6 Dec
2011
190. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN,
 Weissig H, Shindyalov IN, Bourne PE (2000) The protein data
bank. *Nucleic Acids Res* 28:235–242
191. Read RJ, Adams PD, Arendall WB 3rd, Brunger AT, Emsley P,
 Joosten RP, Kleywegt GJ, Krissinel EB, L tteke T, Otwinowski
Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ,
 Vriend G, Zwart PH (2011) A new generation of crystallo-
graphic validation tools for the protein data bank. *Structure* 19:
1395–1412
192. PDB redo (2011) http://www.cmbi.ru.nl/pdb_redo/. Accessed 29
Nov 2011
193. Joosten RP, Vriend G (2007) PDB improvement starts with data
deposition. *Science* 317:195–196
194. Joosten RP, Joosten K, Cohen SX, Vriend G, Perrakis A (2011)
Automatic rebuilding and optimization of crystallographic
structures in the PDB. *Bioinformatics Epub ahead of print*
195. Joosten RP, Salzemann J, Bloch V, Stockinger H, Berglund A,
 Blanchet C, Bongcam-Rudloff E, Combet C, Da Costa AL,
 Deleage G, Diarena M, Fabbretti R, Fettahi G, Flegel V, Gisel
A, Kasam V, Kervinen T, Korpelainen E, Mattila K, Pagni M,

- Reichstadt M, Breton V, Tickle IJ, Vriend G (2009) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Cryst* 42:376–384
196. Joosten RP, Womack T, Vriend G, Bricogne G (2009) Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallogr D Biol Crystallogr* 65:176–185
197. Ban TA (2006) The role of serendipity in drug discovery. *Dialogues Clin Neurosci* 8:244–335
198. Nurnberg HG, Hensley PL, Gelenberg AJ, Fava M, Lauriello J, Paine S (2003) Treatment of antidepressant-associated sexual dysfunction with sildenafil: a randomized controlled trial. *JAMA* 289:56–64
199. Kuipers W, Link R, Standaar PJ, Stoit AR, Van Wijngaarden I, Leurs R, IJzerman AP (1997) Study of the interaction between aryloxypropanolamines and Asn386 in helix VII of the human 5-hydroxytryptamine1A receptor. *Mol Pharmacol* 51:889–896
200. Kuipers W, Oliveira L, Paiva ACM, Rippman F, Sander C, IJzerman AP (1996) In: Findlay J (ed) *Membrane protein models*. Bios Scientific Publishers, Oxford
201. GPCRDB information system for G protein-coupled receptors (2011) <http://www.gpcr.org/7tm/>. Accessed 6 Dec 2011
202. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318:1258–1265
203. Oliveira L, Paiva ACM, Vriend G (1993) A common motif in G-protein-coupled seven transmembrane helix receptors. *J Comp aided Mol Des* 7:649–658
204. Oliveira L, Paiva ACM, Sander C, Vriend G (1994) A common step for signal transduction in G protein-coupled receptors. *Trends Pharmacol Sci* 15:170–172
205. Oliveira L, Paiva ACM, Vriend G (1999) A low resolution model for the interaction of G proteins with G protein-coupled receptors. *Protein Eng* 12:1087–1095
206. Wacker D, Brown MA, Fenalti G, Katritch V, Abagyan R, Cherezov V, Stevens RC (2010) Conserved binding mode of human beta(2) adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J Am Chem Soc* 132:11443–11445
207. Standfuss J, Xie G, Edwards PC, Burghammer M, Oprrian DD, Schertler GFX (2007) Crystal structure of a thermally stable rhodopsin mutant. *J Mol Biol* 372:1179–1188
208. Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG (2008) Conformational thermostabilization of the b1-adrenergic receptor in a detergent-resistant form of the b1-adrenergic receptor in a detergent-resistant form. *Proc Natl Acad Sci USA* 105: 877–882
209. Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, Kobilka BK (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450:383–377
210. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SGF, Choi HJ, DeVree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK (2011) Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature* 469:236–240
211. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res* 31:294–297
212. Horn F, Weare J, Beukers MW, Hörsch S, Bairoch A, Chen W, Edvardsen O, Campagne F, Vriend G (1998) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res* 26:275–279
213. Vroling B, Thorne D, McDermott P, Attwood TK, Vriend G, Pettifer S (2011) Integrating GPCR-specific information with full text articles. *BMC Bioinfo* 12:362
214. Joosten HJ, Han Y, Niu W, Vervoort J, Dunaway-Mariano D, Schaap PJ (2008) Identification of fungal oxaloacetate hydrolyase within the isocitrate lyase/PEP mutase enzyme superfamily using a sequence marker-based method. *Proteins* 70: 157–166
215. Kuipers RK, Joosten HJ, van Berkel WJ, Leferink NG, Rooijen E, Ittmann E, van Zimmeren F, Jochens H, Bornscheuer U, Vriend G, dos Santos VA, Schaap PJ (2010) 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* 78:2013–2101
216. Kourist R, Jochens H, Bartsch S, Kuipers R, Padhi SK, Gall M, Böttcher D, Joosten HJ, Bornscheuer UT (2010) The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *Chembiochem* 11:1635–1643
217. Kuipers R, van den Bergh T, Joosten HJ, Lekan dit Deprez RH, Mannens MM, Schaap PJ (2010) Novel tools for extraction and validation of disease-related mutations applied to Fabry disease. *Hum Mutat* 31:1026–1032
218. Cerdobbel A, De Winter K, Aerts D, Kuipers R, Joosten HJ, Soetaert W, Desmet T (2011) Increasing the thermostability of sucrose phosphorylase by a combination of sequence- and structure-based mutagenesis. *Protein Eng Des Sel* 24:829–834
219. Attwood TK, Kell DB, McDermott P, Marsh J, Pettifer SR, Thorne D (2009) Calling international rescue: knowledge lost in literature and data landslide! *Biochem J* 424:317–333
220. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Letrong I, Teller T, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289:739–745
221. Oliveira L, Hulsen T, Lutje Hulsik D, Paiva AC, Vriend G (2004) Heavier-than-air flying machines are impossible. *FEBS Lett* 564:269–273
222. Wieland K, ter Laak AM, Smit MJ, Kühne R, Timmerman H, Leurs R (1999) Mutational analysis of the antagonist-binding site of the histamine H(1) receptor. *J Biol Chem* 274: 29994–30000
223. Nuclear Receptors 3DM (2011) <https://funken.wur.nl/?sfamid=NR&mode=alignment>. Accessed 9 Dec 2011
224. Stenson PD, Cooper DN (2010) Prospects for the automated extraction of mutation data from the scientific literature. *Hum Genomics* 5:1–4
225. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinfo* 11:548