Using QSAR model for studying heterocycles activity

K K Hammud*, R M Aboob, H A Khalaf, N A M Akosh and F M Hamza

1Ministry of Science and Technology, Directorate of Materials Research, Baghdad-Iraq

*Corresponding: kafaa_khalaf@yahoo.com

Abstract. Different heterocyclic saccharin derivatives that have been previously prepared were tested here for their pharmacokinetic properties with Marvinsketch program and PreADMET website. The quantitative structure–activity relationship (QSAR) calculations included isoelectronic point (pI) as Protonation factor, logP as a partitioning factor that calculated by Consensus and Chemaxon methods, hydrophilic-lipophilic balance (HLB) as a partitioning factor that calculated by Chemaxon, Davies, and Griffins methods, polar surface area (PSA) besides adsorption, distribution, metabolism, excretion, and toxicity (ADMET) to detect the properties. These calculated QSAR characters confirmed that saccharin derivatives were more hydrophilic than saccharin molecule which managed membrane permeability, mutagenicity, and carcinogenicity descriptions. As a conclusion, the tested properties showed a good chance of some of these prepared saccharin derivatives to be good drugs.

Keywords: QSAR, Model, Heterocyclics, Marvinsketch program, saccharin

1. Introduction

Many heterocyclic compounds were multi-step synthesized and tested in vivo and in vitro for drug discovery applications. This time-consuming processes may give unacceptable due to the interference with the metabolism, toxicity, or pharmacokinetic properties. Preclinical discovery of any chemical towards drug performance then improvement charges high range. The quantitative structure – activity relationship QSAR studies give important notes about any chemical structure for discovering the binding target through a receptor or ion channel, for example, the chemical substance dissolves in gastrointestinal tract then passes liver – blood stream pathway or crossing the blood – brain barrier (BBB) to enter central nerve system (CNS) or the brain [1].

To control the efficacy of any chemical as a drug active material, chemical characteristics of this active component must be in the range membrane permeability and protein transportation. These characteristics cover oral bioavailability techniques through in vivo and in vitro with computer – aided prediction (QSAR and adsorption, distribution, metabolism, excretion, and toxicity ADMET) [2].

Now days, many researchers or drug companies worked with earlier steps through combinatorial chemistry to evaluate a heterocyclic compound efficiency on ADMET [3, 4, 5, 6, 7]. ADMET study provide researchers a good understanding of molecular structure capabilities with Lipinski’s rule of five or/ and Jorgensen’s rule of three to be a candidate drug [8]. There are many previous studies on studied drug likeness, lipophilicity, and other important predictors of heterocycles contain N, S, O, Si, and other heteroatoms. These predictors showed the advantages of using these compounds and studying as candidate drug or focused bio-materials. These powerful computational calculations were used to design and explore binding ability of the chemical structure with biological activity [8, 9, 10, 11]. With this...
point of view, this work was to apply QSAR as a model for studying the efficiency of previously prepared saccharin derivatives as a candidate drug through several Marvinsketch program and PreADMET descriptors.

2. Experimental section

Saccharin heterocyclic derivatives were previously prepared and characterized [12, 13] (Figures 1.2, 3). These derivatives were randomly selected for this study and tested by QSAR models.

All saccharin derivatives were calculated by Marvinsketch program Version 18.15.0 [14] to find the following descriptors: Isoelectronic point (pI), as Protonation factor, logP as a partitioning factor that calculated by Consensus and Chemaxon methods, hydrophilic-lipophilic balance (HLB) also as a partitioning factor that calculated by Chemaxon, Davies, and Griffins methods. Polar surface area (PSA) with and without P & S exclusion was also calculated. The difference between presence and absence of each sulfur calculation was 8.38. The comparison depended on the presence of S, Table 1. The online website PreADMET [15] was used to calculate absorption, distribution, metabolism, excretion, and toxicity parameters of these prepared heterocyclics (Table 2).

Figure 1. Structures of Saccharin (SA) and its derivatives (SA-1- SA-8).
Figure 2. Structures of Saccharin derivatives (SA-9-SA15).

Figure 3. Structures of Saccharin derivatives (SA-16-SA-20).
3. Results and Discussion

Any chemical substance (or drug) absorbs by the human body through many steps depending on its interaction with any bio-transporter, its metabolism by a specific enzyme, permeability, and solubility. These factors determine the availability of this substance to be used orally. This process may occur by two pathways (Paracellular and Trancellular). The first pathway (Paracellular) is the slowest because it needs small aqueous pores (3-10 Å) in the small intestine [16]. The second (Transcellular) pathway is faster because drug can pass the cytoplasm or intra-membrane along the gastro-intestinal tract to reach the target [17].

There is a significant relationship between the physiological pH on chemical intake and pI, logP, and HLB data where the solubility of ionizable chemical depends upon this particular pH in gastrointestinal tract [18]. It is known that (pI) is the "pH value at which the molecule carries no electric charge (in a solution)" [19,20]. This means that acidity-basicity character of any solution has a good influence upon electric charge of the tested chemical. Table 1. shows that pI ranged from 3.00 [SA-10, C₆H₅NO₃S] to 7.63 [SA-17, C₁₇H₁₀N₂O₅S] with no pI character of saccharin and these data indicated that acidic solution had the largest impact or high acidity to a slight basicity may gave zero charge of these chemicals.

Lipinski’s rule number five is a guideline for oral delivery in QSAR modeling that limits options of any chemical that considers as a candidate drug [M.Wt less than 500 Daltons, logP less than 5, hydrogen bond acceptor not more 10 while donor not more 5]. In general, the presence of lipophilicity character in any chemical suggests that this chemical has poor ability in both oral absorption and bioavailability because of low solubility in aqueous medium. The standard term lipophilicity is the partition coefficient (log P).

Now Table 1. Shows the descriptors of saccharin heterocyclics. For logP data, both methods used for calculation (Consensus and Chemaxon) showed no significant differences. Consensus method showed logP range (-0.95 [SA-16, C₆H₅N₂O₅S]) to 4.99 [SA-15, C₃₀H₂₄N₆O₆S₂] while Chemaxon method showed (-0.73 [SA-16, C₆H₅N₂O₅S]) to 5.64 [SA-16, C₃₀H₂₄N₆O₆S₂]. Most Saccharin derivatives...
had logP values more than saccharin. However, Negative logP values referred to poor membrane permeability of these high hydrophilic saccharin derivatives.

Table 1. Also shows that logP with hydrogen bond donor and acceptor as predicators of saccharin and its (28) derivatives agreed with Lipinski’s rule five while molecular weight (M.Wt) showed that some of these calculated chemicals were out of this rule. HLB is another calculated character used to predicate the surface active performance depending upon lip- and hydro- groups presence in the tested chemicals. This predicator was calculated by applying three different methods (Chemaxon, Davies, and Griffin) (Table 1.). HLB data were (4.72 – 16.36) for Chemaxon, (1-17.27) for Davies, and (8.83-15.00) for Griffin method. Maximum HLB values for these three methods were semi-identical because of mathematical basis of these test methods.

In addition, HLB values of the parent molecule (saccharin) were 8.95 (Chemaxon), 6.35 (Davies), and 13.00 (Griffin). It seems that HLB variation may be affected by the presence of heterocyclic atoms and hydrogen bonding. Hydrogen bonding, an interaction between electroNegative atom and hydrogen atom in molecular structure, may enhance hydrophilicity than lipophilicity. According to HLB industrial categories, SA-8, SA-3, SA-15, and SA-2 can be considered as water/oil emulsifying agents (HLB range = 3-6), SA-9, SA-1, SA-14, SA-13, SA-5, SA-19, SA-10, SA-24, SA-4, and SA as wetting and spreading agents (HLB range = 7-9), SA-19, SA-10, SA-24, SA-4, SA-12, SA-18, SA-22-SA-20, SA-23, SA-6, SA-21, SA-28, SA-27, and SA-26 as oil / water emulsifying agents (HLB range = 8-16), and finally SA-26, SA-25, and SA-16 can be applied as solubilizing agent (HLB range 15-18), therefore, these saccharin and saccharin derivatives can be good candidates for industry and petrochemicals.

Table 1. pl, logP, HLB, and PSA descriptors of Saccharin heterocycles.
Different ADMET predication methods have selected to quantify chemicals (candidate as a drug) for both oral absorption and toxicity assay as in vitro model. Tables (2-4) show ADMET values for saccharin (SA) and its previously prepared derivatives (SA-1 to SA-28). The results of these 29 calculated chemicals can be summarized in several points:

i. Toxicity is the key improvement of any chemical to drug market. Toxicity data were Algae at, Ames test, Carcino – Mouse, Carcino- Rat, Daphnia- at, hERG-inhibition, medaka- at, minnow – at, TA100 - 10RLI, TA100 – NA, TA1535 - 10RLI, and TA1535 – NA. Numerical values of these predictors were ranged (0.000622- 0.573431) for Algae at, (0.001297 – 2.19115) for Daphnia- at, (1.01E-05 – 6.90708) for medaka- at, and (0.00032- 5.13237) for minnow – at. Mutagenicity and carcinogenicity of these chemicals under test were done and showed that:

a. Ames test: SA-1, SA-2, SA-4 to SA-14, SA-16 to SA-18 were with mutagen predication while the rest were with non- mutagen predication.

b. Carcino – Mouse: SA-7, SA-9, SA-11, SA-17, SA-18, SA-21, SA-22 gave Positive predication while the others gave Negative.

c. Carcino- Rat: all chemicals (SA, SA1 to SA-28) gave Negative predication.

d. hERG-inhibition: there were ambiguous inhibition with SA-10 and SA-25 to SA-28, medium risk inhibition with SA-7, SA-9, and SA-13, and low risk inhibition was predicated for the rest of chemicals under calculation.

e. TA100 - 10RLI, TA100 – NA, TA1535 - 10RLI, and TA1535 – NA: there was mutagenic predication of Ames Salmonella TA100 -10RLI (SA, SA-10, SA-11, SA-12, SA-17, SA-18), TA100-NA (SA-2, SA-4 to SA-9, SA-11, SA-12, SA-16,SA-17), TA1535-10RLI (SA, SA-4, SA-5, SA-7 to SA-9, SA-11, SA-12), and TA1535-NA (SA-4 to SA-7, SA-9, SA-11, SA-12, SA-16).

ii. ADME Prediction for these 29 compounds was done for example through Adsorption [Caco2, Madin – Darby Canine Kidney (MDCK), Human Intestinal Absorption (HIA), Skin Permeability], Distribution [Blood Brain Barrier (BBB), Plasma Protein Binding], Caco-2 or Madin – Darby Canine Kidney (MDCK) is an in vitro Human Intestinal Absorption model to find out chemical permeability so oral absorption or metabolism effected by enzymes. These predictors and others are summarized as below:

a. BBB: SA-10 was with lowest value (0.009632) while SA-8 was with highest value (1.89556).

Chemical may have a direct action in the central Nerve System and to do this job it must has the ability to cross BBB to the target. To avoid chemical side effect, molecular mass (less than 450 Dalton) and Polar Surface Area (PSA, less than 100 Å) are the controllers so BBB may play a good penetration player with M.Wt & PSA values above their limitations. Saccharin derivatives may be penetrating the BBB with molecular mass less than 450 Daltons. In this sequence, saccharin itself and other 16 derivatives (SA-1, SA-2, SA-4, SA-5 to SA-13, SA-16 to SA-18, SA-27) had this character. Also, PSA values (with S calculation) were less than 100 Å in SA, SA-1, SA-2, SA-4 to SA-9, SA-11 to SA-13. Merging both M.Wt and PSA characters gave SA, SA-1, SA-2, SA-4 to SA-9, SA-11 to SA-13 a prediction to breakthrough BBB defense walls.

b. Buffer solubility, mg/L ranged from 0.624376 (SA-20) to 1.867090 (SA-12).
c. Caco-2 cell model was in its lowest value with SA-16 (0.362091) while the highest with SA-8 (21.55).

d. CYP-2C19 inhibition gave no inhibitor character to all 29 compounds.

e. CYP-2C9 inhibition was with SA, SA-1, SA-2, SA-3, SA-8, SA-10, SA-11, SA-14, SA-15, SA-19, SA-20 while the others were non.

f. Only SA-10 and SA-11 were CYP-2D6 inhibitors.

g. For CYP-2D6 Substrate test, SA-10 and SA-11 were CYP-2D6 substrates and SA-25 to SA-28 were weakly while the others were non.

h. For CYP-3A4 inhibition, only SA-3, SA-11, SA-15, and SA-19 to SA-21 were with inhibition character.

i. For CYP-3A4 substrate, non-property was presented in SA, SA-12, and SA-17 and the substrate character was with SA-1, SA-3, SA-5, SA-6, SA-11, SA-14, SA-20, SA-21, SA-25, SA-26, SA-28. The other 18 saccharin derivatives were weakly.

j. HIA predication was found to be ranged from 82.71294 (SA-16) to 99.21234 (SA-1).

k. MDCK, as Adsorption and protein transportation predications that express enzyme action and BBB penetration, was found to be high with SA-7 (55.9885) while the lowest was with SA-14 (0.043435). This mimic of BBB permeability can be considered poor because in most of them MDCK values were less than 25.

l. Pgp inhibition was found to be effective in most of the prepared saccharin derivatives except SA, SA-1, SA-2, SA-4 to SA-13, SA-15, SA-16, SA-17, SA-22.

m. Plasma Protein Binding was remarkable in its results through the lowest value 100 for different saccharin derivatives (SA-1, SA-3, SA-13, SA-15). Drug transport to the target may be achieved through Plasma – Protein Binding in blood [Red Blood Cells (RBCs), Platelets, or Leukocytes]. Also, proteins such albumin, lipoprotein, or glycoprotein may be bind with drug with acidic, basic, or neutral character respectively.

n. Pure Water Solubility in mg/L ranged from 0.0128473 (SA) to 17089.6 (SA-11).

o. Skin Permeability predication ranged from -5.07241 (SA-16) to -2.00797 (SA-3).

p. SKlogD values were (-2.34107 to 3.69156).

q. SKlogP range was -1.24885 (SA-16) to 3.69156 (SA-3).

r. SKlogS buffer ranged from -5.94612 (SA-20) to 0.8475 (SA-12).

s. SKlog S pure ranged from -7.72674 (SA-3) to -1.0622 (SA-11).

Jorgensen's rule of three are another limitation for oral delivery [Qplog S more than -5.7, Caco-2 more than 22 nm/s, and primary metabolites not more than 7]. Lipinski rules were found to be violated in two points with some tested saccharin derivatives but not in hydrogen bond donor or acceptors (as a count). In general, more than 80% had acceptable drug-pharmacokinetic descriptors because the violations of Lipinski rules were:

- M.Wt in Lipinski rule must be less than 500 Daltons and our tested compounds showed saccharin and 23 saccharin derivatives had no violation for this limitations. The violations in molecular mass were marked in SA-3, SA-15, SA-19,SA-20, and SA-21.
- Violation of logP value (not less than 5) was only found in Sa-15 (5.64). This violation is repeated in SA-15 in both molecular mass (M.Wt) and partitioning (logP).

| ID                | SA-1   | SA-2   | SA-3   | SA-4   | SA-5   | SA-6   | SA-7   | SA-8   | SA-9   |
|-------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Alfalfa at        | 0.2902 | 0.11681| 0.097859| 0.00062709| 0.263835| 0.5701 | 0.111399| 0.014199| 0.168171|
| Anisotest C5       |        |        |        |        |        |        |        |        |        |
| Carcinio-Mouse CNC |        |        |        |        |        |        |        |        |        |
| Carcinio-Rat       |        |        |        |        |        |        |        |        |        |
| Daphnia at         |        |        |        |        |        |        |        |        |        |
| SERG inhibition     |        |        |        |        |        |        |        |        |        |
| medaka at          |        |        |        |        |        |        |        |        |        |
| mnmn-9W2           | 1.20023| 0.054736| 0.0105977 | 1.01045e-005| 1.569 | 1.07061 | 3.53564 | 0.094114 | 0.111789 | 0.217813 |
| TA100-10RLI         | 0.471333| 0.165972 | 0.060188 | 0.0009595027 | 0.58517 | 0.61591 | 2.35325 | 0.064379 | 0.023785 | 0.12149 |

Table 2. ADMET data of Saccharin and its (SA-1 to SA-9) derivatives.
### Table 3. ADMET data of Saccharin and its (SA-10 to SA-18) derivatives.

| ID          | SA-10 | SA-11 | SA-12 | SA-13 | SA-14 | SA-15 | SA-16 | SA-17 | SA-18 |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| TA1555 - 10RL | Positive | Negative | Positive | Negative | Positive | Positive | Positive | Negative | Positive |
| TA1555 - 10RL Buffer solubility, mg/L | 0.294017 | 0.388481 | 0.438717 | 0.497927 | 0.720657 | 0.957803 | 1.486750 | 1.89556 | 1.07952 |
| HIA | 94.624318 | 99.21234 | 97.27056 | 96.55931 | 97.43032 | 96.91611 | 89.33103 | 98.55277 | 97.81401 | 98.49994 |
| MDCK | 9.06294 | 3.20876 | 7.53662 | 0.043543 | 13.9024 | 1.9935 | 0.651914 | 55.9885 | 34.8469 | 47.8477 |
| Pgp inhibition | Non | Non | Non | Non | Non | Non | Non | Non | Non |
| Caco2 | 0.368135 | 21.0601 | 18.5932 | 21.1576 | 9.91784 | 12.6362 | 0.373315 | 21.2596 | 21.55 | 21.0477 |
| Buffer solubility, mg/L | 29249.7 | 1128.72 | 271.284 | 3.27713 | 3915.03 | 2523.23 | 1568.39 | 175.163 | 115.731 | 1907.82 |
| CYP - 1F29 | 0.550125 | 0.037432 | 0.087382 | 0.124625 | 0.195370 | 0.270982 | 0.348752 | 0.429723 | 0.510617 | 0.591511 |
| CYP - 2D6 | 0.111730 | 1.52671 | 1.787340 | 3.69156 | 0.417130 | 0.573230 | -0.70192 | 0.2073 | 2.61379 | 1.57865 |
| CYP - 2D6 substrate | Non | Non | Non | Non | Non | Non | Non | Non | Non |
| CYP - 3A4 inhibition | Non | Substrate | Weakly | Substrate | Weakly | Substrate | Non | Weakly | Weakly |
| Plasma Protein | 45.566259 | 100 | 97.467966 | 100 | 34.19429 | 63.87238 | 31.41373 | 78.66534 | 97.5091 | 70.54864 |
| Pure water solubility, mg/L | 6152.07 | 337.128 | 87.0491 | 0.012847 | 7444.22 | 3908.06 | 8004.11 | 322.769 | 41.132 | 863.173 |
| Skin permeability | -5.49903 | -2.63155 | -2.79942 | -2.00797 | -3.44506 | -3.3313 | -4.51533 | -2.40854 | -2.07806 | -2.56695 |
| SKlogD value | 0.510120 | 1.52671 | 1.787340 | 3.69156 | 0.417130 | 0.573230 | -0.70192 | 0.2073 | 2.61379 | 1.57865 |
| SKlogP value | -0.796760 | -2.464600 | -3.085560 | -5.320660 | -1.70219 | -2.00163 | -2.18513 | -3.15675 | -3.41722 | -2.99471 |
| SKlogp Pgp | -1.473860 | -2.99119 | -3.579220 | -7.72674 | -1.44097 | -1.85852 | -1.47323 | -2.8913 | -3.86199 | -2.43915 |

**Notes:**
- **SA** represents the Solubility of compounds.
- **HIA** refers to Human Intestinal Absorption.
- **MDCK** indicates Madin-Darby Canine Kidney cell line.
- **Pgp** signifies P-glycoprotein.
- **Caco2** denotes Caco-2 cell line.
- **BBB** refers to Blood-Brain Barrier.
- **SKlogD** and **SKlogP** are calculated values related to intestinal permeability and protein binding, respectively.
- **SKlogp Pgp** represents Pgp-mediated permeability.

### Additional Information:
- **IOP Publishing** is the publisher of the Journal of Physics: Conference Series.
- The conference details are from 1853(2021) 012002.
- The authors acknowledge the financial support by *(insert funding agencies here)*.
4. Conclusion

Interaction of several saccharin heterocyclic compounds in human body is mainly depending on the ionizable character pI, logP, and HLB data as well as, Lipinski’s rule five and Jorgensen’s rule of three. Results of this study indicated that saccharin derivatives were more hydrophilic than pure saccharin toward low membrane permeability due to hydrogen bonding interactions. This low lipophilicity was the controller of toxicity data that give the drug acceptable pharmacokinetic descriptors, mutagenicity, and carcinogenicity limits. Our calculations showed that saccharin and its derivatives may pass the threshold of BBB defense walls.

5. References

[1] Foata F, Sprenger N, Rochat F, and Damak S 2020 Sci. Rep. 10, 16117.
[2] Waterbeemd H, Lennermäls H, and Artursson P, Eds. 2003 Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability (Wiley –VCH Verlag GmbH & Co. KGaA, Weinheim).
[3] Vijulathaand M. and Kanth S 2007 Central Eur. J. Chem. 5(4), 1064.
[4] Roy K (Ed.) 2019 *In Silico* Drug Design: Repurposing Techniques and Methodologies (Elsevier Inc., New York).

[5] Kang F, Lifongo L, Mbah J, Owono L, Megnassan E, Mbane L, Juson P, Sipple W, and Efange S 2013 *In Silico Pharmacol.* 1, 12.

[6] Bhosle V, Altit G, Autmizguine J, and Chemtob S 2017 Fetal and Neonatal Physiology, Basic Pharmacological Principles, Chapter 18, Volume 1 (Elsevier Inc., New York).

[7] Debnath B and Ganguly S 2014 *Asian J. Pharmac. Clin. Res.* 7(5), 186.

[8] Sangshettia J, Khan F, Chouthe R, Damale M, and Shinde D 2014 *Chinese Chem. Lett.* 25, 1033.

[9] Lilitha P and Sivakamasundari 2010 *Orien. J. Chem.* 26(1), 135.

[10] Lokwani D, Bhandari S, Pujari R, Shastri P, Shelke G, and Pawar V 2011 *J. Enzy. Inhib. Med. Chem.* 26(2), 319.

[11] Barraza S and Denmark S 2018 *J. Am. Chem. Soc.* 140, 66668.

[12] Al-Bayati R, Habib M, and Mekky A 2014 *Int. J. Sci. Technol.* 3(9), 521.

[13] Naser A 2013 *Iraqi National J. Chem.* 50, 199.

[14] www.Chemaxon.com

[15] https://preadmet.bmrc.kr

[16] Tang V and Goodenough D 2003 *Biophys. J.* 84(3), 1660.

[17] Laksitoriini M, Prasasty, Kiptoo P, Siahaan T 2014 *Ther. Deliv.* 5(10), 1143.

[18] Waterbeem H and Testa B (Eds.) 2009 Drug Bioavailability: Estimation of Solubility, Permeability, Absorption, and Bioavailability. Vol. 40, 2nd Edition (Wiley–VCH Verlag GmbH & Co. KGaA, Weinheim).

[19] Moldoveanu S and David V 2017 Properties of analytes and matrices determining HPLC selection. Selection of the HPLC Methods in Chemical Analysis, Chapter 5 (Elsevier Inc., London).

[20] Qiu F and Scrivens G (Eds.) 2018 Accelerated Predictive Stability (APS): Fundamental Pharmaceutical Industry Practices (Elsevier Inc., New York).