Supporting Information

Chirality Sensing and Discrimination of Lysine Derivatives in Water with a Dyn[4]arene

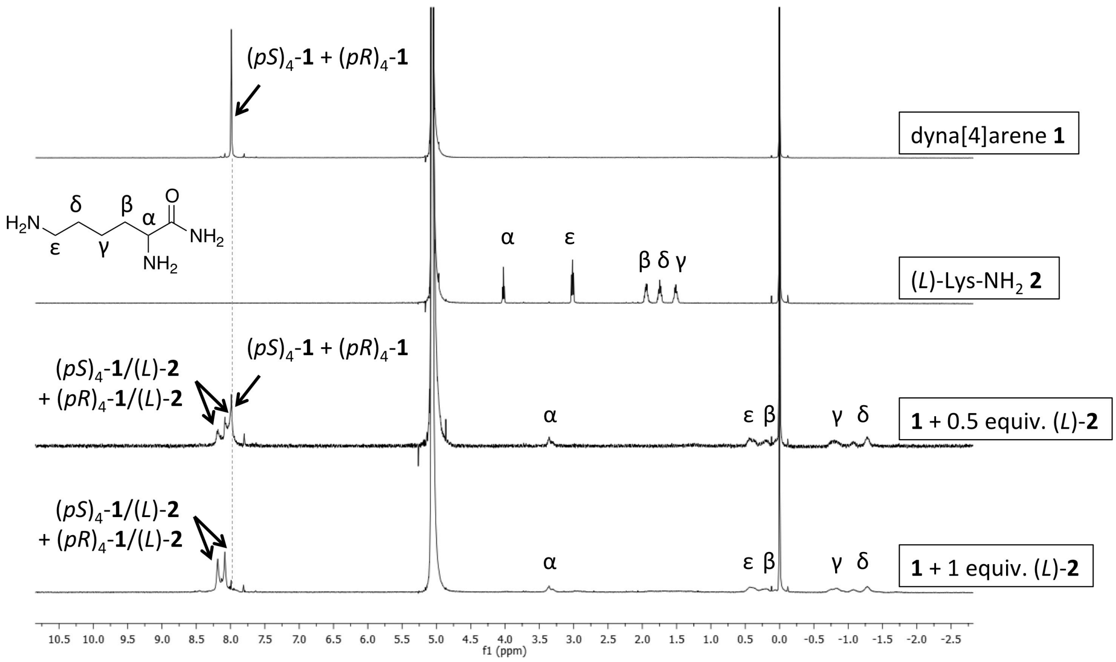
Laurent Vial, Melissa Dumartin, Marion Donnier-Maréchal, Florent Perret, Jean-Patrick Francoia, and Julien Leclaire

NMR titration

The structure of the complex between **1** and (*L*)-**2** was investigated using proton NMR spectroscopy. Conditions: 500 MHz, D_2O, phosphate buffer 200 mM, pH 7.4, 278 K, [**1**]=1 mM (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt was used as internal standard for calibration). The spectra were assigned with COSY experiments.

Upon binding to **1** a strong ring current-induced upfield shift was observed for the protons of the side chain of (*L*)-**2**, while the α-proton was hardly shifted. These observations suggested that the side chain of (*L*)-**2** was located within the cavity of the receptor.

Furthermore, the binding of (*L*)-**2** was slow on the NMR timescale and of a similar order of magnitude for both enantiomeric macrocycles. The formation of two equimolar diastereoisomeric complexes (*L*)-**2**/(pS)$_4$-**1** and (*L*)-**2**/(pR)$_4$-**1** suggested that (*L*)-**2** bound both enantiomers of **1** with a similar affinity.



Statistical details

The ICD amplitude at 318 nm were chosen as input dimensions for the Linear Disciminant Analysis (LDA) procedure. For each analyte, triplicates were obtained at two [guest]/[host **1**] ratio of 1 and 5, respectively.

First, the data were scaled with the "preprocessing.scale()" function of the scikit-learn Python module.[1] Then, the LinearDiscriminantAnalysis class of the module was used as a classifier and was fitted on the input data. The Eigen solver provided the dimensions that maximize the separation between classes, and the data were projected to the resulting two dimensional subspace (no dimensionality reduction was necessary).[2] The confidence intervals for the true means of the classes were plotted at a level of 99%.[3]

For the Jack Knife (or leave-one-out) validation, the data were first scaled with the procedure mentioned above. One replicate was left out of the training set, and the LDA classifier was fitted on the input data. The excluded replicate was then used as a blind sample, and its class was predicted with the "predict()" function of the LinearDiscriminantAnalysis class of the scikit-learn Python module. A correct prediction of the analyte's class increases the Jack Knife score, while an incorrect prediction decreases the score. The procedure was repeated for each replicate of the data sets. For the dipeptides and amino acids data sets, Jack Knife scores of 73.3 % and 100 % were obtained, respectively.

The full Python scripts are available upon simple request.

1. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel,learn B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.,* 2011, **12**, 2825–2830.

2. For more details about the employed algorithms, see: http://scikit-learn.org/stable/modules/lda_qda.html

3. The confidence intervals were calculated according to the following procedure: https://github.com/joferkington/oost_paper_code/blob/master/error_ellipse.py