

A Multivariate Approach to Saccharide Quantitative Structure–Activity Relationships Exemplified by Two Series of 9-Hydroxyellipticine Glycosides

Jörgen Jonsson, Lennart Eriksson,* Sven Hellberg, Michael Sjöström and Svante Wold

Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden

Jonsson, J., Eriksson, L., Hellberg, S., Sjöström, M. and Wold, S., 1989. A Multivariate Approach to Saccharide Quantitative Structure–Activity Relationships Exemplified by Two Series of 9-Hydroxyellipticine Glycosides. – *Acta Chem. Scand.* 43: 286–289

Multivariate saccharide quantitative structure–activity relationships (QSARs) have been developed for two series of 9-hydroxyellipticine glycosides. In order to describe the structural variation of the glycoside moieties, thirteen chromatographic variables were used. Eleven D-glycosides and seven L-glycosides were used in separate QSARs to model 9-hydroxyellipticine glycoside antitumour activity against L 1210 leukemia. The multivariate partial least squares (PLS) method was used to establish the QSARs.

Quantitative structure–activity relationships (QSARs) are useful in drug development¹ and for general hazard assessment in environmental toxicology.^{2,3} QSARs have also received much attention in biology and biochemistry.² In order to establish an informationally useful QSAR for compounds of any complexity, it is recommended that a multitude of chemical-property descriptors is utilized simultaneously.^{4–6} This will allow a more complete description of all major factors related to the biological response.

We have previously applied the multivariate approach to QSAR to β -adrenergic agents, chlorinated aliphatic compounds, halogenated ethers and peptides.⁴ In this paper we report the use of QSARs for a type of compounds that (to our knowledge) has hitherto not been subjected to *multivariate* QSAR modelling, namely, saccharides.

Saccharide QSARs. The concept of structure–activity modelling in saccharide chemistry is not new. In an extensive article, Birch discusses the correlation between saccharide stereochemistry and a threshold sweetness/bitterness activity.⁷ A more traditional univariate structure–activity modelling is used by Rathmore *et al.*⁸ and Yamada *et al.*⁹, who have related saccharide chemistry/stereochemistry to cardiac glycoside activity and anti-complementary activity of polysaccharides, respectively. Another example of the latter is given by Honda *et al.*,¹⁰ who synthesized a series of quaternary ellipticine glycosides acting as antitumour agents. So far, however, we have not found any attempts in the literature aimed at deriving multivariate saccharide QSARs. This may depend on the limited availability of interesting homologous series of saccharides, and on the insufficient characterization of saccharide properties.

In order to develop the multivariate saccharide QSAR,

we used an approach which has previously been applied to the analogous problem of peptide QSARs. In a peptide QSAR, it is essential that the amino acids are appropriately parametrized. In order to describe the amino acids we have derived three scales (z_1 , z_2 and z_3) that, when combined, together reflect the ‘principal properties’ of the amino acids. These scales were derived from a multiproperty matrix consisting of 29 physico-chemical descriptors for each amino acid. The structural variation within a series of peptides is then described by arranging the principal property vectors according to the amino-acid sequence. Thus, it is possible to translate peptides into numbers, and subsequently relate their physico-chemical properties to biological responses.^{4,11,12}

Hence, in order to develop a saccharide QSAR a multivariate characterization of the smallest repetitive carbohydrate units, i.e. the monosaccharides was required. Recently, we have characterized 16 monosaccharides by measuring their retention times in 13 different thin-layer chromatography (TLC) systems.¹³ A principal-components analysis (PCA) showed that four independent factors were necessary to describe the saccharides. The dominant factor describing 90% of the variance in the data, was interpreted as being related to the hydrophobicity of the monosaccharides. In the present study, we use these TLC data to develop QSARs for a series of ellipticine glycosides with antitumour activity reported by Honda *et al.*¹⁰

Characterization of ellipticine glycosides. Ellipticine is an alkaloid which exhibits antitumour properties in tests with experimental animal tumours. Most ellipticine derivatives are insoluble in water, which has resulted in considerable difficulties in the development of compounds for clinical testing. However, when combined with a saccharide moiety, the water solubility of such ellipticine derivatives is

* To whom correspondence should be addressed.

increased, which might influence their bioavailability, permeability and selective toxicity against tumour cells.¹⁰

Honda *et al.* have synthesized 45 ellipticine glycosides and tested their antitumour activity against intraperitoneally implanted L1210 leukemia in mice. Of these 45 glycosides, 36 were 9-hydroxyellipticines, five were 9-methoxyellipticines and four were 'pure' ellipticines. In order to obtain a homogeneous class for QSAR modelling, the inclusion of compounds was restricted to the 36 9-hydroxyellipticine glycosides. However, among these compounds another 18 derivatives had to be omitted to obtain a consistent training set for the saccharide QSAR. Thus, six glycosides were left out because they were anomeric (α/β) mixtures, six because they were furanosyl derivatives and a further six since characteristics of their glycoside moieties were not available. However, three of the α/β mixed compounds were used as test set objects, see below. The omitted furanosyl derivatives, three D- and three L-glycosides, were insufficient for QSAR development. To summarize: 18 9-hydroxyellipticine glycosides remained; the glycoside part of these compounds consisted of aldopentoses and aldohexoses in their pyranosyl form. All other complementary conformational details are given in Table 1.

Within the selected series of 9-hydroxyellipticine glycosides, the backbone part (i.e. the ellipticine moiety) remained constant for all compounds (Fig. 1). Hence, the structural variability is related to changes in the glycoside moiety. For the structural description of the glycoside part of the selected compounds, we used the previously measured monosaccharide TLC data.¹³ In total, 13 different

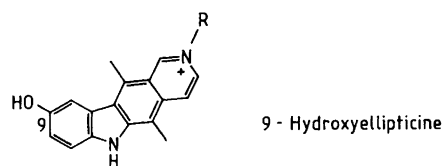


Fig. 1. The 9-hydroxyellipticine backbone. The different glycosides were obtained by combining the backbone with diverse monosaccharides at the 'R'-position (Table 1).

chromatographic variables were used. Moreover, in order to discriminate between α - and β -forms, one extra qualitative (zero/one) variable was included.

Statistical analysis

An approximately symmetrical distribution of the observed responses and the chromatographic retention data was achieved by converting the data into log values.

The saccharide QSARs were established using the multivariate PLS (partial least-squares projections to latent structures) methodology.^{14,15} PLS relates chemical structural descriptors (here: TLC descriptors plus the qualitative variable mentioned) to biological-response data (here: antitumour activity) by multivariate projections of the two data-blocks so that the projection scores of the two blocks are correlated. In this way, a quantitative relationship between chemical descriptors and biological activity is obtained. In the PLS computation, the number of significant components (i.e. projection dimensions) is determined by cross-validation.¹⁶ The PLS method has been extensively described in the literature.^{14,15}

Results

As a first step, we divided the 18 9-hydroxyellipticine derivatives into two classes. This division gave α -D/ β -D glycosides in one class and α -L/ β -L-glycosides in the other. Two separate QSARs were then developed, one for each class (see below, next section). However, to check for the appropriateness of the subdivision, a QSAR model was proposed that simultaneously included all 18 9-hydroxyellipticine glycosides. This QSAR gave no significant explanation of the biological activity. Hence we conclude that it is difficult to model the present compounds in the same QSAR.

The PLS analysis of the α -D/ β -D class (with 11 compounds) resulted in three significant (with cross-validation) model dimensions. In total, these PLS components described 84% of the variance in the antitumour activity variable (y) and 91% of the variance in the structural (chromatographic and qualitative) descriptors (x). The corresponding cross-validated result was 71% for the antitumour activity (y). The first PLS component described 9% of the antitumour activity variance, the second 66%, and the third 9% (11, 48 and 12%, cross-validated), respectively.

Table 1. Observed and calculated antitumour activity for the 9-hydroxyellipticine glycosides.

Glycoside moiety (R) ^a	Antitumour activity ^b	
	Calculated	Observed
1 α -D-Arabinosyl	2.89	2.78
2 α -D-Lyxosyl	2.64	2.81
3 β -D-Ribosyl	2.13	2.05
4 β -D-Fucosyl	2.20	2.09
5 β -D-Allosyl	2.33	2.40
6 β -D-Galactosyl	2.62	2.86
7 β -D-Glucosyl	2.09	1.72
8 α -D-Mannosyl	2.10	1.94
9 α -D-Talosyl	2.22	2.33
10 2-Acetamido-2-deoxy- β -D-galactosyl	0.96	0.90
11 2-Acetamido-2-deoxy- β -D-glucosyl	1.31	1.59
12 α -L-Arabinosyl	2.93	2.93
13 α -L-Lyxosyl	3.02	2.90
14 β -L-Ribosyl	2.99	2.97
15 β -L-Fucosyl	2.67	2.77
16 β -L-Galactosyl	1.79	1.76
17 β -L-Glucosyl	1.83	1.76
18 α -L-Mannosyl	2.43	2.56

^aR marks the substitution position on the 9-hydroxyellipticine backbone (see Fig. 1). All glycosides are in the pyranosyl form.

^bThe antitumour activity is expressed as increased mean life span, log (ILS %).¹⁰

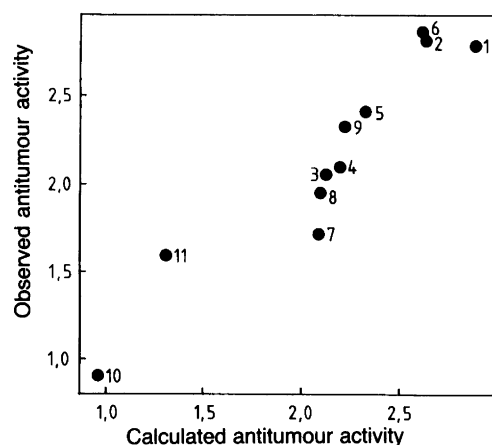


Fig. 2. Correlation plot for the D-glycosides. The observed antitumour activity is plotted versus the calculated activity. Notation as in Table 1.

In Fig. 2, it can be seen that there is fairly good agreement between the observed antitumour activity of the D-glycosides and the corresponding calculated activity (see also Table 1). There is no clear separation of α -D (Nos. 1, 2, 8 and 9) and β -D (Nos. 3–7, 10 and 11) glycosides, which justifies them being modelled in the same QSAR. For the second class with α -L and β -L glycosides (with 7 compounds), the PLS analysis also resulted in three significant PLS components. Together these components explained 94% of the antitumour-activity (y) variance (91% cross-validated) and 95% of the variance in the structural data (x). The PLS components described 70, 13 and 11% of the biological activity variance (73, 10 and 8% cross-validated), respectively. The observed and calculated antitumour activities are plotted in Fig. 3. For this class of L-glycosides, the *pentapyranosyl* glycosides (Nos. 12–14) are seen to have higher antitumour potency than the *hexapyranosyl* glycosides (Nos. 15–18).

For comparative purposes, the two QSAR calculations were repeated, this time using *only* the qualitative (α/β)

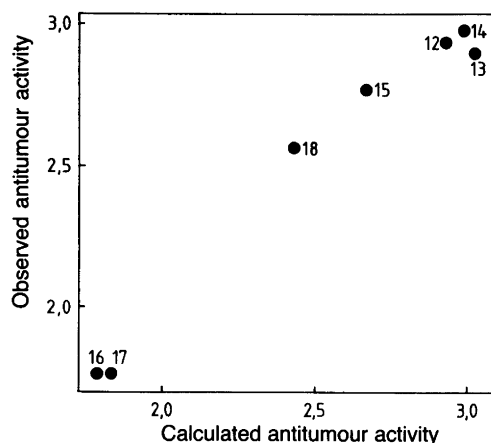


Fig. 3. Correlation plot for the L-glycosides. The observed antitumour activity is plotted versus the calculated activity. Notation as in Table 1.

variable as structural descriptor. For the D-glycosides, a barely significant one-dimensional PLS model resulted, that explained 10% of the antitumour-activity variance. The PLS model for the L-glycosides was not significant at all. The insignificant first PLS component accounted for only 7% of the variance in the biological-activity variable. Since both QSARs had poor modelling capability, it was concluded that the TLC descriptors contribute most of the information in the present saccharide-QSAR development.

The derived QSARs were validated using three α/β mixed pyranosyl glycosides as a test set. One of these (L-rhamnopyranosyl glycoside) did not fit the L-QSAR model, hence was not meaningful to predict its biological response. The other two test compounds (D- and L-xylopyranosyl glycoside) fitted fairly well to their models, which enabled meaningful predictions. Since the two compounds were anomeric α/β mixtures (29/71), the predictions were calculated both for the α and the β form, resulting in a predicted biological-activity interval.

For the D-xylopyranosyl glycoside the activity was predicted to be in the interval 2.05 (β)–2.52 (α) with an actual observed activity of 2.93. For the L-xylopyranosyl glycoside the predicted interval was 2.29 (β)–2.69 (α) with an observed value of 2.60. Thus the predicted activity interval of the latter compound is in accordance with the observed activity. The prediction error for the D-glycoside is approximately 30% of the activity range of the compounds studied.

Discussion

In aqueous solution monosaccharides exist in an equilibrium between open-chain structures and different ring conformations.¹⁷ Thus, the previously derived TLC variables probably describe some average characteristics for the monosaccharides, a fact that, at first glance, might raise the question of whether these variables may be used in saccharide QSARs. It is evident that these variables cannot discriminate between, for instance, an imagined enantiomeric pair of some β -D and β -L glycosides. However, owing to the initial selection of ellipticine derivatives, the subsequent division into α -D/ β -D and α -L/ β -L classes, and the inclusion of a qualitative variable, the TLC descriptors are indeed appropriate for the present QSAR application. These descriptors contain information concerning the hydrophobicity/hydrophilicity of the monosaccharides, which might explain why they are capable of modelling the antitumour activity of the 9-hydroxyellipticine glycosides.

The suitability of subdividing the 9-hydroxyellipticine derivatives into separate classes of D- and L-glycosides is best seen by comparing the first PLS components of the two established QSAR models. Both components explain substantial parts of the variance in respective blocks of structural descriptors (x), but completely different proportions of the variance in the antitumour activity variable (y). For the D-glycosides, only 9% of the activity variance is accounted for by the first PLS component, while among the L-glycosides the corresponding proportion is 70%. These

results strongly indicate that the biological activity models for the two classes are indeed different.

Both the saccharide QSARs show fairly good predictive power. The predicted antitumour-activity interval for the L-xylopyranosyl glycoside matches the observed activity. The predicted interval for the D-xylopyranosyl glycoside is, however, not equally good. Nevertheless, when compared with the total antitumour-activity range of the compounds studied, the prediction is quite reasonable.

We wish to emphasize that the present study, the first attempt at a multivariate saccharide QSAR, should be regarded as a preliminary illustration of the construction of quantitative relationships between saccharide properties and a biological activity. Here we have used measured data of monosaccharides to model relatively simple glycoside moieties of ellipticine derivatives, an application far removed from the complex situations that might prevail among oligosaccharides. In our opinion, however, the results are encouraging and indicate the possibility of developing multivariate saccharide QSARs.

In drug development, QSAR models are often used at late stages, if used at all, and the synthesized and tested analogues are rarely selected with the aim of giving well balanced training sets for QSAR modelling. However, we strongly believe that QSAR models could play a more central and active role in drug design. This could be achieved using statistical designs such as factorial or fractional factorial designs¹⁸ for selecting the compounds to be synthesized for screening. For such training sets, QSAR models with improved predictive capability can be expected compared with those from more or less randomly compiled training sets. Recently, we have shown the efficiency of fractional factorial designs for the selection of training-set compounds for different series of peptides, alcohols and halogenated hydrocarbons.¹⁹ Here, principal properties were used as design variables. Analogously, we suggest that the principal properties derived for monosaccharides¹³ be used as design variables for obtaining efficient training sets for saccharide QSARs.

Acknowledgements. We are grateful for the financial support from the Swedish Natural Science Research Council (NFR).

References

1. Seydel, J. K. and Schaper, K.-J. *Chemische Struktur und Biologische Aktivität von Wirkstoffen*, Verlag Chemie, Weinheim 1979.
2. Turner, L., Choplin, F., Dugard, P., Hermens, J., Jaekch, R., Marsmann, M. and Roberts, D. *Toxicol. in Vitro* 1 (1987) 143.
3. Kaiser, K. L. E. *QSAR in Environmental Toxicology*, Proc. Workshop on QSAR in Environ. Toxicol., Hamilton, Ontario, Canada 1983, Reidel, Dordrecht, Holland 1984.
4. Hellberg, S. *QSAR a Multivariate Approach*, Thesis, Umeå University, Umeå, Sweden 1986.
5. Wold, S. and Dunn, W. J., III. *J. Chem. Inf. Comp. Sci.* 23 (1983) 6.
6. Wold, S., Eriksson, L., Jonsson, J., Hellberg, S. and Sjöström, M. *Proc. Sem. Use of QSAR for Chemicals Screening*, Report No. 8 (1988), Swedish National Chemicals Inspectorate, Stockholm, 1988.
7. Birch, G. G. *Crit. Rev. Food. Sci. Nutr.* 9 (1976) 57.
8. Rathore, H., From, A. H. L., Ahmed, K. and Fullerton, D. S. *J. Med. Chem.* 29 (1986) 1945.
9. Yamada, H., Nagai, T., Cyong, J. C. and Otsuka, Y. *Carbohydr. Res.* 144 (1985) 101.
10. Honda, T., Kato, M., Inoue, M., Shimamoto, T., Shima, K., Nakanishi, T., Yoshida, T. and Noguchi, T. *J. Med. Chem.* 31 (1988) 1295.
11. Wold, S., Eriksson, L., Hellberg, S., Jonsson, J., Sjöström, M., Skagerberg, B. and Wikström, C. *Can. J. Chem.* 65 (1987) 1814.
12. Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S. *J. Med. Chem.* 30 (1987) 1126.
13. Eriksson, L., Jonsson, J., Sjöström, M. and Wold, S. *Acta Chem. Scand., Ser. B* 42 (1988). 504.
14. Dunn, W. J., III, Wold, S., Edlund, U., Hellberg, S. and Gasteiger, J. *Quant. Struct.-Act. Relat.* 3 (1984) 131.
15. Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. *Chemometrics: Mathematics and Statistics in Chemistry*, NATO ISI Series C 138, Reidel, Dordrecht, Holland 1984, pp. 4-96.
16. Wold, S. *Technometrics* 20 (1978) 379.
17. Stryer, L. *Biochemistry*, Freeman, San Francisco, USA 1981, pp. 257-258.
18. Box, G. E. P., Hunter, W. G. and Hunter, J. S. *Statistics for Experimenters*, Wiley, New York, USA 1978.
19. Tosato, M. L., Marchini, S., Passerini, L., Pino, A., Eriksson, L., Hellberg, S., Jonsson, J., Sjöström, M., Skagerberg, B. and Wold, S. (1988). *Submitted for publication.*

Received October 24, 1988.