Available online at www.sciencedirect.com
ScienceDirect
Tetrahedron Letters
Tetrahedron Letters 47 (2006) 9017–9020

A new tool for the rational design of methylbiotin hosts

Fernando Herranz, M. Dolores Santa María and Rosa M. Claramunt*

Departamento de Química Orgánica y Bio-Orgánica, UNED, Facultad de Ciencias, Senda del Rey 9, E-28040 Madrid, Spain

Received 29 August 2006; revised 19 October 2006; accepted 24 October 2006

Abstract—To study the binding mode of biotin related compounds with artificial hosts, we have developed a new tool to be used as a guide to test their behaviour prior to their synthesis. In that way, we have considered a set of 23 complexes comprising biotin and urea derivatives with synthetic hosts to develop a Partial Least Squares Cross-Validated (PLS-CV) model. The data, for such a model, are the binding constants (K_b) of each complex and the interaction energies ($-E_{min}$) calculated by molecular mechanics with AMBER and OPLS force fields. The predictive power of the model has been verified.

1. Introduction

In the host–guest chemistry field, one of the main purposes is to mimic the recognition of biologically active compounds. Biotin is perhaps the most appealing natural molecule being known that this ligand forms very strong complexes with avidin (K_b 2.5×10^{13} M^{-1}) and streptavidin (K_b 1.7×10^{15} M^{-1}).[1] The challenge of mimicking its behaviour is a difficult task that has not been attained at the moment. One of the main reasons is the traditional approach, where new hosts are synthesized without previous tests, in most cases just by intuition.

In previous papers, we have proved that by a dynamic interplay between experiments and molecular modelling it is possible to approximate the objective to get an artificial complex that will mimic the behaviour of the protein avidin towards biotin.[2–5] The use of molecular modelling to test new designed hosts saves time and allows to reach a deeper understanding of the main variables directing this kind of interactions.

2. Results and discussion

We have demonstrated how the theoretical enthalpies ($-E_{min}$) provided by molecular mechanics calculations using Monte Carlo conformational search of the most stable structures of the host–guest complexes are well correlated with the experimental binding constants ($\mathrm{Ln}\,K_b$) determined by means of NMR titrations. Moreover, our results confirm that for such close related compounds and in chloroform as solvent, the entropic factor must be similar or rather close for all cases.[2–5]

However, if the aim is to obtain a model with predictive power, a different approach must be undertaken. In a simple linear regression model, the parameter r^2 indicates how well correlated the data are, but not how well predicted would be for a new compound not used in the model generation. Because of that, we decided to use the Partial Least Squares (PLS) Cross-Validated (CV) methodology.[6,7] PLS is a method for constructing predictive models when the factors are many and/or highly collinear. The emphasis is made on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables.[7] In our case the number of factors, the energies ($-E_{min}$) calculated with AMBER and OPLS force fields, are small but they are highly correlated (Fig. 1). In the CV method the predictive power of the model is denoted by the value of the q^2 parameter, the closer to 1 the higher predictive power. This type of methodology, quite common in QSAR studies, is used for the first time here in the field of the host–guest chemistry as far as we know.

The data used in the model generation are compiled in Table 1. These data are the binding constant values K_b and the interaction energies $-E_{min}$ calculated with AMBER and OPLS force fields for the complexes formed between the hosts depicted in Figure 2 and

* Corresponding author. Fax: +34 91 398 83 72; e-mail: rclaramunt@ccia.uned.es

0040-4039/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetlet.2006.10.120

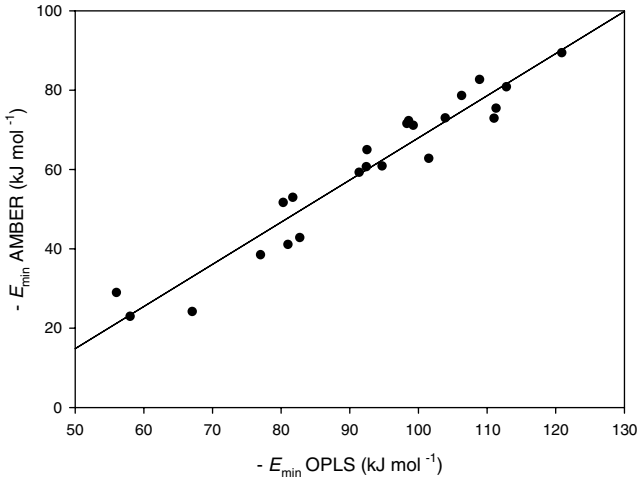
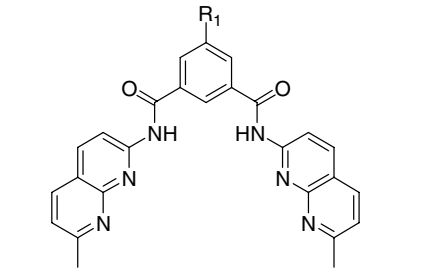


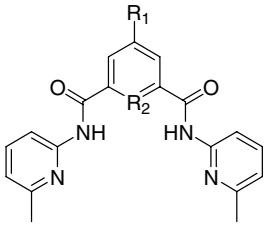
Figure 1. Correlation between AMBER and OPLS energies.

Table 1. Data for model development

Complex	K_b	$-E_{min}$	
		AMBER	OPLS
I:**1**	3.50×10^4	78.64	106.29
I:**2**	3.00×10^1	24.20	67.05
I:**3**	9.50×10^3	71.13	99.24
I:**4**	6.00×10^3	72.33	98.56
I:**5**	1.20×10^3	75.45	111.30
II:**1**	9.75×10^2	60.70	92.40
II:**2**	5.00	29.00	56.00
II:**3**	1.45×10^3	51.70	80.30
II:**4**	2.30×10^3	53.00	81.70
II:**5**	2.37×10^3	65.00	92.50
III:**1**	3.60×10^3	71.60	98.30
III:**2**	5.00	23.00	58.00
III:**3**	1.41×10^2	38.50	77.00
III:**4**	1.00×10^2	41.10	81.00
III:**5**	2.74×10^2	62.80	101.50
IV:**1**	4.00×10^3	72.90	111.00
IV:**2**	3.40×10^1	42.80	82.70
IV:**3**	4.80×10^3	60.90	94.70
IV:**4**	5.70×10^3	59.30	91.40
IV:**5**	6.10×10^3	73.00	103.90
V:**1**	1.48×10^5	89.40	120.90
V:**3**	3.30×10^4	80.80	112.80
V:**4**	2.10×10^4	82.70	108.90



R_1 = -H ; **I**
R_1 = -CONH-(7-methylnaphthyridin-2-yl); **V**



R_1= -H; R_2 = -CH ; **II**
R_1= -Cl; R_2=N; **III**
R_1 = -CONH-(6-methylpyridin-2-yl); R_2=CH; **IV**

Figure 2. Hosts used to generate the model.

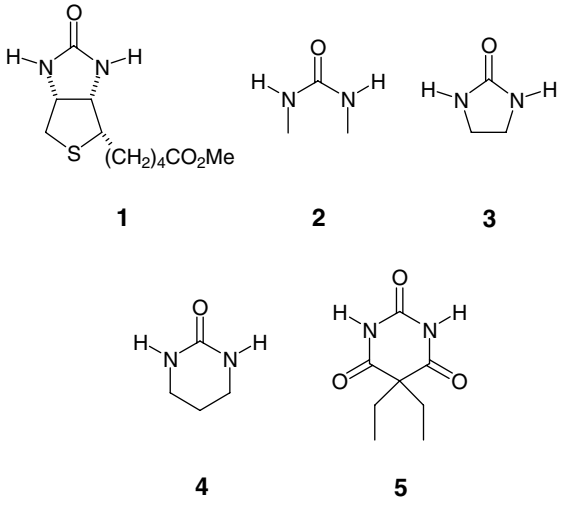


Figure 3. Guests used to generate the model.

the guests,[3,4,8,9] methylbiotin (**1**), *N*,*N*′-dimethylurea (**2**), 2-imidazolidone (**3**), *N*,*N*′-trimethyleneurea (**4**) and barbital (**5**), shown in Figure 3.

In order to obtain the most probable structures for the complexes, as well as the associated energies, we carried out Monte Carlo conformational searches using as starting points the minimized structures of hosts and guests. We performed all the calculations with both AMBER and OPLS force fields (Macromodel software)[10] with the GB/SA model for chloroform,[2–5] except for host **III** in which the solvent model was changed to water.[3]

The correlation between the experimental binding constants and the values predicted with our model is shown in Figure 4. This model is constructed by the use of 'The unscrambler' software with the first principal component of the analysis.[11] The value of r^2 is **0.936** and of q^2 is **0.899**, thus the model posseses a strong predictive capability. We have then calculated the standard deviation on error prediction ($SDEP_{ex}$) for an external set of complexes not used in the model generation, by keeping out the following complexes: **I**:**3**, **II**:**5**, **III**:**3**, **IV**:**3** and **V**:**4**. On this way we have obtained a quite small value for the $SDEP_{ex}$ of **0.5**.

The experimental and predicted values for the external set of host–guest complexes are compiled in Table 2. Although in the PLS generation, complexes **I**:**5**, **III**:**5** and **IV**:**5** become outliers being excluded from the model, prediction of Ln K_b value for complex **II**:**5** is as

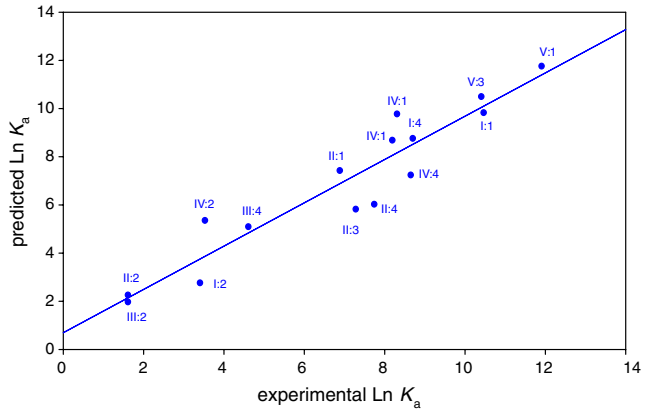


Figure 4. Correlation between experimental and predicted values for Ln K_b.

Table 2. Experimental and predicted values for the external set of complexes

Complex	Experimental Ln K_b	Predicted Ln K_b
I:**3**	9.159	8.727
II:**5**	7.771	7.751
III:**3**	4.949	4.592
IV:**3**	8.476	7.629
V:**4**	9.952	10.34

good as for the other complexes of the external set (Table 2).

By means of this novel approach we have designed 10 new hosts, compiled in Figure 5, based on the compounds whose properties as hosts we had characterized previously.[2–5]

It becomes clear that in order to obtain stronger interactions with biotin and related compounds, the efforts should be directed towards developing hosts that bind the guest by multiple interacting points. This must be achieved with structures where a compromise between preorganization and flexibility is reached.

The Monte Carlo conformational search of the complexes formed by methylbiotin (**1**) and the new hosts **VI**–**XV** affords the most probable structures and the associated energies (Table 3), which after introduction in the model will provide an estimation of the binding constants K_b (Table 4).

From the values compiled in Table 4, the most promising hosts are **XIII**, **XIV** and **XV** and so their preparation will be undertaken to have the experimental validity. All complexes, but **VI**:**1** and **VII**:**1**, show a K_b equal or larger than that of complex **V**:**1** (Table 1), which shows for the moment the strongest interaction between a synthetic receptor and a biotin derivative. The use of a cyclic host, preorganized for the assembling with the guest but at the same time flexible, affords an even higher interaction through several non-covalent bonds. Such type of interaction network is the key point to

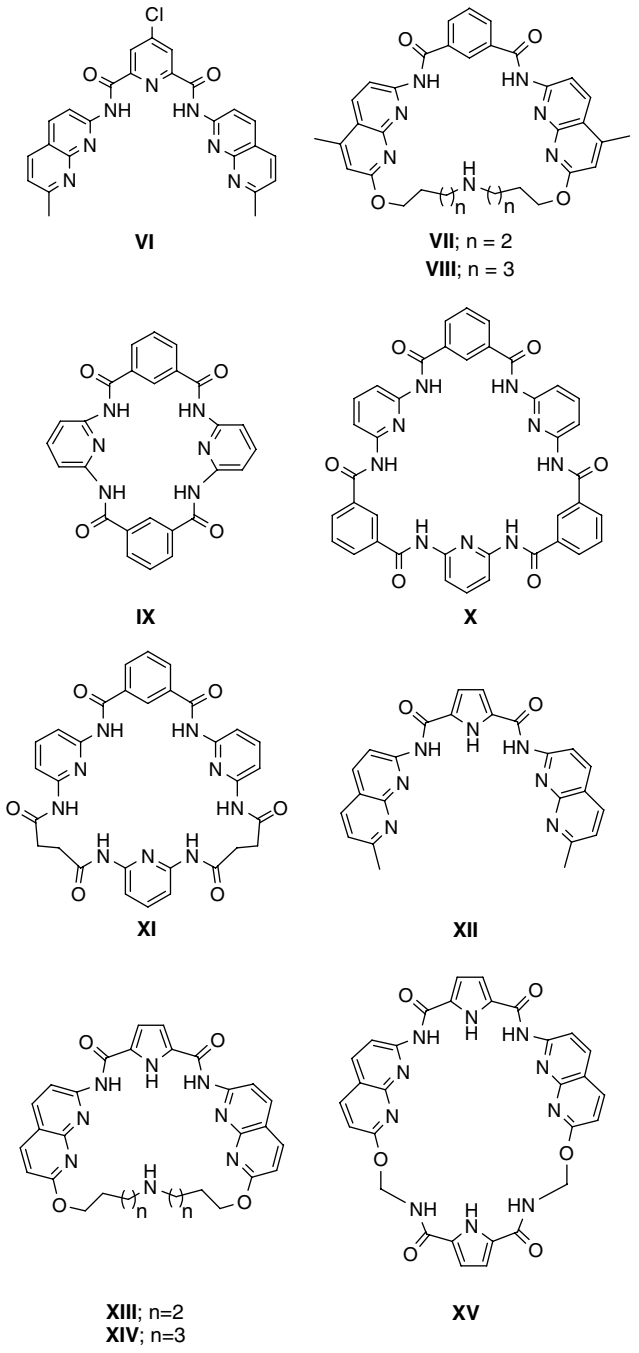


Figure 5. New hosts **VI**–**XV**.

Table 3. Interaction energy values $-E_{min}$ (kJ mol^{-1}) obtained with AMBER and OPLS for the complexes of the new hosts

Complex	AMBER	OPLS
VI:**1**	75.70	109.81
VII:**1**	70.40	122.60
VIII:**1**	91.70	127.90
IX:**1**	81.40	127.90
X:**1**	91.50	131.40
XI:**1**	85.70	132.70
XII:**1**	89.80	124.70
XIII:**1**	95.40	163.50
XIV:**1**	103.60	161.10
XV:**1**	103.20	156.70

Table 4. Predicted K_b for the new complexes

Complex	Ln K_b	Std. deviation	K_b (M^{-1})
VI:**1**	9.885	0.7400	1.96×10^4
VII:**1**	10.489	1.8530	3.59×10^4
VIII:**1**	12.472	1.3590	2.61×10^5
IX:**1**	11.712	1.0970	1.22×10^5
X:**1**	12.729	1.9900	3.37×10^5
XI:**1**	12.402	1.1610	2.43×10^5
XII:**1**	12.083	1.1130	1.77×10^5
XIII:**1**	15.512	1.1720	5.45×10^6
XIV:**1**	15.930	1.3740	8.29×10^6
XV:**1**	15.558	1.6390	5.71×10^6

approximate the binding mode of biotin in the biological environment.

We have proved that the use of compounds showing C_3 symmetry, like hosts **IV** and **V**, gives rise to a three times larger binding constants due to the higher probability and, therefore a more favourable entropic cost, of finding the right way for complexation.[4] Then, another important factor to be considered is the symmetry of the host. Host **IX** presents a C_2 symmetry and host **X** a C_3 symmetry, so the K_b for these two compounds is expected to be around two and three times higher than the predicted values, as modelling does not account for the probability factor.[4] As a title of example, the Monte Carlo predicted structures for complexes **X**:**1** and **XV**:**1**, using the AMBER force field, are shown in Figures 6 and 7.

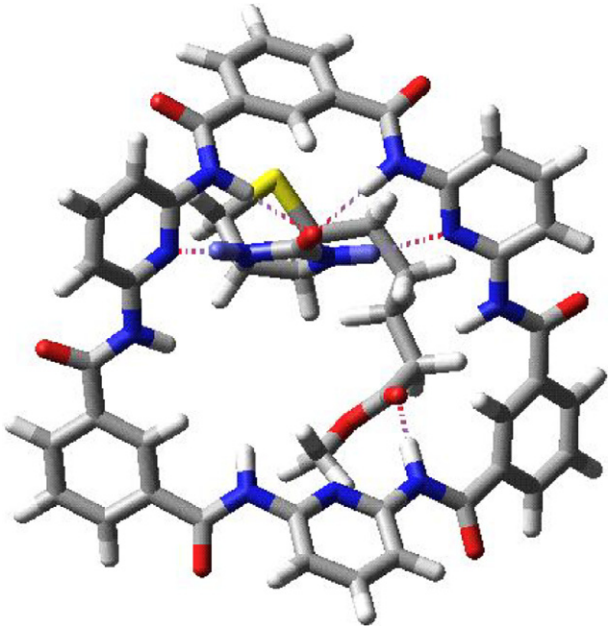


Figure 6. Structure of complex **X**:**1**.

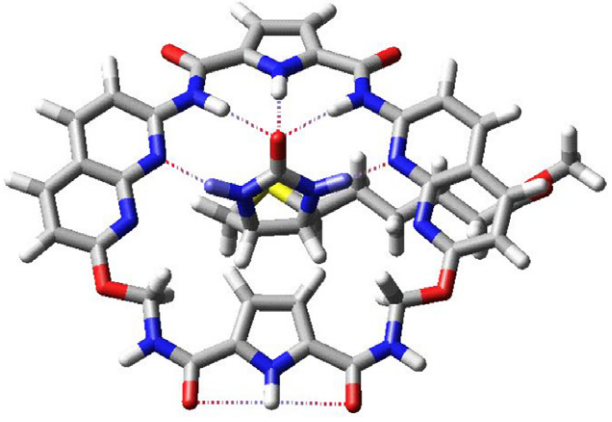


Figure 7. Structure of complex **XV**:**1**.

3. Conclusions

We have developed a Partial Least Squares Cross-Validated model by using the experimental binding constants of 23 complexes and the interaction energies obtained by Monte Carlo conformational search with AMBER and OPLS force fields.

By means of this new model we have been capable to provide an estimation of the binding constants for 10 modelled complexes between new designed hosts **VI**–**XV** and methylbiotin (**1**), and to find the most probable structures for the host–guest interaction.

In summary, the Partial Least Squares Cross-Validated model is a practical and extremely useful tool in the field of host–guest chemistry of biotin analogues.

Acknowledgements

Thanks are given to MCyT of Spain for financial support, project numbers BQU2003-00976 and CTQ2006-02586.

References and notes

1. Dixon, R. W.; Radmer, R. J.; Kuhn, B.; Kollman, P. A.; Yang, J.; Raposo, C.; Wilcox, C. S.; Kulm, L. A.; Stayton, P. S.; Behnke, C.; Le Trong, I.; Stenkamp, R. *J. Org. Chem.* **2002**, *67*, 1827–1837.
2. Claramunt, R. M.; Herranz, F.; Santa María, M. D.; Jaime, C.; de Federico, M.; Elguero, J. *Biosens. Bioelectron.* **2004**, *20*, 1242–1249.
3. Claramunt, R. M.; Herranz, F.; Santa María, M. D.; Pinilla, E.; Torres, M. R.; Elguero, J. *Tetrahedron* **2005**, *61*, 5089–5100.
4. Herranz, F.; Santa María, M. D.; Claramunt, R. M. *J. Org. Chem.* **2006**, *71*, 2944–2951.
5. Herranz, F.; Santa María, M. D.; Claramunt, R. M. *Molecules* **2006**, *11*, 478–485.
6. Frank, I. E.; Friedman, J. H. *Technometrics* **1993**, *395*, 109–148.
7. Höskuldson, A. *J. Chemom.* **1988**, *2*, 211–228.
8. Goswami, S. P.; Mukherjee, R. *Tetrahedron Lett.* **1997**, *38*, 1619–1622.
9. Goswami, S. P.; Dey, S. *J. Org. Chem.* **2006**, *71*, 7280–7287.
10. MacroModel, Schrödinger LLC, 2004. <http://www.schrodinger.com/Products/macromodel.html>.
11. CAMO Software, Inc. web: <http://www.camo.com>.