Beilstein-Institut
Cracking the Sugar Code by Navigating the Glycospace
June 27th – July 1st, 2011, Potsdam, Germany

Representing Carbohydrates by Pseudoreceptor Models for Virtual Screening in Drug Discovery

Max Pillong and Gisbert Schneider*

Swiss Federal Institute of Technology (ETH), Department of Chemistry and Applied Biosciences, Zürich, Switzerland

E-Mail: *gisbert.schneider@pharma.ethz.ch

Received: 9th January 2012/Published: 11th July 2012

Abstract

Carbohydrates provide a rich source of structural diversity that could be increasingly useful for innovative drug design. We suggest a representation of monosaccharides based on their pharmacophoric properties (pseudoreceptor model) to enable quantitative similarity searching with the aim to identify sugar bioisosters and functionally equivalent scaffolds for synthesis. We present a bioinformatical comparison of carbohydrate structures based on pseudoreceptor models. A similarity matrix was computed for 19 monosaccharide structures. As an outcome of this preliminary analysis, one might consider both glucose and deoxyribose as 'universal' sugars with regard to their receptor interaction potential. Potential applications of pharmacophore feature representations of carbohydrate structures in bioinformatics are discussed. A recent case study is reviewed that led to the identification of aminoglycoside scaffold replacements with antibacterial potential by pseudoreceptor-based virtual screening of a large compound library.

Introduction

One of the early and pivotal steps in drug discovery is the identification of structurally novel chemical entities (NCE) that exhibit a desired effect on a pharmaceutically relevant target molecule. Computer-assisted molecular design and 'virtual screening' technology is being increasingly used for this purpose with carbohydrate moieties offering a potentially substantial increase in structural diversity of candidate compounds [1]. Cipolla *et al.* recently

affirmed that "[the] polyfunctionality of carbohydrates stimulate[s] their use as scaffolds for the generation of libraries by combinatorial decoration with different pharmacophores."[2] As a prerequisite for success in this endeavour, accurate representations of molecular structure and computation of relevant properties is essential. Current virtual screening approaches like, e.g., automated ligand-receptor docking or pharmacophore similarity searching typically rely on heuristic or empirical methods for conformer generation and property estimation, offering the advantage to computationally sieve through several millions of compounds [3]. As a result, prioritized virtual hits are suggested for biochemical activity determination. This strategy has been successfully applied to many hit finding exercises in early-phase drug discovery, by complementing experimental high-throughput screening as a source for new pharmacologically appealing compounds [4]. While mere hit retrieval seems feasible by virtual screening the next step in the drug discovery pipeline, namely automated computer-assisted hit-to-lead optimization, requires further attention and method development [5].

Potentially different representations of molecular structures, their dynamics and properties are required for each step of chemical optimization [6]. Therefore, it will be necessary to develop and adapt appropriate molecular 'descriptors' of compounds of pharmaceutical interest. Although the role of carbohydrate moieties for ligand-receptor interactions has been realized and partially addressed in computational medicinal chemistry [7, 8], specifically in the field of lectin-sugar interactions [9] and natural product analysis [10, 11], carbohydrates have rarely been explicitly considered for virtual screening or computer-assisted drug design, and consequently our understanding of apt descriptors is limited. Current bioinformatical analyses of biologically relevant polycarbohydrate structures, e.g. glycan sequences, usually exclude modelling of potential receptor interactions [12]. Here, we suggest a 'pseudoreceptor' representation of monosaccharides as a conformation-sensitive descriptor for molecular modelling [13]. Recently, we successfully applied this theoretical concept to finding small synthetic inhibitors of bacterial protein biosynthesis by taking aminoglycosides as query compounds for virtual screening of large compound libraries [14].

Carbohydrate-protein Interactions and Pseudoreceptor Modelling

Sugar-protein interactions are of critical importance in a wide range of biochemical processes and functions. Complex formation is largely mediated by weak interactions, which give rise to specific association between the guest (carbohydrate) and host (macromolecular receptor) molecule. The hydroxyl groups of carbohydrates can act as both hydrogen-bond donors and acceptors, and form multiple hydrogen-bridges with a receptor structure. Tight hydrogen-bond networks have indeed been identified from crystal structures of carbohydrate-protein complexes (Figure 1A). In addition, partially positively charged carbons of carbohydrate ring systems can interact with π-electron systems and form CH-π interactions between carbohydrate ligands and their macromolecular receptors, typically *via* arene host systems (Figure 1B) [15]. In proteins, the amino acid side chains of Trp, Phe and Tyr are

frequently found to recognize carbohydrate ligands [16, 17]. Using density functional theory (DFT), the average energy contribution of such an interaction has been estimated to be approximately -2.5 kcal × mol^{-1} [18, 19]. Its length is similar to the average hydrogen-bond length (approx. 3 Å) observed in drug-protein interactions [20], with an optimal distance for the CH-π dispersion interaction of approximately 3.7 Å [18].

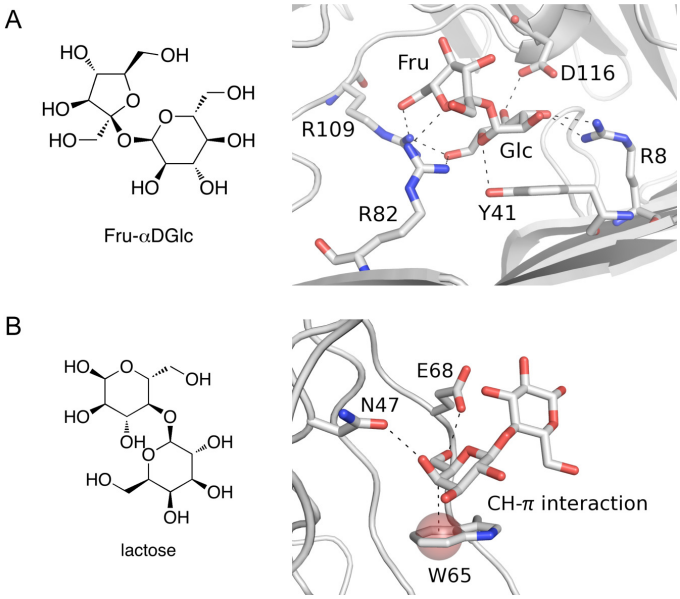


Figure 1. Examples of carbohydrate-protein interactions. **(A)** Hydrogen-bond formation between maltoporin (LamB) and Fru-αDGlc (PDB [21] ID: 1af6 [22]); **(B)** CH-π interaction between human carbohydrate binding protein S-Lac lectin (L-14-II) and lactose (PDB ID: 1hlc [23]).

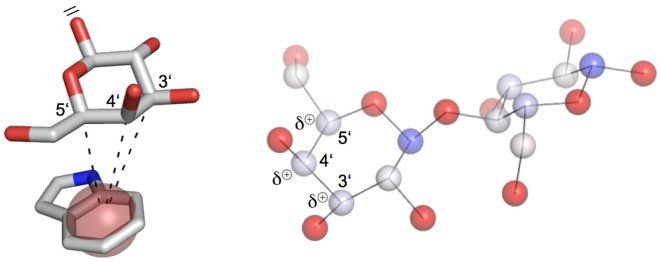


Figure 2. Cartoon representation of the CH-π interaction between the aromatic residue W65 of S-Lac lectin and partially positively charged carbon atoms of lactose (PDB ID: 1hlc). Dashed lines are for illustration only. The right panel represents an energy-minimized conformation of lactose with computed partial charges color coded in blue (positive), white (neutral), and red (negative); color intensity corresponds to the absolute charge value. Charges were computed with MOPAC (http://openmopac.net/) with PM6-DH+ [24] and the COSMO implicit solvation model [25].

Generally, such polar interaction patterns are appealing features for designing drugs with high binding potency and selectivity. If one is interested in comparing the interaction potential ('pharmacophoric potential') of carbohydrates with other drug-like compounds by computational means, these features have to be adequately represented as numerical, typically vectorial, molecular 'descriptors'. Mathematical descriptor comparison enables virtual screening for compounds with similar properties but different architecture ('scaffold-hopping', bioisosteric replacement) and compound clustering.

One such descriptor is termed 'pseudoreceptor', which may be seen as an idealized receptor pocket that is constructed around a ligand conformation or a conformation ensemble. The idea is to place virtual receptor atoms (potential pharmacophoric points, PPP) by applying geometric constraints for receptor-ligand interactions (Figure 3). For example, the geometry of a hydrogen-bridge is defined by the distance r between the donor and acceptor base atom, and the angles α and ω (Figure 3A). Similar definitions are applied for arene-arene, lipophilic and ionic interactions, to name just the most prominent examples [26, 27]. For rapid similarity analysis and database searching the spatial constellation of PPPs may be converted to a linear vector representation. We here used the PRPS pseudoreceptor model [28] and limited the number of PPPs to –OH mediated hydrogen-bridges only, owed to the fact that PRPS in its current version does not consider CH-π interactions.

Briefly, PRPS models were translated into an alignment-free correlation vector representation. Such a pseudoreceptor-derived correlation vector encodes the distance-based frequency of pairs of pseudoatom features present in the model (Eq. 1). Pseudoatom pairs with a distance up to 15 Å (in 1 Å increments) were annotated. This resulted in a vector giving the number *Freq* of pseudoatom pairs with features x and y at distance d.

$$Freq_d(x,y) = \sum_i^x \sum_j^x \delta_d^{i,j}, \text{ where} \tag{1}$$

the Kronecker delta δ computes to 1 whenever a pair of the pseudoatoms i and j exists at distance d. Similarity analysis of monosaccharides was performed by computing the Euclidean distance between their respective pseudoreceptor correlation vectors. This general idea of property encoding for molecular similarity analysis was introduced by Moreau and Broto [29], and has later been adapted to various applications ranging from structural modelling [30] and database searching [31] to molecular *de novo* design [32].

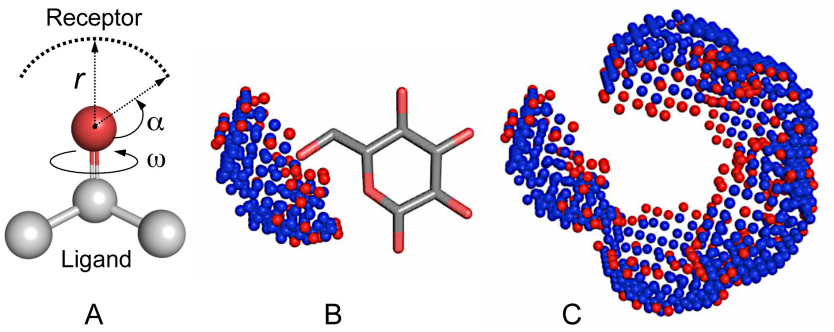


Figure 3. Principle of pseudoreceptor generation. **(A)** Geometric constraints for placing potential hydrogen-bond donor points around a ligand carbonyl oxygen atom, **(B)** Potential hydrogen-bond acceptor (red) and -donor (blue) points placed around a hydroxyl group of glucose, **(C)** A pseudoreceptor model of glucose.

Similarity of Monosaccharides Based on Pseudoreceptor Models

Representing bioactive compounds by their pharmacophores provides the means to sort them according to their interaction potential with macromolecular receptors. We performed such an analysis for 19 selected monosaccharides that are abundant in natural products and form carbohydrate components of bacterial cell walls and glycosylated proteins [33 – 35]. Sugar conformations were used as presented in Figure 4. Prior to PRPS pseudoreceptor generation we subjected all compounds to energy relaxation to eliminate potentially remaining strain. Then, each structure was represented as a vector coding for the corresponding pseudoreceptor model. Subsequent hierarchical clustering of the vectors using Ward's method [36] led to the tree presented in Figure 5A. It is apparent that due to different receptor interaction potential caused by varying numbers of hydroxyl groups the C_5 carbohydrates (orange circles) are separated from C_6 carbohydrates (blue circles). Glucose and deoxyribose occupy prominent positions in the tree, located between the bulk of C_5 and C_6 sugar molecules. One might therefore consider both glucose and deoxyribose as 'universal' monosaccharides with regard to their receptor interaction potential expressed by the pseudoreceptor model. The most distant cluster pairs are (mannose, gulose) and (xylose, arabinose). One might also speculate that while mannose and gulose (xylose and arabinose, respectively) seem exchangeable, a substitution of, e. g., mannose by xylose is structurally or functionally more severe. Such predictions require rigorous chemical and biological testing.

Figure 5B presents a matrix of pairwise similarities between the 19 monosaccharides, again highlighting a potentially special role of glucose and deoxyribose. Such a similarity table could serve as a basis for quantifying the similarity of glycans with regard to potential receptor interactions. In analogy to amino acid substitution matrices [37] this matrix could

be used for generating glycan alignments, which would be complementary to existing similarity schemes based on sugar types, as e. g. produced by the software KCaM (KEGG Carbohydrate Matcher) [38, 39].

We furthermore see a requirement for consideration of CH-π interactions in pharmacophore modelling tools such as our PRPS pseudoreceptor approach. Their critical importance in specific carbohydrate-protein association has been recognized but not fully harvested by bioinformatics and computer-assisted molecular modelling and design. It is likely that the dendrogram and similarity matrix presented in Figure 5 will be subject to partial modification when CH-π interactions are considered as PPPs. In this regard our present study should be considered as preliminary only. More detailed computational investigations are necessary to determine the impact of CH-π interaction modelling on the outcome of molecular similarity analyses.

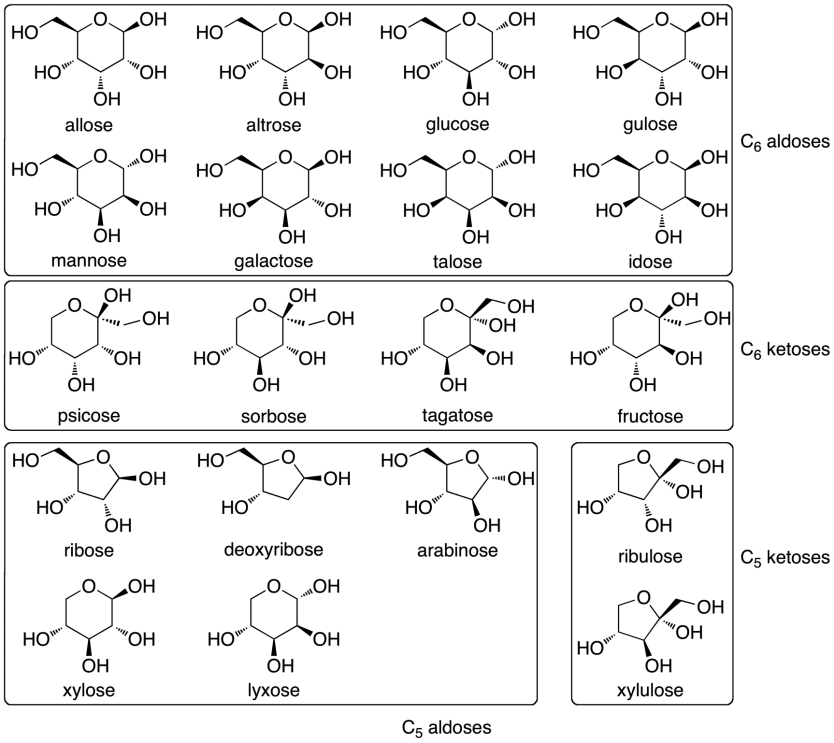


Figure 4. Repertoire of monosaccharides used in this study.

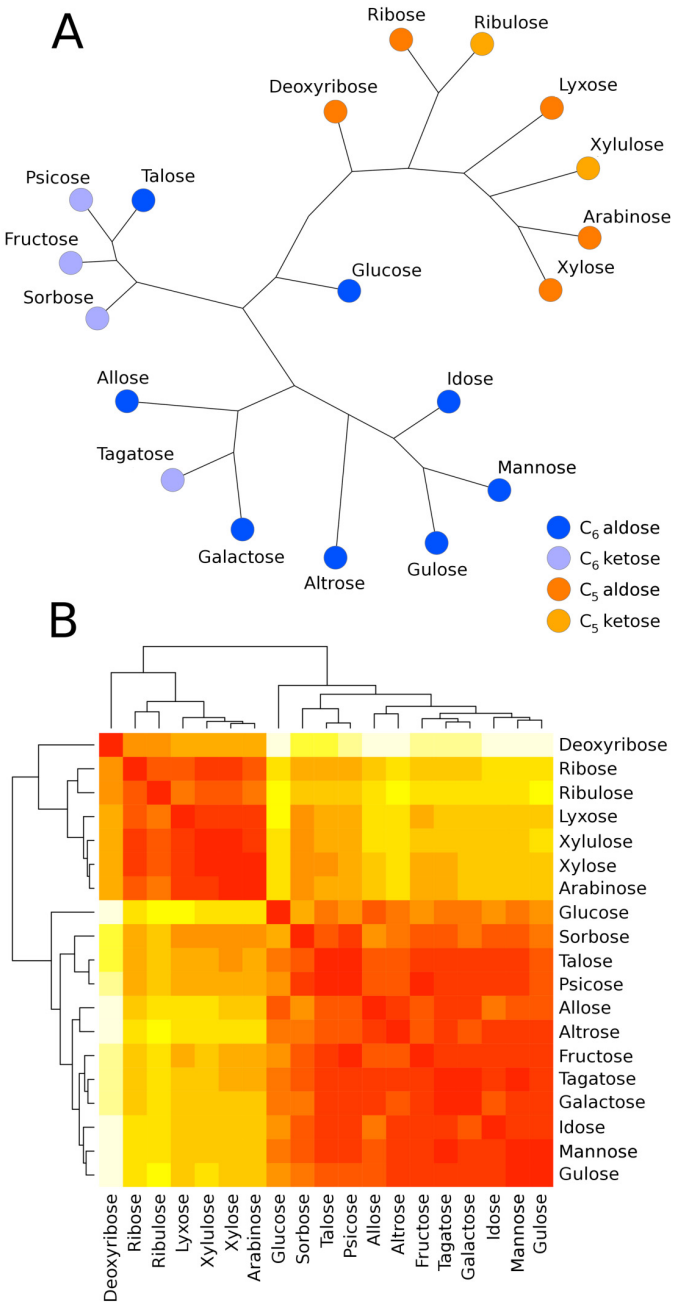


Figure 5. Clustering of carbohydrates represented by PRPS pseudoreceptor models. **(A)** Ward clustering; **(B)** hierarchical tree and similarity analysis (heat map coloring from white to red indicates greater similarity). Carbohydrates considered: β-D-allopyranose, β-D-altropyranose, α-D-arabinofuranose, β-D-desoxyribofuranose, β-D-fructofuranose, α-D-galactopyranose, β-D-glucopyranose, β-D-gulopyranose, β-D-idopyranose, α-D-lyxopyranose, α-D-mannopyranose, β-D-psicopyranose, β-D-ribofuranose, α-D-ribulofuranose, β-D-sorbopyranose, α-D-tagatopyranose, α-D-talopyranose, α-D-xylofuranose, β-D-xylulofuranose.

``Scaffold-hopping'' Using an Ensemble Pseudoreceptor

In a recent drug discovery project [14], we performed pseudoreceptor-based virtual screening to find replacements of aminoglycoside antibiotics that are able to block *in vitro* transcription and translation. This proof-of-concept study demonstrates the applicability of pseudoreceptor representations of carbohydrate derivatives to performing 'scaffold-hops', *i.e.* finding isofunctional compounds with different architecture [40]. The pharmacophore model matched an alternative chemotype, namely the vanilloid derivative **1**, which possesses lower structural complexity and greater synthetic accessibility than the reference aminoglycosides.

The pseudoreceptor was constructed for eight aminoglycosides bound to their 16S RNA target (ribosomal A-site) (Figure 6). Crystal structure conformations of the ligands were aligned to form a conformational ensemble. Then, a database of screening compounds was converted to PRPS vectors (single compound conformations were computed), and compounds were sorted by decreasing similarity to the reference pseudoreceptor model. From the top-ranking candidates, ten compounds were selected, ordered and tested in an *in vitro* transcription/translation assay, five of which exhibited inhibitory activity. The most potent hit **1** is a vanilloid derivative (Scheme 1), with striking similarity to capsaicin (methyl vanillyl nonenamide from chili pepper), which is also known as an antibiotic agent [41]. It is of note that aminoglycoside antibiotics, in turn, also block vanilloid receptor 1 (TRPV1) [42]. These observations corroborate the ensemble pseudoreceptor as a common feature that allows for scaffold replacements, which is relevant not only for the replacement of sugar moieties in case of poor synthetic accessibility, but in particular for identifying carbohydrates and carbohydrate derivatives as building blocks in medicinal chemistry.

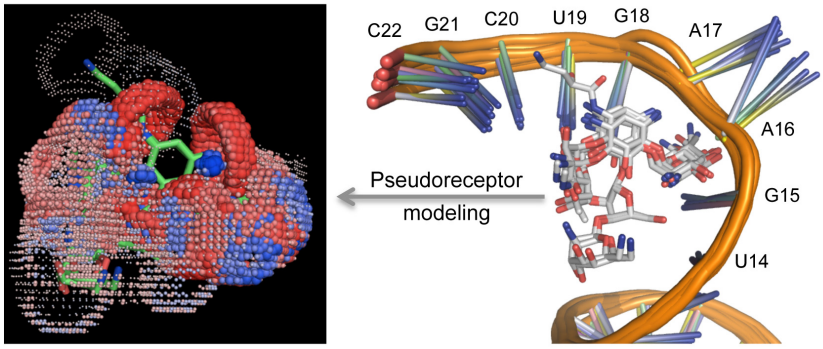
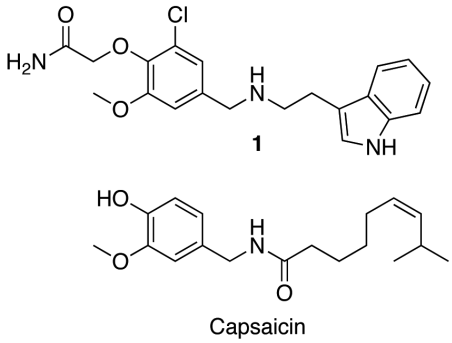


Figure 6. Ensemble pseudoreceptor model (left) generated from the alignment of eight aminoglycosides bound to the ribosomal A-site (right). PDB identifiers: gentamicin C1A (2et3) [43], paromomycin (1j7t) [44], tobramycin (1lc4) [45], neomycin (2et4) [43], neamin (2f4s) [46], kanamycin A (2esi) [43], geneticin (1mwl) [47], amikacin (2g5q) [48].



Scheme 1. Bioactive compound **1** found by pseudoreceptor screening as a replacement for aminoglycoside antibiotics. Compound **1** exhibits close structural similarity to capsaicin.

Conclusions

'Fuzzy' (permissive) pharmacophore representations of molecular structures allow for retrieving alternative chemotypes with lower structural complexity and greater synthetic accessibility than the reference compound(s). This concept was pursued to group monosaccharides according to their pharmacophore similarity. Structures were represented by simplistic pseudoreceptor models, and a monosaccharide similarity matrix was obtained that might be suited as a basis for property-based alignments, virtual compound screening and the design of bioactive compounds. The pseudoreceptor approach used in this preliminary study complements the many existing tools for pharmacophore-based hit and lead structure identification [49]. Without doubt, other molecular representations that allow for abstraction from the atomistic chemical structure and molecular constitution will also be suitable for this purpose, possibly even outperform our specific PRPS implementation. We motivate the use of property- and pharmacophore-based descriptors for carbohydrate modelling, specifically glycan analysis, to identify potential targets (macromolecular receptors) and consider molecular similarity not only by comparing sugar types but also by taking into consideration functionally relevant features. Ideally, virtual screening similarity metrics based on structural fingerprints are combined with metrics that are based on functionally relevant molecular features and properties [50 – 52]. For further progress in this field, CH-π interactions should be added to the set of pharmacophore feature types to more adequately model carbohydrate-target interactions.

Acknowledgements

G.S. is grateful to the Beilstein-Institut for a travel grant and invitation to attend the *2nd Beilstein Symposium on Glyco-Bioinformatics* held in Potsdam in 2011. ETH Zürich and the Deutsche Forschungsgemeinschaft (FOR1406, TP4) are thanked for financial support. The Chemical Computing Group Inc., Montreal, Canada, provided a research license of MOE.

References

[1] Le, G.T., Abbenante, G., Becker, B., Grathwohl, M., Halliday, J., Tometzki, G., Zuegge, J., Meutermans, W. (2003) Molecular diversity through sugar scaffolds. *Drug Discov. Today* **8:**701 – 709.
doi: http://dx.doi.org/10.1016/S1359-6446(03)02751-X.

[2] Cipolla, L., La Ferla, B., Airoldi, C., Zona, C., Orsato, A., Shaikh, N., Russo, L., Nicotra, F. (2010) Carbohydrate mimetics and scaffolds: sweet spots in medicinal chemistry. *Future Med. Chem.* **2:**587 – 599.
doi: http://dx.doi.org/10.4155/fmc.10.8.

[3] Tang, Y.T., Marshall, G.R. (2011) Virtual screening for lead discovery. *Methods Mol. Biol*. **716:**1 – 22.
doi: http://dx.doi.org/10.1007/978-1-61779-012-6_1.

[4] Sotriffer C (ed.) Virtual Screening: Principles, Challenges, and Practical Guidelines. Wiley-VCH, Weinheim (2011).

[5] Schneider, G. (2010) Virtual screening – and endless staircase? *Nat. Rev. Drug Discov.* **9:**273 – 276.
doi: http://dx.doi.org/10.1038/nrd3139.

[6] Köppen, H. (2009) Virtual screening – what does it give us? *Curr. Opin. Drug Discov. Devel*. **12:**397 – 407.

[7] Woods, R.J., Tessier, M.B. (2010) Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Curr. Opin. Struct. Biol*. **20:**575 – 583.
doi: http://dx.doi.org/10.1016/j.sbi.2010.07.005

[8] Roldós, V., Cañada, F.J., Jiménez-Barbero, J. (2011) Carbohydrate-protein interactions: a 3D view by NMR. *ChemBioChem* **12:**990 – 1005; 1293.

[9] Neumann, D., Lehr, C.M., Lenhof, H.P., Kohlbacher, O. (2004) Computational modeling of the sugar-lectin interaction. *Adv. Drug Deliv. Rev.* **56:**437 – 457. doi: http://dx.doi.org/10.1016/j.addr.2003.10.019.

[10] Ertl, P., Schuffenhauer, A. (2008) Cheminformatics analysis of natural products: lessons from nature inspiring the design of new drugs. *Prog. Drug Res.* **66:**217, 219 – 235.

[11] Grabowski, K., Baringhaus, K.H., Schneider, G. (2008) Scaffold diversity of natural products: inspiration for combinatorial library design. *Nat. Prod. Rep.* **25:**892 – 904. doi: http://dx.doi.org/10.1039/b715668p.

[12] Aoki-Kinoshita, K.F. (2008) An introduction to bioinformatics for glycomics research. *PLoS Comput Biol.* **4:**e1000075. doi: http://dx.doi.org/10.1371/journal.pcbi.1000075.

[13] Tanrikulu, Y., Schneider, G. (2008) Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nat. Rev. Drug Discov.* **7:**667 – 677. doi: http://dx.doi.org/10.1038/nrd2615.

[14] Urbanek, D.A., Proschak, E., Tanrikulu, Y., Becker, S., Karas, M., Schneider, G. (2011) Scaffold-hopping from aminoglycosides to small synthetic inhibitors of bacterial protein biosynthesis using a pseudoreceptor model. *Med. Chem. Commun*. **2/3:**181 – 184. doi: http://dx.doi.org/10.1039/c0md00207k.

[15] Muraki, M. (2002) The importance of CH/pi interactions to the function of carbohydrate binding proteins. *Protein Pept. Lett*. **9:**195 – 209. doi: http://dx.doi.org/10.2174/0929866023408751.

[16] Drickamer, K. (1997) Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* **5:**465 – 468. doi: http://dx.doi.org/10.1016/S0969-2126(97)00202-5.

[17] Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J., Hilgenfeld, R. (2001) CH-pi-interactions in proteins. *J. Mol. Biol*. **307:**357 – 377. doi: http://dx.doi.org/10.1006/jmbi.2000.4473.

[18] Raju, R.K., Ramraj, A., Vincent, M.A., Hillier, I.H., Burton, N.A. (2008) Carbohydrate-protein recognition probed by density functional theory and *ab initio* calculations including dispersive interactions. *Phys. Chem. Chem. Phys*. **10:**6500 – 6508. doi: http://dx.doi.org/10.1039/b809164a.

[19] Kozmon, S., Matuška, R., Spiwok, V., Koča, J. (2011) Three-dimensional potential energy surface of selected carbohydrates' CH/π dispersion interactions calculated by high-level quantum mechanical methods. *Chemistry* **17:**5680 – 5690. doi: http://dx.doi.org/10.1002/chem.201002876.

[20] Liu, Z., Wang, G., Li, Z., Wang, R. (2008) Geometrical preferences of the hydrogen bonds on protein-ligand binding interface derived from statistical surveys and quantum mechanics calculations. *J. Chem. Theory Comput*. **4:**1959 – 1973. doi: http://dx.doi.org/10.1021/ct800267x.

[21] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28:**235 – 242. doi: http://dx.doi.org/10.1093/nar/28.1.235.

[22] Wang, Y.F., Dutzler, R., Rizkallah, P.J., Rosenbusch, J.P., Schirmer, T. (1997) Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin. *J. Mol. Biol.* **272:**56 – 63. doi: http://dx.doi.org/10.1006/jmbi.1997.1224.

[23] Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., Rini, J.M. (1993) X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-Å resolution. *J. Biol. Chem.* **268:**27034 – 27038.

[24] Korth, M. (2010) Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields. *J. Chem. Theory Comput*. **6:**3808 – 3816. doi: http://dx.doi.org/10.1021/ct100408b.

[25] Klamt, A., Schüürmann, G. (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans.* **2:**799 – 805. doi: http://dx.doi.org/10.1039/p29930000799.

[26] Böhm, H.J. (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des*. **6:**61 – 78. doi: http://dx.doi.org/10.1007/BF00124387.

[27] Bissantz, C., Kuhn, B., Stahl, M. (2010) A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **53:**5061 – 5084, 6241. doi: http://dx.doi.org/10.1021/jm100112j.

[28] Tanrikulu, Y., Proschak, E., Werner, T., Geppert, T., Todoroff, N., Klenner, A., Kottke, T., Sander, K., Schneider, E., Seifert, R., Stark, H., Clark, T., Schneider, G. (2009) Homology model adjustment and ligand screening with a pseudoreceptor of the human histamine H4 receptor. *ChemMedChem* **4:**820 – 827. doi: http://dx.doi.org/10.1002/cmdc.200800443.

[29] Moreau, G., Broto, P. (1980). The autocorrelation of a topological structure: A new molecular descriptor. *Nouv. J. Chim*. **4:**359 – 360.

[30] Bauknecht, H., Zell, A., Bayer, H., Levi, P., Wagener, M., Sadowski, J., Gasteiger, J. (1996) Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists. *J. Chem. Inf. Comput. Sci.* **36:**1205 – 1213.
doi: http://dx.doi.org/10.1021/ci960346m.

[31] Fechner, U., Franke, L., Renner, S., Schneider, P., Schneider, G. (2003) Comparison of correlation vector methods for ligand-based similarity searching. *J. Comput. Aided Mol. Des.* **17:**687 – 698.
doi: http://dx.doi.org/10.1023/B:JCAM.0000017375.61558.ad.

[32] Schneider, G., Clément-Chomienne, O., Hilfiger, L., Schneider, P., Kirsch, S., Böhm, H.J., Neidhart, W. (2000) Virtual screening for bioactive molecules by evolutionary *de novo* design. *Angew. Chem. Int. Ed. Engl*. **39:**4130 – 4133.
doi: http://dx.doi.org/10.1002/1521-3773(20001117)39:22<4130::AID-ANIE4130>3.0.CO;2-E.

[33] Caffall, K.H., Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr. Res.* **344:**1879 – 900.
doi: http://dx.doi.org/10.1016/j.carres.2009.05.021.

[34] Bazaka, K., Crawford, R.J., Nazarenko, E.L., Ivanova, E.P. (2011) Bacterial extracellular polysaccharides. *Adv. Exp. Med. Biol*. **715:**213 – 226.
doi: http://dx.doi.org/10.1007/978-94-007-0940-9_13.

[35] Mariño, K., Bones, J., Kattla, J.J., Rudd, P.M. (2010) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **6:**713 – 723.
doi: http://dx.doi.org/10.1038/nchembio.437.

[36] Ward, J.H. (1963) Hierachical grouping to optimize an objective function. *J. Am. Statist. Assoc.* **58:**236 – 244.
doi: http://dx.doi.org/10.2307/2282967.

[37] Henikoff, S., Henikoff, J.G. (2000) Amino acid substitution matrices. *Adv. Protein Chem*. **54:**73 – 97.
doi: http://dx.doi.org/10.1016/S0065-3233(00)54003-0.

[38] Aoki, K.F., Yamaguchi, A., Ueda, N., Akutsu, T., Mamitsuka, H., *et al.* (2004) KCaM (KEGG Carbohydrate Matcher): A software tool for analyzing the structures of carbohydrate sugar chains. *Nucleic Acids Res*. **32:**W267 –W272.
doi: http://dx.doi.org/10.1093/nar/gkh473.

[39] Aoki, K.F., Mamitsuka, H., Akutsu, T., Kanehisa, M. (2005) A score matrix to reveal the hidden links in glycans. *Bioinformatics* **21:**1457 – 1463.
doi: http://dx.doi.org/10.1093/bioinformatics/bti193.

[40] Schneider, G., Schneider, P., Renner, S. (2006) Scaffold-hopping: how far can you jump? *QSAR Comb. Sci.* **25:**1162 – 1171.
doi: http://dx.doi.org/10.1002/qsar.200610091.

[41] Ozçelik, B., Kartal, M., Orhan, I. (2011) Cytotoxicity, antiviral and antimicrobial activities of alkaloids, flavonoids, and phenolic acids. *Pharm. Biol*. **49:**396 – 402.
doi: http://dx.doi.org/10.3109/13880209.2010.519390.

[42] Raisinghani, M., Premkumar, L.S. (2005) Block of native and cloned vanilloid receptor 1 (TRPV1) by aminoglycoside antibiotics. *Pain* **113:**123 – 133.
doi: http://dx.doi.org/10.1016/j.pain.2004.09.042.

[43] François, B., Russell, R.J., Murray, J.B., Aboul-ela, F., Masquida, B., Vicens, Q., Westhof, E. (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res*. **33:**5677 – 5690.
doi: http://dx.doi.org/10.1093/nar/gki862.

[44] Vicens, Q., Westhof, E. (2001) Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* **9:**647 – 658.
doi: http://dx.doi.org/10.1016/S0969-2126(01)00629-3.

[45] Vicens, Q., Westhof, E. (2002) Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding a site. *Chem. Biol*. **9:**747 – 755.
doi: http://dx.doi.org/10.1016/S1074-5521(02)00153-9.

[46] Murray, J.B., Meroueh, S.O., Russell, R.J., Lentzen, G., Haddad, J., Mobashery, S. (2006) Interactions of designer antibiotics and the bacterial ribosomal aminoacyl-tRNA site. *Chem. Biol.* **13:**129 – 138.
doi: http://dx.doi.org/10.1016/j.chembiol.2005.11.004.

[47] Vicens, Q., Westhof, E. (2003) Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide. *J. Mol. Biol*. **326:**1175 – 1188.
doi: http://dx.doi.org/10.1016/S0022-2836(02)01435-3.

[48] Kondo, J., François, B., Russell, R.J., Murray, J.B., Westhof, E. (2006) Crystal structure of the bacterial ribosomal decoding site complexed with amikacin containing the gamma-amino-alpha-hydroxybutyryl (haba) group. *Biochimie* **88:**1027 – 1031.
doi: http://dx.doi.org/10.1016/j.biochi.2006.05.017.

[49] Langer, T. (2010) Pharmacophores in drug research. *Mol. Inf.* **29:**470 – 475. doi: http://dx.doi.org/10.1002/minf.201000022.

[50] Willett, P. (2011) Similarity searching using 2D structural fingerprints. *Methods Mol. Biol.* **672:**133 – 158. doi: http://dx.doi.org/10.1007/978-1-60761-839-3_5.

[51] Zhou, J.Z. (2011) Chemoinformatics and library design. *Methods Mol. Biol.* **685:**27 – 52. doi: http://dx.doi.org/10.1007/978-1-60761-931-4_2.

[52] Klenner, A., Hartenfeller, M., Schneider, P., Schneider, G. (2010) 'Fuzziness' in pharmacophore-based virtual screening and *de novo* design. *Drug Discov. Today Technol*. **7:**e237-e244. doi: http://dx.doi.org/10.1016/j.ddtec.2010.10.004.