QSAR Study Regarding the Inhibitory Activity of Some Iminosugars Against α-glucosidase

LASZLO TARKO*, ANCA HIRTOPEANU*
Romanian Academy, Center of Organic Chemistry "C.D.Nenitzescu", 202B Spl. Independentei, 060023 Bucharest

Iminosugars are known as potent inhibitors of glycosidases, being thus interesting for treating carbohydrate mediated diseases. As α–glucosidase from yeast is similar to human α-glucosidase, a QSAR study to identify the molecular features relevant for the inhibitory activity against α-glucosidase from yeast is important for the rational design of therapeutic leads. In our study the calibration set includes 26 iminosugars (including 5-6 atoms in cycle) with known inhibitory activity. The prediction set includes 12 iminosugars. There are two "activity cliffs" and an outlier in the calibration set. Excluding the outlier, the predictive quality of the best QSAR is high enough ($r^2 = 0.8816$; $F = 54.6$; $r^2_{CV} = 0.7119$). The molecular features inferred as having largest influence on activity are the number of atoms, the number of OH groups and the presence/absence of double bonds in the heterocycle (absence is favorable). Three iminosugars in the prediction set were identified as compounds recommended for synthesis.

Keywords: inhibitory activity, iminosugars, α-glucosidase, PRECLAV

Iminosugars mimic carbohydrates, acting as inhibitors of various glycosidases, which are the enzymes responsible for catalyzing the glycosidic bonds in complex carbohydrates and glycoconjugates [1]. Therefore iminosugars are considered to be powerful agents against various carbohydrate mediated diseases such as cancer, viral infections (hepatitis, HIV) and diabetes [2]. The inhibitory activity against a certain type of glycosidase depends on the structure of the iminosugar (number of cycle atoms and stereochemistry) and on the substituents on the cycle. While most structure – activity relationship studies were centered on deoxynojirimycin derivatives, very few articles report the synthesis and different glycosidase inhibitory activity of a series of pyrrolidine iminosugars with varying stereochemistry and substituents on nitrogen [3-5]. Recently two QSAR studies were published in literature regarding the glycosidase inhibitory activity of iminosugars [6, 7].

It is known the fact that α-glucosidase from yeast and human á-glucosidase I share a high identity

in their catalytically active domains, have similar substrate speciûcity, pH optimum, and inhibitor sensitivity [8, 9].

We present here a QSAR study regarding the inhibitory activity against yeast α-glucosidase of a series of iminosugars. The aim is to identify the molecular features (significant molecular fragments included) having the highest impact (favorable or unfavorable) on the inhibitory activity.

The calibration set includes 26 iminosugars having inhibitory activity against yeast α-glucosidase. The number of molecules is rather low, but in order to get reliable results it is important for the biological activity to be observed in identical conditions, even if it was determined by different research groups. The prediction set includes 12 molecules having unknown inhibitory activity, the purpose being to identify the ones most likely to present biological activity.

Methods and formulas

The analyzed molecules (calibration set + prediction set) are presented in table 1. The molecules **27**-**38** were included in the prediction set. We used literature data for the observed values of the dissociation constant K_i measured in similar conditions (*p*H 6.8), ranging between [0.14, 7,500]. The "dependent property" is the inhibitory activity A_K, defined by formula (1). The value of *k* factor in formula (1) is the minimum value of K_i (Ki = 0.14 for compound **2**). The observed values A_{Kobs} of activity, used in computations, are within [0, 4.729] range.

$$A_{Kobs} = \log (K_i/k) \quad (1)$$

The minimum energy geometry, for each molecule, was obtained by PCModel v. 9.0 software [27], using MMX force field [28]. Then the geometry was more rigorously optimized using the quantum mechanics program MOPAC v. 8.137W [29], with the PM6 method [30]. In the MOPAC analysis the keywords "pm6 pulay gnorm=0.2 geo-ok mmok bonds vectors" were used to set the parameters for geometry optimization.

In the next step, the programs MOPAC, DRAGON v. 5.4 [31] and PRECLAV v. 1305 [32-35] computed, for each molecule, the value of almost 1700 molecular descriptors.

A specific procedure [36] identified the "significant" molecular fragments. Actually, if the bond order value of the chemical bond between two heavy atoms (different from hydrogen) is high enough, these atoms are included in the same molecular fragment. Consequently, the conjugation of a certain molecular fragment with the neighboring fragment(s) is low. The percentages, in weight, of "significant" fragments, are well correlated with the values of activity. The significance of a positive Pearson correlation sign is that "a large mass percentage of this fragment increases the activity value". Consequently, a negative correlation sign signifies that "a large mass

* *email: ltarko@cco.ro; ahirtopeanu@yahoo.com*

percentage of this fragment decreases the activity value". Often, the identified fragments are not the same as the classical chemical groups.

A specific procedure identified the "significant" descriptors. Then, the PRECLAV program computed QSARs (2):

$$A_K = C_0 + \sum_{i=1}^{p} C_i \times D_i \quad (2)$$

where

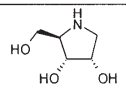
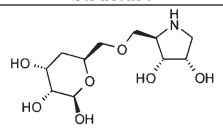
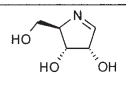
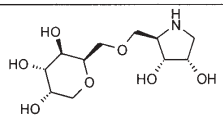
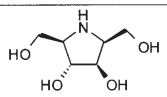
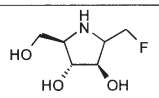
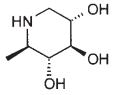
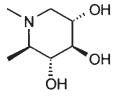
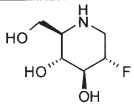
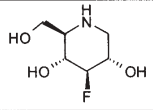
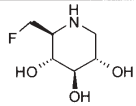
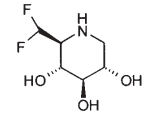
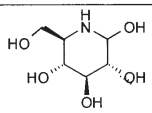
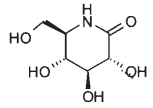
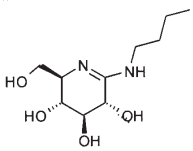
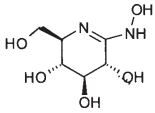
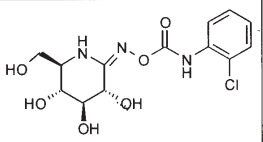
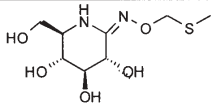
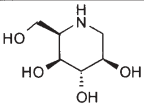
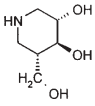
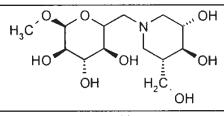
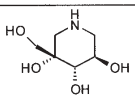
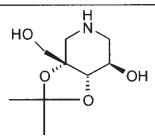
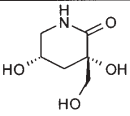
A_K = inhibitory activity;
C_0 = intercept;
C_i = weighting factors;
D_i = the values of "significant" descriptors;
p = number of descriptors.

The descriptors included in the best QSAR are named "predictors". Some specific formulas [32, 35] computed the quality of the equations, identified the "high outliers" in the calibration set and computed the relative "utility" of predictors within [0, 1000] range. The predictors which present high value of "utility" can be considered very useful for estimating the activity, because they correlate well with the activity and do not correlate with other predictors. Thus, each "useful" predictor offers a different kind of information from the other predictors.

After computing the A_{calc} values of the activity for the prediction set molecules, PRECLAV sorts these molecules according to the computed values. The program computed the average value A_{calc}^{m} and the standard deviation s of the estimated values. The program considers "high estimated values" the values fulfilling the criterion (3) and "low estimated values" the values fulfilling the criterion (4).

$$A_{calc} > A_{calc}^{m} + 0.5 \times \sigma \quad (3)$$
$$A_{calc} < A_{calc}^{m} - 0.5 \times \sigma \quad (4)$$

Table 1
THE ANALYZED MOLECULES AND THE VALUE OF THE INHIBITORY ACTIVITY AGAINST α-GLUCOSIDASE

No.	Structure	A_{Kobs}	Ref	No.	Structure	A_{Kobs}	Ref
1		2.216	[10]	3		3.836	[11]
2		0.000	[10]	4		3.632	[11]
5		2.757	[12]	6		2.610	[13]
7		4.047	[14]	8		4.104	[14]
9		4.155	[13]	10		4.252	[13]
11		2.133	[13]	12		4.729	[15]
13		1.653	[16]	14		3.895	[17]
15		1.808	[18]	16		1.316	[17]
17		1.385	[17]	18		2.553	[17]
19		2.269	[19]	20		2.788	[20] [21]
21		2.625	[20] [21]	22		3.216	[22] [23]
23		4.155	[23]	24		4.216	[24]

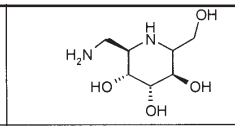
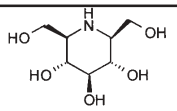
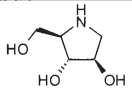
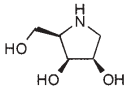
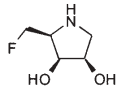
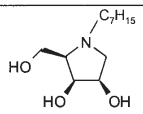
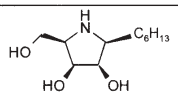
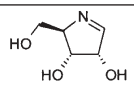
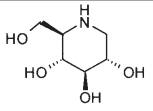
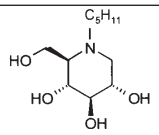
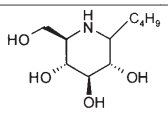
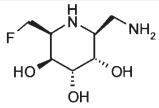
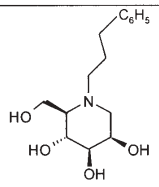
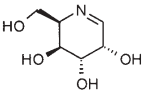
25		2.345	[25]	**26**		3.808	[26]
27		?		**28**		?	
29		?		**30**		?	
31		?		**32**		?	
33		?		**34**		?	
35		?		**36**		?	
37		?		**38**		?	

Table 1
CONTINUED

Therefore, a certain molecule in the prediction set presents high, moderate or low activity not in an absolute manner, but compared to the other molecules in the prediction set. As a rule, the molecules having high estimated values are recommended for synthesis. In addition, the computed average A_{calc}^{m} activity of the molecules in the prediction set can be compared with the computed activity of the molecules from the calibration set.

Results and discussions

As a rule, similar structures have similar properties. If the property is the biochemical activity this statement is named QSAR axiom. Frequently, this axiom is challenged because some similar structures present non-similar activities and some non-similar structures present similar activities. The presence of such "activity cliffs" highlights the infringement of the QSAR axiom for some pairs of molecules.

There are two "activity cliffs" in the analyzed calibration set:

- the molecules **11** and **12** having high similarity of structures vs. low similarity of activities
- the molecules **13** and **17** having low similarity of structures vs. high similarity of activities

The "significant" molecular fragments are -CH=N- (included in molecule **2, 32** and **38**; r = - 0.5249) and -N=C-NH- (included in molecules **15-18**; r = - 0.4486)**.**

From the point of view of the best QSAR, using all molecules in the calibration set, the molecule **11** is an outlier (A_{Kcalc} = 4.059). Maybe the reported low A_{Kobs} value for this molecule was not correctly determined. The quality of this QSAR, which includes four predictors, is low enough (r^2 = 0.7591 F = 17.3 r^2_{CV} = 0.7647, where F is the Fisher function). Unusually, the value of the cross-validated square of Pearson linear correlation factor r^2_{CV} is higher than the value of the square of the Pearson linear correlation factor r^2.

Table 2
THE COMPUTED VALUE OF THE INHIBITORY ACTIVITY WITHOUT THE OUTLIER **11**

No. (Table 1)	A_{Kcalc}	No. (Table 1)	A_{Kcalc}	No. (Table 1)	A_{Kcalc}
1	1.600	**14**	3.264	**27**	3.212
2	-0.021	**15**	1.899	**28**	3.475
3	3.937	**16**	1.828	**29**	3.196
4	3.292	**17**	1.281	**30**	4.119
5	2.878	**18**	2.551	**31**	2.444
6	3.060	**19**	3.009	**32**	-0.684
7	4.369	**20**	2.851	**33**	3.609
8	4.056	**21**	2.775	**34**	3.885
9	3.419	**22**	3.607	**35**	3.615
10	3.379	**23**	3.820	**36**	4.301
11	-	**24**	4.112	**37**	3.396
12	4.839	**25**	2.349	**38**	-1.232
13	2.073	**26**	4.143		

The best QSAR (formula (2)), in the absence of the outlier has:

C_0 = 0.4383

C_1 = 108.8781

D_1 is the molecular volume weighted dipole moment

U = 1000

C_2 = - 3.5662

D_2 is a 3D-MoRSE (DRAGON) descriptor [37] weighted by Sanderson electronegativities

U = 839

C_3 = 1.6608

D_3 is an unweighted 3D-MoRSE (DRAGON) [37] descriptor

U = 595

The quality of this QSAR, which includes three predictors, is high enough (r^2 = 0.8816;

F = 54.6; r^2_{CV} = 0.7119).

Minimum correlation of predictors with activity: for D_3 (r^2 = 0.2251)

Maximum intercorrelation of predictors: for pair D_2/ D_3 (r^2 = 0.0461)

The utility of predictors for description of inhibitory activity is very high for D_1 and D_2 and moderate for D_3.

The computed values A_{Kcalc}, in the absence of the outlier **11,** are presented in table 2. The average of A_{Kcalc} in calibration set is 2.975 ± 1.109. The average of A_{Kcalc} in the prediction set is 2.778 ± 1.813.

The prediction set molecules having "high" computed activity, according to formula (3), are **30, 34** and **36**. The prediction set molecules having "low" computed activity, according to formula (4), are **32** and **38.**

Conclusions

A QSAR study was performed to determine the molecular features relevant to the inhibitory activity against yeast α-glycosidase for a series of iminosugars. In the study the calibration set included 26 iminosugars with 5-6 atoms in the cycle having the α-glucosidase inhibitory activity determined in identical conditions and the predicition set included 12 iminosugars. For reliable conclusions the following parameters were used: a) the physical meaning of predictors, according to MOPAC/PRECLAV/DRAGON documentation; b) the mathematical sign of predictors in the best QSAR; c) the value of predictors' utility; d) the structure and correlation of "significant" molecular fragments.

The molecular features (inferred as) having largest influence on activity are

- the number of atoms in the heterocycle (optimum value seems to be 6)

- the number of OH/CH_2OH groups on the cycle (optimum value seems to be 3)

- the presence/absence of double bonds in the cycle (absence is favorable)

The influence of presence/absence of large R (alkyl), X (halogen) and/or N (tri-substituted nitrogen) groups is debatable as the number and the diversity of molecules in the prediction set is rather low. The influence of the molecular lipophilicity seems to be low, because the logP descriptor is not a predictor.

Acknowledgments: Financial support from the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, through project number PN-II-RU-TE-2011-3-0298 is gratefully acknowledged.

References

1. ASANO, N., NASH, R.J., MOLYNEUX, R.J., FLEET, G.W., Tetrahedron Asymm., **11**, 2000, p. 1645.
2. BOUILLON, M.E., PYNE, S.G., Tetrahedron Lett., **55**, 2014, p. 475.
3. SAOTOME, C., WONG, C.H., KANIE, O., Chem. Biol., **8**, 2001, p. 1061.
4. AYERS, B.J., NGO, N., JENKINSON, S.F., FERNANDO MARTINEZ, R., SHIMADA Y., ADACHI, I., WEYMOUTH-WILSON, A.C., KATO, A., FLEET, G.W.J, J. Org. Chem., **77**, 2012, p. 7777.
5. CHANG, Y.F., GUO, C.W., CHAN, T.H., PAN, Y.W., TSOU, E.L., CHENG, W.C., Mol. Divers., **15**, 2011, p. 203.
6. KHALAF, R.A., ABDULA, A.M., MUBARAK, M.S., TAHA, M.O., J Mol. Model., **17**, 2011, p. 443.
7. MOORTHY, N.S.H.N., NATERCIA, F.B., RAMOS, M.J., FERNANDES, P.A., Bioorg. Med. Chem., **20**, 2012, p. 6945.
8. BAUSE, E., ERKENS, R., SCHWEDEN, J., JAENICKE, L., FEBS Lett., **206**, 1986, p. 208.
9. SCHWEDEN, J., BORGMANN, C., LEGLER, G., BAUSE, E., Ach. Biochem. Biophys., **248**, 1986, p.335.
10. WITTE, J.F., MCCLARD, R.W., Tetrahedron Lett., **32**, 1991, p. 3927.
11. MIKKELSEN, G., CHRISTENSEN, T.V., BOLS, M. LUNDT, I., Tetrahedron Lett., **36**, 1995, p. 6541.
12. LEGLER, G., KORTH, A., BERGER, A., EKHART, C.W., GRADNIG, G., STUTZ, A.E., WITHERS, S.G., WRODNIGG, T., Carbohydr. Res., **250**, 1993, p. 67.
13. ANDERSEN, S.M., EBNER, M., EKHART, C.W., GRADNIG, G., LEGLER, G., LUNDT, I., STUTZ, A.E., Carbohydr. Res., **301**, 1997, p. 155.
14. KAJIMOTO, T., LIU, K.K.C., PEDERSON, R.L., ZHONG, Z., ICHIKAWA, Y., PORCO, J.A., WONG, C.H., J. Am. Chem. Soc., **113**, 1991, p. 6187.
15. SZAREK, M.A., WU, X., SZAREK, W.A., Carbohydr. Lett., **299**, 1997, p. 165.
16. TSURUOKA, T., FUKUYASU, H., ISHI, M., USUI, T., SHIBAHARA, S., INOUYE, S., J. Antibiot. **49**, 1996, p. 155.
17. HOOS, R., VASELLA, A., RUPITZ, K., WITHERS, S.G., Carbohydr. Res., **298**, 1997, p. 291.
18. LEGLER, G., FINKEN, M.-T., Carbohydr. Res., **292**, 1996, p. 103.
19. FOWLER, P.A., HAINES, A.H., TAYLOR, R.J.K., CHRYSTAL, E.J.T., GRAVESTOCK, M.B., Carbohydr. Res., **246**, 1993, p. 377.
20. JESPERSEN, T.M., DONG, W., SIERKS, M.R., SKRYDSTRUP, T., LUNDT, I., BOLS, M., Angew. Chem., **106**, 1994, p. 1858.
21. DONG, W., JESPERSONE, T., BOLS, M., SKRYDSTRUP, T., SIERKS, M.R., Biochemistry, **35**, 1996, p. 2788.
22. ICHIKAWA, M., IGARASHI, Y., ICHIKAWA, Y., Tetrahedron Lett., **36**, 1995, p. 1767.
23. PERSSON, M.P., BUTT, W.M., JORGENSEN, M., CHRISTENSEN, P., HANSEN, L.T., BOLS, M., Tetrahedron Lett., **37**, 1996, p. 2097.
24. HANSEN, A., TAGMOSE, T.M., BOLS, M., Tetrahedron, **53**, 1997, p. 697.
25. WONG, C.-H., PROVENCHER, L., PORCO, J.A., JUNG, S.-H., WANG, Y.-F., CHEN, L., WANG, L., STEENSMA, D.H., J. Org. Chem., **60**, 1995, p. 1492.
26. HOLT, K.E., LEEPER, F.J., HANDA, S., J. Chem. Soc., Perkin Trans. I, **3**, 1994, p. 231.
27. *** PCModel (last version) - available from Serena Software, http://www.serenasoft.com
28. MM2 (improved version), see Allinger N. L., J. Am. Chem. Soc., **99**, 1977, p. 8127
29. *** MOPAC - available from http://www.openmopac.net/
30. STEWART, J.J.P., J. Mol. Model., **13**, 2007, p. 1173.
31. DRAGON - available from Talete srl., http://www.talete.mi.it
32. PRECLAV - available from the Center of Organic Chemistry – Bucharest, E-mail: ltarko@cco.ro or tarko_laszlo@yahoo.com
33. TARKO, L., Rev. Chim. (Bucuresti), **56**, 2005, p. 639
34. TARKO, L., LUPESCU, I., GROPOSILA – CONSTANTINESCU, D., ARKIVOC, **X**, 2005, p. 254
35. TARKO, L., STECOZA, C.E., ILIE, C., CHIFIRIUC, M.C., Rev. Chim.(Bucharest), **60**, no. 5, 2009, p. 476
36. TARKO, L., Rev. Chim. (Bucuresti), **55**, no. 6, 2004, p. 539
37. SCHUUR, J.H., SELZER, P., GASTEIGER, J., J. Chem. Inf. Comput. Sci., **36**, 1996, p. 334

Manuscript received: 2.12.2014