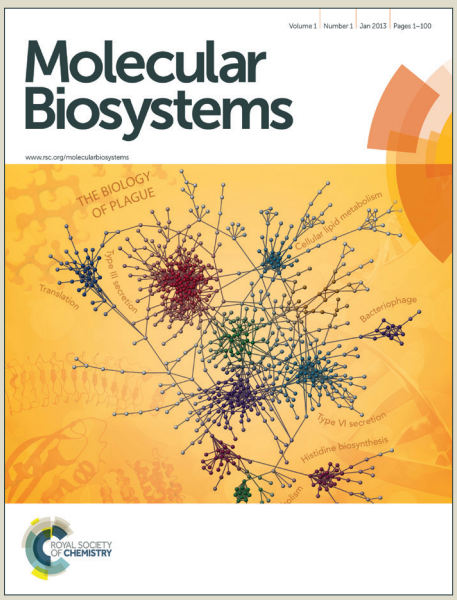
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

A Systematic Study of Chemogenomics for Carbohydrates†‡

Jiangyong Gu, Fang Luo, Lirong Chen,[*a] Gu Yuan, Xiaojie Xu[*a]

Abstract

Chemogenomics focuses on the interactions between biologically active molecules and protein targets for drug discovery. Carbohydrates are the most abundant compounds in natural products. Comparing with other drugs, the carbohydrate drugs show weaker side effects. Searching for multi-target carbohydrate drugs can be regarded as a solution to improve therapeutic efficacy and safety. In this work, we collected 60344 carbohydrates from the Universal Natural Products Database (UNPD) and explored the chemical space of carbohydrates by principal component analysis. We found that there is a large quantity of potential lead compounds among carbohydrates. Then we explored the potential of carbohydrates in drug discovery by using a network-based multi-target computational approach. All carbohydrates were docked to 2389 target proteins. The most potential carbohydrates for drug discovery and their indications were predicted based on a docking score-weighted prediction model. We also explored the interactions between carbohydrates and target proteins to find the pathological networks, potential drug candidates and new indications.

Introduction

Chemogenomics is used to predict interactions between biologically active molecules and protein targets[1, 2]. Based on the assumption that similar ligands bind similar receptors[2], knowledge on the protein-ligand interactions can be used to identify novel ligands for a given target or a novel target for a given ligand.

Carbohydrates are the most abundant compounds in natural products. They are mainly distributed in the surface of cell and involved in many important biological recognition processes such as cell-cell communication, bacterial adhesion, viral infection and masking of immunological epitopes[3-5]. Although carbohydrates play important roles in a lot of recognition processes, there are relatively few carbohydrates or carbohydrate-derived drugs in the area of therapeutics. One main reason is that their inherent drawbacks of poor pharmacokinetic properties[6]. Due to their high polarity, it is difficult to cross the enterocyte layer in the small intestine, which is a prerequisite for oral availability.

Though their inherent drawbacks, carbohydrates are still offering exciting new therapeutic opportunities. For example, carbohydrates have strong target specificity owing to the important role in cellular recognition processes. It means that carbohydrates can help drugs arrive the target accurately and act properly[7]. Moreover, carbohydrates or carbohydrate-derived small molecules can be effective drugs by modify their structures to make them more "drug-like"[8]. With the development of this new class of small-molecule drugs, which are called glycomimetics[9], these compounds can mimic the bioactive function of carbohydrates by improving the poor pharmacokinetic properties. In addition, a few FDA-approved drugs contain carbohydrate moieties as part of their structures. And generally, removal of the sugar eliminates the therapeutic effect of the drug[7]. Also due to the active site of interaction between carbohydrate and target is almost on the surface of the cell, comparing with other drugs, the carbohydrate drugs have weaker side effects.

Over the past decades, the investment by the global pharmaceutical industries on drug research and development has increased heavily, but the number of new drugs approved has significantly declined[10]. The main reason can be attributed to the side effects or toxicity of the candidate drug compounds observed in clinical trials[11, 12]. In recent

years, more and more evidences have shown that many drugs can exert their activities by modulating multi-targets[13, 14]. Inappropriate modulation could induce strong side effects[15, 16]. Therefore, searching for multi-target drugs can be regarded as one of the solutions to improve therapeutic efficacy and safety.

However, the rational design and searching of multi-target drugs is still a considerable challenge for pharmaceutical industry. These mainly attribute to the new methods which need to consider target combinations and to identify important lead compounds. The combinational effect could be larger than the sum of individual effects. But the number of possible combinations can increase exponentially, which makes it very limiting and expensive to validate by experimental approach[17]. In this condition, computational methods that can deal with lots of data provide a more promising and desirable strategy[18]. With the help of powerful computer we can predict a large number of new drug-target interactions by constructing drug-target networks to explore the mechanism behind the combinations at the molecular level.

These recent developments of computational methods resulted in network pharmacology, which was a novel paradigm with potential to provide more global comprehension of drug action, including drug resistance and side effects in the context of biological networks and pathways[19, 20]. The network-based representation and analysis seemed very valid in complex diseases, as it can provide new therapeutic views for drug repositioning[21]. It can also be used to design drugs based on network targeting[22].

Recently, lots of studies have been performed to analyze multi-target drug discovery by network-based methods[22-27]. For example, Li *et al*. used a biological network-based multi-target computational estimation scheme to evaluate the anticoagulant activity of a series of argatroban intermediates[28]. Others revealed the interactions between proteins and ligands, or predicted the potential drug candidate compounds and targets[29, 30].

We hereby explored the interactions between carbohydrates and target proteins to find the pathological network, potential drug candidates and new indications. The carbohydrates here included all kinds of sugars, the compounds which contained carbohydrate moieties as part of their structures, glycomimetic and carbohydrate-like compounds whose hydroxyl was substituted by other group.

In 2002, our lab established a 3D structure database of components from Chinese traditional medicinal herbs[31]. And recently we constructed the Universal Natural Products Database (UNPD) which contained 208213 natural products from plants, animals and microorganisms[32]. Here, based on the largest non-commercial and freely available database, we collected 60344 carbohydrates through searching the fragments of carbohydrate from UNPD. We also explored the potential of carbohydrates in drug discovery by using a network-based multi-target computational approach. These attempts may offer new opportunities to understand the pharmacological properties of carbohydrate compounds and provide benefit for drug discovery from carbohydrates for treating complex diseases.

Methods

1. Collection of Carbohydrates and Approved Drugs

The carbohydrates were collected from UNPD through searching the fragments of carbohydrate. Carbohydrate can exist in either a straight-chain or ring form. Moreover, the ring form sugars such as glucose and ribose play important roles in cellular recognition processes and most of the FDA-approved carbohydrate-based drug are derived from pentose and hexose[33]. Therefore, most of the fragments we chose are the ring form pentose, hexose and heptose or their analogs. Then, we removed the duplicates. We also collected the straight-line chain carbohydrate by structure similarity searching and set the tanimoto coefficient to 0.5. The number of carbohydrates and the fragments were listed in File S1 (ESI‡). We used the absolute configuration for each compound and deleted the salts or adducts. Finally, we found 60344 carbohydrates in total. All of the structures of carbohydrates were minimized in MMFF94 force field. The structures of FDA-approved drugs were downloaded from

DrugBank[34].

2. Calculation of Molecular Descriptors of Carbohydrates

The carbohydrates and FDA-approved drugs were screened by "rule of five"[35]. And there were 11751 out of 60344 compounds which obeyed the "rule of five". For carbohydrates and FDA-approved drugs, we calculated AlogP, molecular weight (MW), Number of hydrogen bond donors (NHBD), number of hydrogen bond acceptors (NHBA), number of rotatable bonds (NRB), molecular volume (MV) and molecular surface area (MSA) in Discovery Studio (Fig. 1 and Table 1).

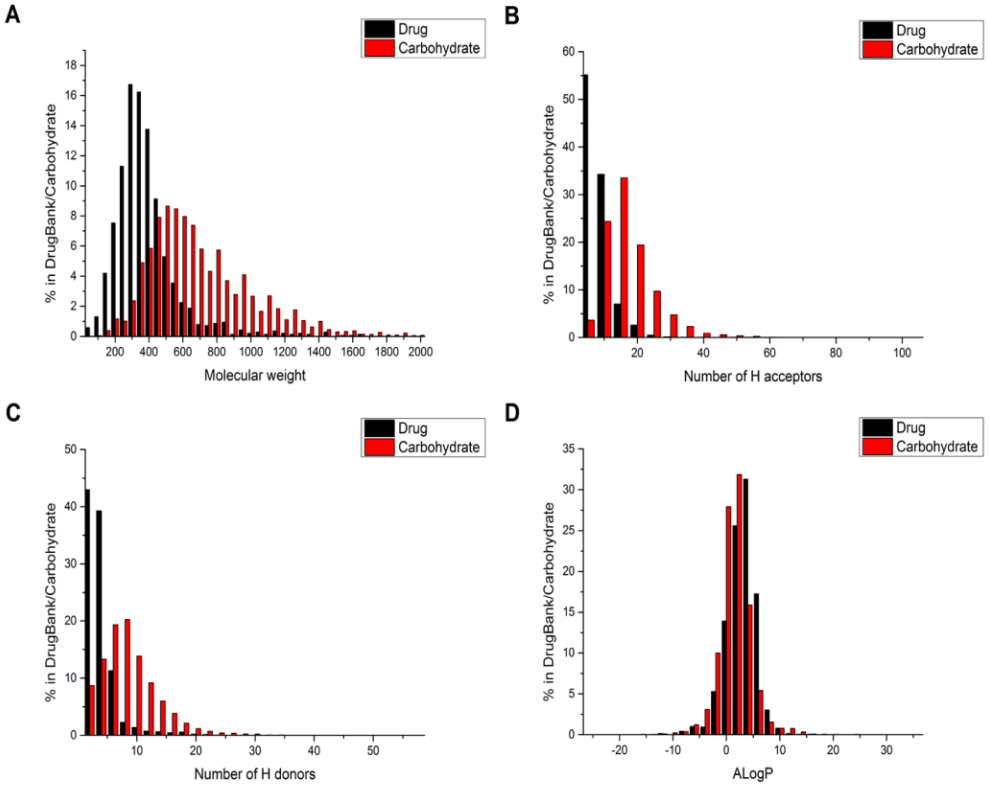


Fig. 1 Distribution of four molecular descriptors of carbohydrates and approved drugs.

Table 1. Statistics of molecular descriptors of carbohydrates and FDA-approved drugs.

Descriptors	Carbohydrates				Approved drugs			
	Mean	Median	Min	Max	Mean	Median	Min	Max
AlogP	0.498±3.212	0.374	-26.173	34.758	1.899±2.814	2.164	-12.843	14.242
MW	682.8±324.8	610.5	58.1	3760.7	360.8±199.1	322.1	6.9	1639.9
NRB	9.7±7.5	9	0	102	5.5±4.9	4	0	44
NHBA	14.3±7.9	12	1	104	5.2±4.2	4	0	51
NHDA	7.1±4.9	6	0	57	2.3±2.6	2	0	23
MV	449.2±212.4	393.4	39.8	2076.5	238.0±128.1	217.8	6.8	1053.3
MSA	647.3±304.4	570.3	71.4	3242.1	347.7±186.4	312.5	16.6	1586.9

3. Chemical Space Analysis

Principal component analysis was employed in library analysis module of Discovery Studio to compare the structural diversity between carbohydrates and drugs. The variance of the data called first principal component was maximized on the first coordinate. The rest of it on the second coordinate, and so on. We built the principal component analysis model with the above seven descriptors. The variances of PC1, PC2 and PC3 for carbohydrates and drugs in Fig. 2 were 0.486, 0.236, and 0.132, respectively.

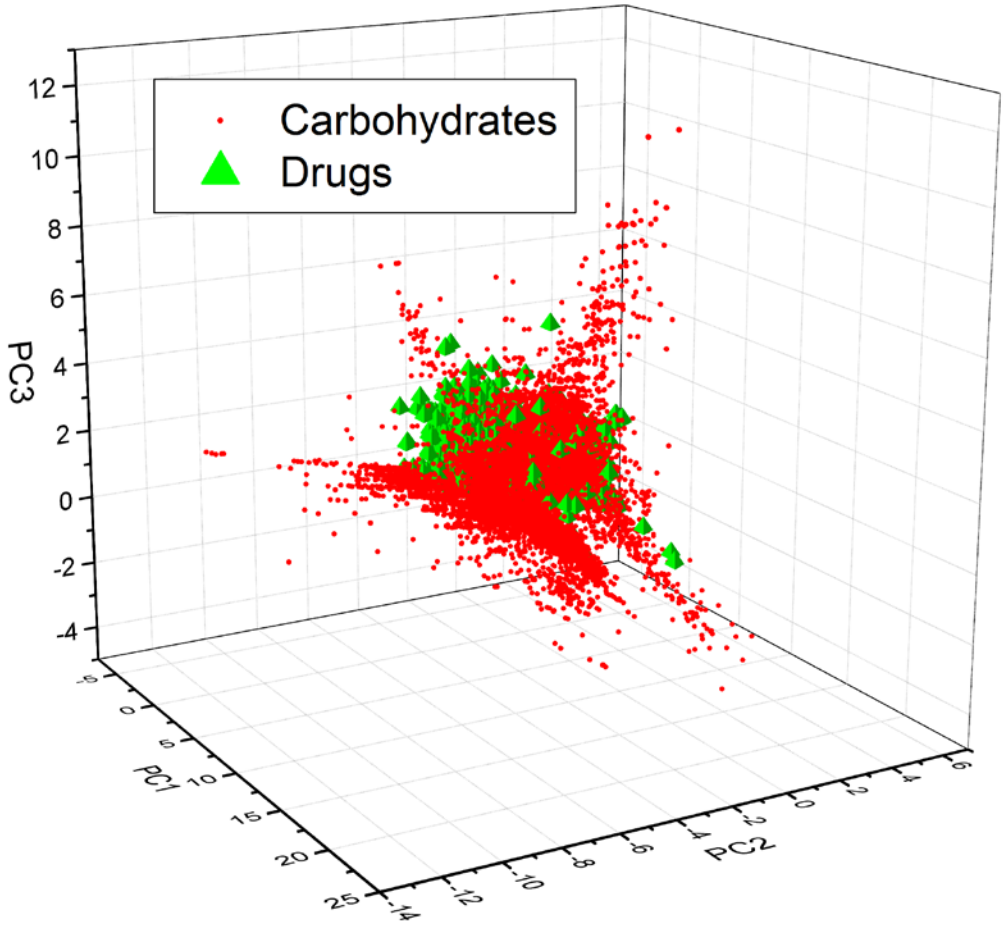


Fig. 2 The distribution in chemical space according to principal component analysis of carbohydrates and FDA-approved drugs. The red dots and green triangles represent carbohydrates and FDA-approved drugs, respectively.

4. Construction of Drug-Target Network Based on Docking Data

We used the crystal or NMR structures of 2389 target proteins in RCSB Protein Data Bank to screen potential lead compounds. All of the downloaded structures were protein-ligand complexes, which were target proteins of approved drugs in DrugBank or other human proteins. The hetero atoms were removed and the hydrogen atoms were added in Discovery Studio. The original ligands of complex structures were used to define the active site and as reference compounds to compare the affinity of carbohydrates to targets accordingly. The docking was performed by autodock4.01[36] in DOVIS 2.0[37], and parameters were listed in File S1 (ESI‡). The binding site was defined as a 40×40×40 Å cube centered on the space which the original ligand occupied with a spacing of 0.375 Å. In order to improve the accuracy of predicted results and make data processing convenient, we selected the molecules that the docking score was higher than 9 and higher than that of the original ligand of complex structure to construct the drug-target network. The drug-target network based on docking result was constructed in

Cytoscape[38]. Also, we used the network analysis plugin to calculate the network properties and node centralities. To find the potential lead-like and drug-like molecules, here we constructed two drug-target networks: one was based on the docking data of carbohydrate molecules which obeyed the "rule of five" and the other was based on the docking data of all carbohydrates.

Results and Discussion

1. Statistics of Molecular Descriptors of Carbohydrates and FDA-approved Drugs

Seven important molecular descriptors of carbohydrates in UNPD and FDA-approved drugs in DrugBank were listed in Table 1. Obviously, except for AlogP, the statistical means and standard deviations of other descriptors of carbohydrates were larger than those of FDA-approved drugs, so the more diverse chemical structures of natural products would provide more polypharmacology through interacting with multiple target proteins[39].

Lipinski's "rule of five"[35] was derived from statistic data of oral drugs and was often used in virtual screening from large compound libraries. The content of it was in general, an orally active drug has no more than one violation of the following criteria: not more than 5 hydrogen bond donors, not more than 10 hydrogen bond acceptors, molecular mass less than 500 Daltons and octanol-water partition coefficient log P not greater than 5.

We checked the satisfied condition for "rule of five" of carbohydrates and found that only 11751 (19.5%) out of 60344 obeyed the rules while 1065 drugs (77.2%) out of 1380 obeyed the rules. The big difference was owing to the structural features of carbohydrate: polyhydroxy meant that it was very soluble in water and difficult to be satisfied with the first three rules.

We compared carbohydrate molecules with FDA-approved drugs in properties of "rule of five" in Fig. 1. The histograms of the distribution of each descriptor showed that as for the molecular weight, only less than 50% molecules was in the range between 0 to 500, also the distribution of NHBA and NHBD of most carbohydrate were not in the range of "rule of five". The log P was an exception, most values of the log P in carbohydrate were between -5 and 5, which was similar to that of drugs. Due to the large difference of distribution in MW, NHBA and NHBD between carbohydrates and drugs, we chose the molecules which obeyed the "rule of five" to explore their network pharmacology.

2. Distribution of Drug-like Chemical Space

The drug-like chemical space is important for drug discovery[40-42]. To get a better understanding of carbohydrates and FDA-approved drugs, we used principal component analysis to give visual illustration in chemical space. The 3D plot in Fig. 2 provided the distribution information clearly. It showed that carbohydrates had vast diversity. From Fig. 2 we also can see that there were obvious overlaps between the two molecular datasets in chemical space, which indicated that the carbohydrates contained lots of drug-like compounds. And some carbohydrate compounds may have desired drug-like properties.

3. Network Pharmacology Based on All Carbohydrates

Network pharmacology was proposed by Hopkins in 2007[43] and it could use network analysis methods to explore the interaction of molecules in biological networks. It can help us understand the interaction mechanism and predict the drug efficacy[44]. It is thought to be an important and potential method to find and develop multi-target drugs in drug discovery[45].

By using Autodock4 all carbohydrates were docked to 2389 targets and screened according to docking score. Then we constructed the drug-target network (Fig. 3). The network showed that most carbohydrate molecules targeted at one or two target proteins. In a network, an "edge" was an association, interaction, or any other well-defined relationship. The degree of a node was the number of edges connected to it. The betweenness centrality of a node was a measure of a node's importance in a network. It was equal to the number of shortest paths from all vertices to all others that pass through that node. Degree and betweenness centrality were two

important parameters to assess the node in the network. For example, the nodes with high betweenness centrality can be regarded as key nodes of a network. These network parameters also can be used to measure directly the importance of proteins or molecules.

There were several carbohydrates which have many targets, such as UNPD2675 (179 targets), UNPD119313 (136 targets), UNPD95242 (125 targets) and UNPD52311 (103 targets). UNPD2675 (seldomycin factor 5) was one of the most active aminoglycoside antibiotics isolated from fermentation broth of *Streptomyces hofunensis* by using of a cationic exchange resin[46]. UNPD119313 was also an aminoglycoside antibiotic.

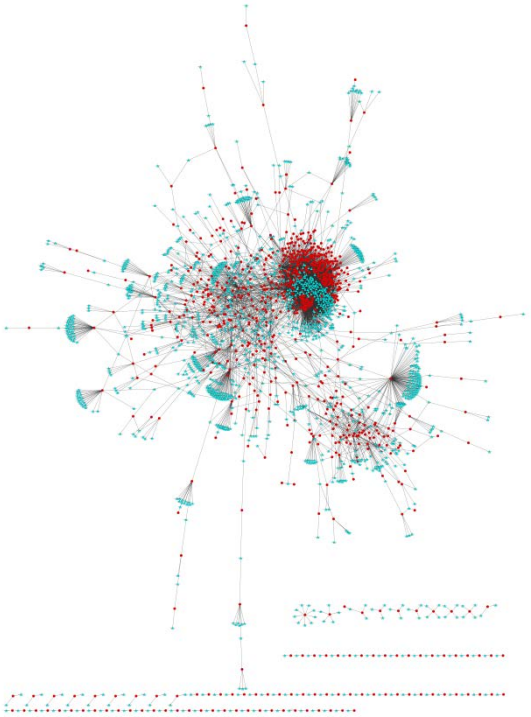


Fig. 3 Drug-target network of carbohydrates and their computational targets. Red circles and cyan triangles correspond to target proteins and carbohydrates, respectively.

The drug-target network contained 2386 nodes and 9931 edges, in which cyan nodes represented potential drug compounds and red notes represented potential targets (Fig. 3.). Table 2 listed the degree and betweenness of the candidate compounds. The degree of nodes provided an opportunity for us to find highly connected molecules or proteins which may play important roles in the drug-target interaction network.

Table 2 Chemical information and network parameters of carbohydrates.

UNPD ID	Name	CAS NO.	Degree	Betweenness
UNPD2675	Seldomycin 5	56276-26-7	179	0.07112299
UNPD119313	N/A	N/A	136	0.03954833
UNPD95242	Mutamicin 5	55750-88-4	125	0.02347492
UNPD52311	Gentamicin V-2	60352-78-5	103	0.01177308
UNPD41602	Seldomycin 2	54333-78-7	89	0.01822394
UNPD65790	Fortimicin E	71772-09-3	84	0.01757209
UNPD174494	Mutamicin 2	54830-48-7	80	0.00606319
UNPD70326	Mutamicin 1b	54830-51-2	72	0.01369594
UNPD59628	Sannamycin J	83997-42-6	60	0.00500115
UNPD123829	Neomycin C	66-86-4	50	0.0021831
UNPD6619	Fortimicin AN	74918-34-6	40	0.00152173

4. Network Pharmacology Based on the Carbohydrates Which Obey the "Rule of Five"

The above drug-target network provided the information of all carbohydrates and predicted several compounds which may be good candidates of lead-like carbohydrates in drug discovery. Due to the high polarity of carbohydrates, the drug-like candidates were screened by the "rule of five". We used the molecules which obeyed the "rule of five" to construct a new drug-target network and expected to find more drug-like candidates. The drug-target network and candidates were shown in Fig. 4 and Table 3.

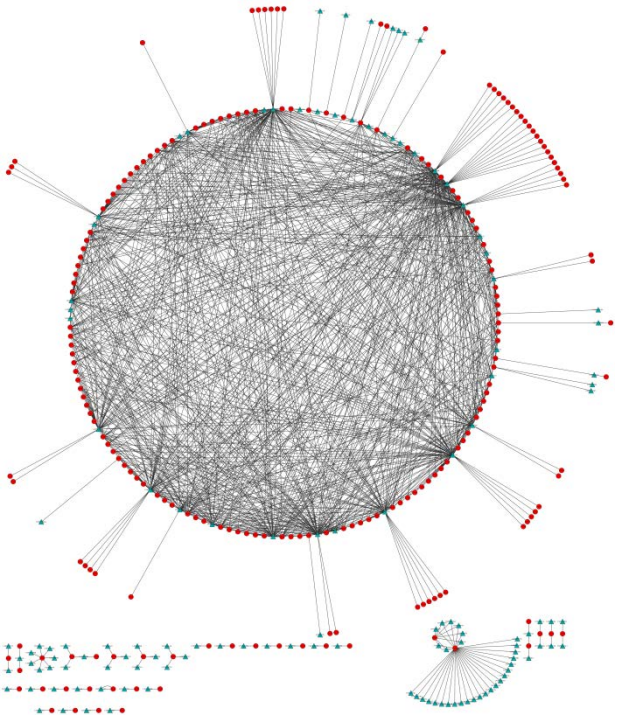


Fig. 4 Drug-target network of carbohydrates which obey the rule of five and their computational targets. Representations of the symbols are the same with Fig. 3.

The network (Fig. 4) contained 339 nodes and 1011 edges. The centralization and heterogeneity analysis showed the network centralization and heterogeneity are 0.197 and 1.871, respectively, indicating that a few nodes were more central than the others in this network, i.e., the drug–target space was biased toward certain compounds and proteins. Such as UNPD141357, UNPD179573 and UNPD78875, degree of which was 72, 67 and 66, respectively. High degree meant that these compounds can interact with more targets. The degree and betweenness of candidates were listed in Table 3.

Table 3. Chemical information and network parameters of carbohydrates.

UNPD ID	Name	CAS NO.	Degree	Betweenness
UNPD141357	Antibiotic KA 6606 XIV	81749-22-6	72	0.15561437
UNPD179573	Sporaricin B	68743-78-2	67	0.11885893
UNPD78875	Antibiotic KA 6606 V	75829-53-7	66	0.16812401
UNPD190270	Antibiotic KA 6606 XVIII	88595-56-6	65	0.11339506
UNPD59628	Sannamycin J	83997-42-6	60	0.08723644
UNPD27668	Antibiotic KA 6606 VIII	81768-56-1	59	0.11335026
UNPD166516	Antibiotic KA 6606 XIX	88643-89-4	50	0.08333396
UNPD191985	Antibiotic KA 6606 XIII	81749-23-7	46	0.06427514
UNPD77950	Sannamyicin C	73522-71-1	46	0.04051971
UNPD155286	Istamycin A0	72503-80-1	44	0.06905324
UNPD143900	Istamycin C0	83860-42-8	37	0.03515297
UNPD27844	Istamycin B0	82443-85-4	32	0.02465623
UNPD83155	Istamycin B	72523-63-8	28	0.01281024
UNPD14597	Sannamycin A	72503-79-8	27	0.02075767
UNPD56972	Calystegine N1	NOT AVAILABLE	17	0.05119082
UNPD115521	Sporaricin A	68743-79-3	16	0.00172078

5. Predicted Diseases for Carbohydrates

Carbohydrate drugs have been used to treat some diseases for many years, such as diabetes, influenza virus infections, thrombosis, Gaucher's disease and osteoarthritis[6]. There were a few FDA-approved drugs derived from carbohydrates. Here we predicted the potential carbohydrates based on above networks which could be good lead candidates for diseases. Generally, the compounds could interact with several target proteins which would relate to a few diseases. We constructed a docking score-weighted prediction network model to predict the relationship between the carbohydrates and some diseases in Fig. 5. T is the set of targets related to a disease. The prediction coefficient for each carbohydrate was listed in Table 4 and File S1 (ESI‡).

$$\text{Prediction coefficient} = \sum_{i \in T} score_i$$

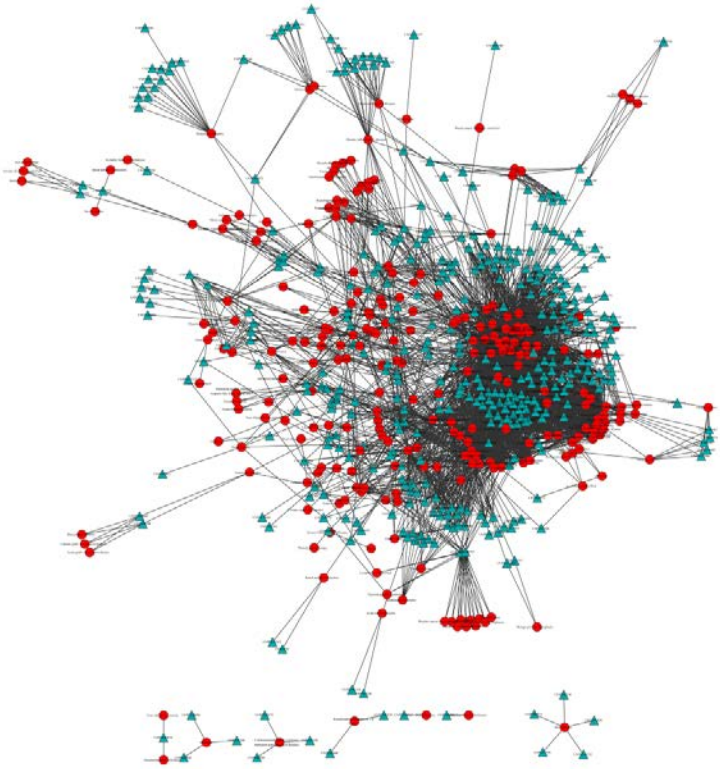


Fig. 5 Potential compound-potential disease network. Red circles and cyan triangles correspond to diseases and carbohydrates, respectively.

Table 4 indicated that most of the predicted compounds can interact with the target proteins of asthma and rheumatoid arthritis. UNPD2675 had a high prediction coefficient and it maybe a good lead candidate for Type II diabetes mellitus. UNPD2675 and UNPD137471 would have large possibility as drugs for asthma and Type II diabetes mellitus. The top rank carbohydrates which had high prediction coefficient for diseases were shown in Table 4 and Fig. 6.

Table 4. Prediction coefficient of carbohydrates for diseases.

Carbohydrates	Prediction coefficient	Diseases
UNPD2675	48.5	Type II diabetes mellitus
UNPD2675	45.77	Asthma
UNPD137471	42.85	Asthma
UNPD190270	39.71	Asthma
UNPD250	38.73	Asthma
UNPD56182	34.05	Rheumatoid arthritis, unspecified
UNPD95242	33.85	Asthma
UNPD17936	33.81	Rheumatoid arthritis, unspecified
UNPD209061	33.54	Asthma
UNPD53084	33.45	Asthma
UNPD52311	33.44	Rheumatoid arthritis, unspecified
UNPD25089	33.39	Asthma
UNPD95242	33.02	Type II diabetes mellitus
UNPD209061	25.03	Chronic lymphocytic leukemia
UNPD52311	23.02	Multiple sclerosis
UNPD56182	22.98	Multiple sclerosis

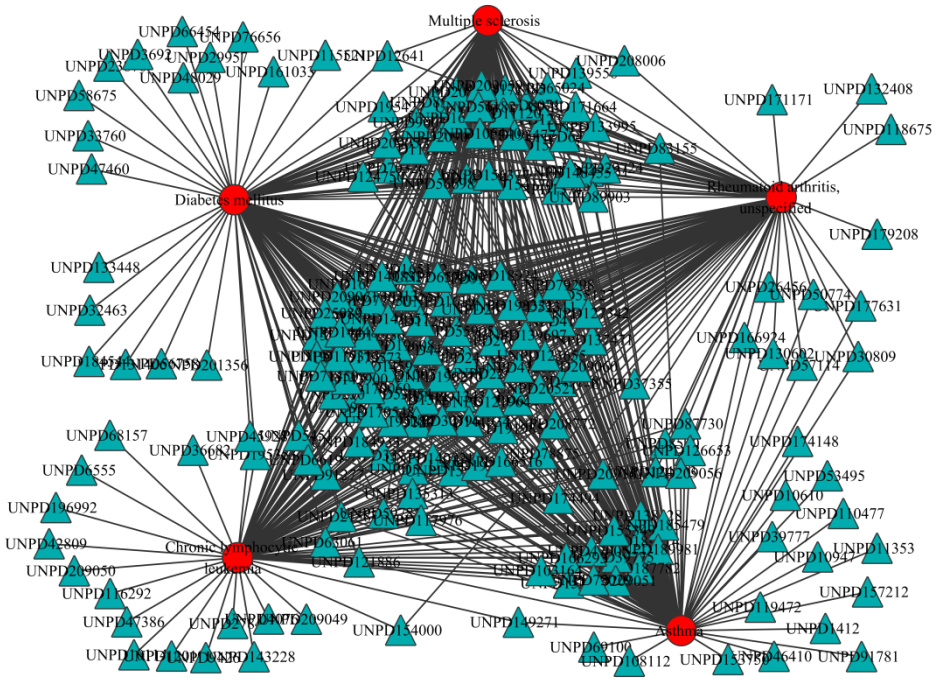


Fig. 6 Potential compound-potential disease network: Multiple sclerosis, Diabetes mellitus, Asthma, Rheumatoid arthritis and Chronic lymphocytic leukemia. Representations of the symbols were the same to Fig. 5.

Conclusions

In summary, we have constructed the network of interactions between carbohydrates and target proteins. We have constructed the network of carbohydrates which obeyed the “rule of five” and related disease targets. We found some potential carbohydrate candidate molecules and some new interactions between the molecules and proteins. Furthermore, the networks revealed that carbohydrates can be used to be potential lead candidates for multi-target therapies. Moreover, the network indicated that carbohydrates had potential therapeutic effects against complex diseases, such as asthma, rheumatoid arthritis and Type II diabetes mellitus. Our results provided new insights to understand the pharmacological properties of carbohydrate compounds and were beneficial for drug discovery for treating complex diseases.

Acknowledgements

This work was financially supported by National Key Special Project of Science and Technology for Innovation Drugs (Grant No. 2012ZX09501001-004 and 2013ZX09402202) and the Enterprise Academician Workstation of Jiangsu Province, PR China (Grant No. BM2011027). The calculations were performed on TianHe-1(A) at National Supercomputer Center in Tianjin, PR China.

Notes and references

[a] Beijing National Laboratory for Molecular Sciences, State Key Lab of Rare Earth Material Chemistry and Applications, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P. R. China. Fax: (+86)010-62751708

E-mail: lirongc@pku.edu.cn; xiaojxu@pku.edu.cn

† The first two authors should be regarded as joint first authors.

‡ Electronic Supplementary Information (ESI) available. See DOI: 10.1039/b000000x/

1. D. Rognan, *Br. J. Pharmacol.*, 2007, **152**, 38-52.
2. T. Klabunde, *Br. J. Pharmacol.*, 2007, **152**, 5-7.
3. Y. Van Kooyk and G. A. Rabinovich, *Nat. Immunol.*, 2008, **9**, 593-601.
4. R. D. Cummings, *Mol Biosyst*, 2009, **5**, 1087-1104.
5. A. Varki and J. B. Lowe, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor (NY), 2nd edn., 2009.
6. B. Ernst and J. L. Magnani, *Nat. Rev. Drug Discov.*, 2009, **8**, 661-677.
7. A. K. Anatole, in *Glycobiology and Drug Design*, American Chemical Society, 2012, vol. 1102, ch. 1, pp. 3-22.
8. M. C. Galan, D. Benito-Alifonso and G. M. Watt, *Org. Biomol. Chem.*, 2011, **9**, 3598-3610.
9. J. L. Magnani and B. Ernst, *Discov. Med.*, 2009, **8**, 247-252.
10. J. W. Scannell, A. Blanckley, H. Boldon and B. Warrington, *Nat. Rev. Drug Discov.*, 2012, **11**, 191-200.
11. H. Kitano, *Nat. Rev. Drug Discov.*, 2007, **6**, 202-210.
12. A. L. Hopkins, *Nat. Chem. Biol.*, 2008, **4**, 682-690.
13. M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin and B. K. Shoichet, *Nat. Biotechnol.*, 2007, **25**, 197-206.
14. A. L. Hopkins, *Nature*, 2009, **462**, 167-168.
15. T. Klabunde and A. Evers, *ChemBioChem*, 2005, **6**, 876-889.
16. A. Vedani, M. Dobler and M. A. Lill, *Basic Clin. Pharmacol. Toxicol.*, 2006, **99**, 195-208.
17. S. J. Haggarty, K. M. Koeller, J. C. Wong, R. A. Butcher and S. L. Schreiber, *Chem. Biol.*, 2003, **10**, 383-396.
18. S. S. Ortega, L. C. Cara and M. K. Salvador, *Drug Metabol. Drug Interact.*, 2012, **27**, 199-207.
19. A. D. Boran and R. Iyengar, *Curr Opin Drug Discov Devel*, 2010, **13**, 297-309.
20. S. Zhao and R. Iyengar, *Annu. Rev. Pharmacol. Toxicol.*, 2012, **52**, 505-521.
21. A. Pujol, R. Mosca, J. Farres and P. Aloy, *Trends Pharmacol. Sci.*, 2010, **31**, 115-123.
22. E. L. Leung, Z. W. Cao, Z. H. Jiang, H. Zhou and L. Liu, *Brief Bioinform*, 2013, **14**, 491-505.
23. B. Yang, J. Zhang, Y. Yin and Y. Zhang, *Biomed Res Int*, 2013, **2013**, 401649.
24. H. Seo, W. Kim, J. Lee and B. Youn, *Int. J. Oncol.*, 2013, **43**, 1737-1744.
25. X. Robin, P. Creixell, O. Radetskaya, C. Santini, J. Longden and R. Linding, *Clin. Pharmacol. Ther.*, 2013, DOI: 10.1038/clpt.2013.171.
26. J. Harrold, M. Ramanathan and D. Mager, *Clin. Pharmacol. Ther.*, 2013, DOI: 10.1038/clpt.2013.176.
27. L. Xie, L. Xie, S. L. Kinnings and P. E. Bourne, *Annu. Rev. Pharmacol. Toxicol.*, 2012, **52**, 361-379.
28. Q. Li, X. Li, C. Li, L. Chen, J. Song, Y. Tang and X. Xu, *PLoS One*, 2011, **6**, e14774.
29. S. Tian, Y. Li, D. Li, X. Xu, J. Wang, Q. Zhang and T. Hou, *J. Chem. Inf. Model.*, 2013, **53**, 1787-1803.
30. H. Liu, J. Wang, W. Zhou, Y. Wang and L. Yang, *J. Ethnopharmacol.*, 2013, **146**, 773-793.
31. X. Qiao, T. Hou, W. Zhang, S. Guo and X. Xu, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 481-489.
32. J. Gu, Y. Gui, L. Chen, G. Yuan, H. Z. Lu and X. Xu, *PLoS One*, 2013, **8**, e62839.
33. A. A. Klesov, Z. J. Witczak and D. Platt, *Carbohydrate drug design*, American Chemical Society ;

Distributed by Oxford University Press, Washington, DC London ; New York, 2006.
34. C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo and D. S. Wishart, *Nucleic Acids Res.*, 2011, **39**, D1035-1041.
35. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv Drug Deliv Rev*, 2001, **46**, 3-26.
36. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785-2791.
37. X. Jiang, K. Kumar, X. Hu, A. Wallqvist and J. Reifman, *Chem. Cent. J.*, 2008, **2**, 18.
38. M. Kohl, S. Wiese and B. Warscheid, *Methods Mol. Biol.*, 2011, **696**, 291-303.
39. A. Lagunin, D. Filimonov and V. Poroikov, *Curr. Pharm. Des.*, 2010, **16**, 1703-1717.
40. C. M. Dobson, *Nature*, 2004, **432**, 824-828.
41. C. Lipinski and A. Hopkins, *Nature*, 2004, **432**, 855-861.
42. A. C. Rigby, *Comb. Chem. High Throughput Screen.*, 2009, **12**, 927-928.
43. A. L. Hopkins, *Nat. Biotechnol.*, 2007, **25**, 1110-1111.
44. J. Hoeng, R. Deehan, D. Pratt, F. Martin, A. Sewer, T. M. Thomson, D. A. Drubin, C. A. Waters, D. de Graaf and M. C. Peitsch, *Drug Discov. Today*, 2012, **17**, 413-418.
45. H. B. Engin, A. Gursoy, R. Nussinov and O. Keskin, *Curr. Pharm. Des.*, 2013.
46. R. E. Carney, J. B. McAlpine, M. Jackson, R. S. Stanaszek, W. H. Washburn, M. Cirovic and S. L. Mueller, *J. Antibiot. (Tokyo)*, 1978, **31**, 441-450.