Supporting Information

© Wiley-VCH 2005

69451 Weinheim, Germany

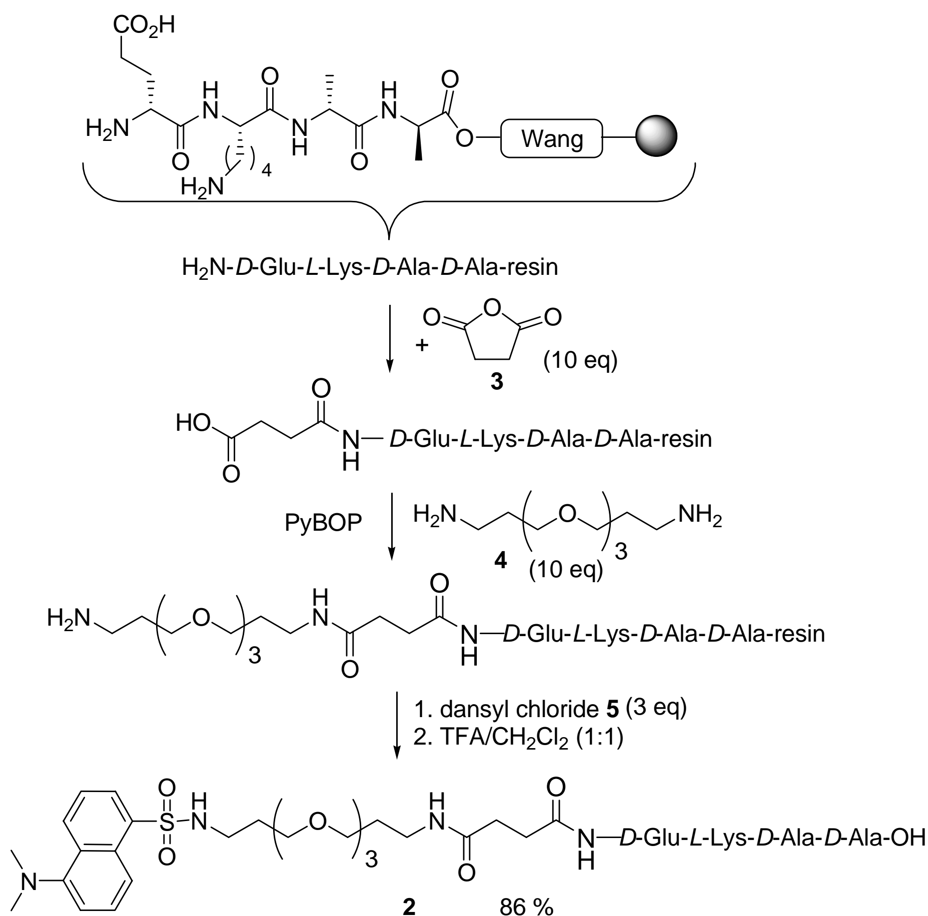
Charge Interactions do the Job: A Combined Statistical and Combinatorial Approach to Find Artificial Receptors for Tetrapeptide Binding in Water.

Carsten Schmuck,* Martin Heil, Josef Scheiber and Knut Baumann

Experimental Section

General remarks: Reaction solvents were dried and distilled under argon before use. All other reagents were used as obtained from either Aldrich or Fluka. ^{1}H and ^{13}C NMR shifts are reported relative to the deuterated solvents. Peak assignments are based on either DEPT, 2D NMR studies and/or comparison with literature data. IR spectra were recorded using samples prepared as tablets (KBr). Melting points are not corrected.

Synthesis of the dansylated tetrapeptide substrates 2:



The fluorophore-labeled tetrapeptide **2** was synthesized on Wang resin by a standard protocol: Wang resin (300 mg, 1.11 mmolg$_{-1}$, 0.33 mmol) was swollen in CH_2Cl_2/DMF (8:2) for 1.5 h, and the first amino acid was coupled to the resin by use of Fmoc-D-Ala (2.0 equiv) respectively, diisopropylcarbodiimide (DIC, 2.0 equiv), and dimethylaminopyridine (DMAP, 0.1 equiv) in CH_2Cl_2/MeOH (8:2), with a reaction time of 20 h. The coupling step was repeated twice. The Fmoc group was cleaved with piperidine in DMF (20 %). The other three amino acids (D-Ala, L-Lys and D-Glu) were coupled by use of the following procedure: 2.5 equiv of amino acid and 2.5 equiv of (benzotriazol-1-yloxy)trispyrrolidinophosphonium hexafluorophosphate (PyBOP) in DMF containing 3% N-methyl morpholine (NMM; 10 mL) for 4 h. The free amino function of the tetrapeptide was then coupled with succinic anhydride (10.0 equiv) in CH_2Cl_2/DMF mixture (8:2) for 3 h. The resulting carboxylic acid was coupled with 1,13-diamino-4,7,10-trioxaundecane with PyBOP (10.0 equiv) in DMF (reaction time 3 h). The resulting amine was treated with dansyl chloride (3.0 equiv) in DMF for 20 h. Cleavage from the resin was performed with a CH_2Cl_2/TFA mixture (1:1) for 2 h. The solvent was evaporated, and the resulting red oil was treated with dry diethyl ether to obtain a light green solid that was analytically pure.

Yield: 152 mg, 46 %, 0.15 mmol

Melting point: 172 °C

^{1}H-NMR (600 MHz, DMSO-d6): δ = 1.19 (d, $^3J_{H\text{-}H}$ = 7.1 Hz 3 H; CH_3), 1.23 (d, $^3J_{H\text{-}H}$ = 7.3 Hz 3 H; CH_3),.1.21-1.35 (m, 2 H; CH_2), 1.50-1.60 (m, 6 H; CH2), 1.63 (m, 1 H; CH_2), 1.72 (m, 1 H; CH_2), 1.83 (m, 1 H; CH_2), 2.20-2.40 (m, 6 H; CH_2), 2.71 (m, 2 H; CH_2), 2.82 (q, $^3J_{H\text{-}H}$ = 6.0 Hz, 2 H; CH_2), 3.03 (m, 2 H; CH_2), 3.09 (s, 6 H; CH_3), 3.21-3.24 (m, 4 H; CH_2), 3.32-3.37 (m, 4 H;, CH_2), 3.39-3.43 (m, 4 H;, CH_2), 3.44-3.50 (m, 2 H;, CH_2), 4.13-4.20 (m, 3 H; CH), 4.27 (m, 1 H; CH), 7.73 (m, 3 H; ar-CH), 7.87 (t, $^3J_{H\text{-}H}$ = 5.1 Hz, 1 H; NH), 7.96 (s, 3 H; NH_3^+), 7.98 (d, $^3J_{H\text{-}H}$ = 7.7 Hz, 1 H; NH), 8.02 (t, $^3J_{H\text{-}H}$ = 7.7 Hz, 1 H; NH), 8.06-8.15 (m, 3 H; CH), 8.17 (d, $^3J_{H\text{-}H}$ = 7.5 Hz, 1 H; NH), 8.56 (t, $^3J_{H\text{-}H}$ = 5.3 Hz, 1 H; NH), 8.78 (d, $^3J_{H\text{-}H}$ = 7.3 Hz, 1 H; NH);

^{13}C-NMR (150 MHz, DMSO-d6): δ = 17.3, 18.3 (CH_3), 22.4, 25.9, 26.0, 26.6, 27.2, 29.5, 29.6, 30.4, 30.9, 30.9, 31.0, 36.0, 38.6 (CH_2), 40.2 (CH_3), 45.8, 47.4, 47.8, 52.4 (CH), 67.2, 68.0, 69.3, 69.4, 69.5, 69.6 (CH_2), 125.1, 125.3, 125.5, 127.8, 128.6, 129.0, 129.1 (CH), 136.6, 145.5, 171.2, 171.5, 171.7, 172.0, 172.3, 174.0, 174.1 (C_q);

MS (ESI, DMSO/H_2O) m/z: 953 [M^+], 864 [$C_{40}H_{62}N_7O_{12}S^+$], 665 [$C_{32}H_{50}N_5O_8S^+$], 536 [$C_{26}H_{38}N_3O_7S^+$], 454 [$C_{22}H_{34}N_3O_5S^+$];

HR-MS (pos. ESI) 953.467 (calculated for $C_{43}H_{68}N_8O_{14}S^+$ 953.465);

FT-IR (KBr disk): $\tilde{\nu}$ [cm^{-1}] = 3430 (s), 3286 (s), 2938 (m), 2868 (m), 1732 (m), 1653 (s), 1541 (m), 1458 (w), 1396 (w), 1320 (w), 1212 (w), 1143 (w), 795 (w)

Synthesis of receptors 3 and 4 (general protocol): The synthesis was performed on Rink amide resin following a standard protocol: Rink amide resin (300 mg, 0.74 mmol/g, 0.22 mmol) was swollen in DMF for 1.5 h. The Fmoc protection group was removed by agitation with piperidine in DMF (20 %) for 20 min. Coupling conditions for the amino acids: 2.5 eq of Fmoc amino acid; 2.5 eq PyBOP in DMF containing NMM 3 % (10 mL). The mixture was shaken for 3.5 h to ensure quantitative coupling. The yield was monitored indirectly by the UV absorption of the Fmoc-piperidine adduct at 300 nm. The attachment of the 5-guanidiniocarbonylpyrrol-2-carboxylic acid was performed under related conditions: 2.5 eq carboxylic acid, 2.5 eq PyBOP and DMF containing 5 % NMM, and elongated reaction times of 24 h. To provide a quantitative coupling the last step was repeated. The product was cleaved from solid support by shaking the resin with a CH_2Cl_2/TFA mixture (5:95). The solvent was evaporated and the remaining oil treated with dry ether. To obtain the hydrochloride salt the resulting white solid was dissolved in 1 ml of methanol and acidified with 4 mL of 0.1 N hydrochloric acid and lyophilized. This was repeated three times to ensure complete exchange of trifluoroacetate for chloride.

Receptor 3: CBS-KKF Yield: 151 mg, 98 %, 0.21 mmol

Melting point: 247 °C (decomposition)

^{1}H-NMR (400 MHz, DMSO-d6): δ = 1.16-1.28 (m, 2 H; CH_2), 1.29-1.42 (m, 2 H; CH_2), 1.44-1.75 (m, 8 H; CH_2), 2.66-2.71 (m, 2 H; CH_2), 2.73-2.79 (m, 2 H; CH_2), 2.80-2.86 (m, 1 H; CH), 2.98-3.02 (m, 1 H; CH_2), 4.12-4.17 (m, 1 H; CH), 4.40-4.45 (m, 2 H; CH), 6.91 (s, 1 H; py-CH), 7.12 (s, 1 H; NH_2), 7.18-7.25 (m, 5 H; ar-CH), 7.45 (s, 1 H; NH_2), 7.60 (m, 1 H; py-CH), 7.88 (d, $^3J_{H\text{-}H}$ = 8.1 Hz, 1 H; NH), 7.92 (bs, 6 H; NH_3^+) 8.27 (d, $^3J_{H\text{-}H}$ = 7.8 Hz, 1 H; NH), 8.51 (s, 2 H; gua-NH_2), 8.67 (d, $^3J_{H\text{-}H}$ = 7.6 Hz, 1 H; NH), 8.71 (s, 2 H; gua-NH_2), 12.17 (s, 1 H; gua-NH), 12.52 (s, 1 H; py-NH);

^{13}C-NMR (150 MHz, DMSO-d6): δ = 22.3, 22.7, 26.6, 26.7, 31.2, 31.3, 37.8, 38.7, 40.2 (CH_2), 53.0, 53.0, 53.8 (CH), 113.8, 116.0 (py-CH), 125.8, 132.5 (py-C_q), 126.4, 128.2, 129.4 155.7 (gua-C_q), 137.9, 159.1, 159.9, 171.3, 171.8, 173.0 (C_q);

MS (ESI, DMSO/H_2O) m/z: 599 [M^+], 377 [$C_{21}H_{37}N_6O_3^+$], 300 [$M+H^{++}$], 179 [$C_7H_7N_4O_2^+$];

FT-IR (KBr disk): $\tilde{\nu}$ [cm^{-1}] = 3326 (s), 3061 (s), 2949 (m), 1702 (s), 1654 (s) 1541 (m), 1472 (w), 1276 (w), 1198 (w), 815 (w), 754 (w).

Receptor 4: CBS-KYK Yield: 110 mg, 69 %, 0.15 mmol

Melting point: 209 °C

^{1}H-NMR (400 MHz, DMSO-d6): δ = 1.20-1.32 (m, 4 H; CH_2), 1.47-1.69 (m, 6 H; CH_2), 1.65-1.70 (m, 2 H; CH_2), 2.69-2.77 (m, 5 H; CH_2), 2.90-2.96 (m, 1 H; CH), 4.33-4.43 (m, 2 H; CH), 6.59 (d, $^3J_{H-H}$ = 8.6 Hz, 2 H; ar-CH), 6.91 (s, 1 H; py-CH), 7.01 (d, $^3J_{H-H}$ = 8.6 Hz, 2 H; ar-CH), 7.06 (s, 1 H; NH_2), 7.25 (s, 1 H; NH_2), 7.61 (m, 1 H; py-CH), 7.87 (d, $^3J_{H-H}$ = 8.1 Hz, 1 H; NH), 7.97 (bs, 6 H; NH_3^+) 8.20 (d, $^3J_{H-H}$ = 8.1 Hz, 1 H; NH), 8.53 (s, 2 H; gua-NH_2), 8.66 (d, $^3J_{H-H}$ = 7.6 Hz, 1 H; NH), 8.72 (s, 2 H; gua-NH_2), 9.18 (bs, 1H; OH), 12.15 (s, 1 H; gua-NH), 12.51 (s, 1 H; py-NH);

^{13}C-NMR (100 MHz, DMSO-d6): δ = 22.3, 22.6, 26.7, 26.8, 31.3, 31.536.3, 38.7 (CH_2), 52.4, 53.2, 54.6 (CH), 113.8, 116.0 (py-CH), 115.0, 130.2 (ar-CH), 125.8, 132.5 (py-C_q), 155.8 (gua-C_q), 127.9, 156.0, 159.9, 171.0, 171.6, 173.5 (C_q);

MS (ESI, DMSO/H_2O) m/z: 615 [$M+H^+$], 437 [M^+- $C_7H_9N_4O_2$], 308 [$M+H^{++}$], 179 [$C_7H_7N_4O_2^+$];

FT-IR (KBr disk): $\tilde{\nu}$ [cm^{-1}] = 3421 (s), 2938 (m), 1710 (s), 1654 (s) 1558 (m), 1477 (w), 1268 (w), 1194 (w), 821 (w), 758 (w).

Synthesis of the acetylated tetrapeptide N-Ac-EKAA-OH: The acetylated tetrapeptide N-Ac-EKAA-OH was synthesized on Wang resin by a standard protocol: Wang resin (300 mg, 1.11 mmol g^{-1}, 0.33 mmol) was swollen in CH_2Cl_2/DMF (8:2) for 1.5 h, and the first amino acid was coupled to the resin by use of Fmoc-D-Ala (2.0 equiv), diisopropylcarbodiimide (DIC, 2.0 equiv), and dimethylaminopyridine (DMAP, 0.1 equiv) in CH_2Cl_2/MeOH (8:2), with a reaction time of 20 h. The coupling step was repeated twice. The Fmoc group was cleaved with piperidine in DMF (20 %). The other three amino acids (D-Ala, L-Lys and D-Glu) were coupled by use of the following procedure: 2.5 equiv of amino acid and 2.5 equiv of (benzotriazol-1-yloxy)trispyrrolidinophosphonium hexafluorophosphate (PyBOP) in DMF containing 3% N-methyl morpholine (NMM; 10 mL) for 4 h. The resulting free amino functions were acetylated using 10 equiv of acetic acid anhydride (340 mg, 3.3 mmol) and DMAP (0.1 eq) in a mixture of CH_2Cl_2/DMF (8:2) for 8 h. The resin was washed thoroughly with CH_2Cl_2, methanol, diethylether and again CH_2Cl_2 to remove traces of DMF. Cleavage of the product from the resin was achieved by treatment with 50 % TFA in CH_2Cl_2 and precipitation by adding dry ether to the solution. The white solids were lyophilized twice with water containing 2 mL of 0.1 N HCl to ensure the presence of a hydrochloric salt. The obtained white solids were analytically pure.

Tetrapeptide N-Ac-EKAA-OH Yield: 52 %, 82 mg, 0.17 mmol

Melting point: 187 °C

^{1}H-NMR (600 MHz, DMSO-d6): δ = 1.18 (d, $^3J_{H-H}$ = 7.0 Hz 3 H; CH_3), 1.26 (d, $^3J_{H-H}$ = 7.3 Hz 3 H; CH_3), 1.21-1.35 (m, 4 H; CH_2), 1.49-1.56 (m, 4 H; CH2), 1.63 (m, 1 H; CH_2), 1.72 (m, 1 H; CH_2), 1.83 (m, 1 H; CH_2), 1.84 (s, 3 H; CH_3), 2.23 (m, 2 H; CH_2), 2.72 (m, 2 H; CH_2), 4.15-4.22 (m, 3 H; CH), 4.30 (m, 1 H; CH), 7.93 (b s, 3 H; NH_3^+), 8.04 (d, $^3J_{H-H}$ = 7.9 Hz, 1 H; NH), 8.04 (d, $^3J_{H-H}$ = 7.6 Hz, 1 H; NH), 8.09 (d, $^3J_{H-H}$ = 7.6 Hz, 1 H; NH), 8.13 (d, $^3J_{H-H}$ = 7.3 Hz, 1 H; NH);

^{13}C-NMR (150 MHz, DMSO-d6): δ = 16.9, 18.2, 22.4 (CH_3), 22.1, 26.4, 27.1, 30.1, 31.0, 38.4 (CH_2), 47.4, 47.6, 52.1, 52.4 (CH), 169.8, 171.2, 171.6, 171.9, 174.0, 174.1 (C_q);

MS (ESI, DMSO/H_2O) m/z: 460 [M^++H], 396 [M^++Na-$C_3H_6NO_2$], 389 [$C_{18}H_{33}N_2O_7^+$], 290 [$C_{13}H_{24}NO_6^+$], 203 [$C_8H_{15}O_5$-Na^+];

FT-IR (KBr disk): $\tilde{\nu}$ [cm^{-1}] = 3422 (s), 3074 (s), 2922 (s), 1728 (m), 1654 (s), 1542 (m), 1459 (w), 1378 (w), 1213 (m), 1170 (m), 1046 (w) 826 (w)

Qualitative solid phase Screening. To probe the entire receptor library **1** qualitatively for its binding properties, aliquots of the 512 resin bound deprotected receptors **1** were pooled and the combined mixture incubated with a 5 μM solution of the tetrapeptide substrate **2** in 20 μM bis-tris-buffer of pH = 6.0 in water. After the supernatant solution was washed off, the beads were screened under UV light using a fluorescence microscope. The selective binding of the tetrapeptide substrate **2** by some - but not all - of the 512 receptors **1** can be observed as indicated by the strong fluorescence activity of individual beads (Figure below). Only those beads, on which the attached receptor is capable to bind the peptide even in water, show the characteristic fluorescence of the dansyl group. All the other receptors which do not bind the peptide under these experimental conditions remain dark. The percentage of beads in the mixture taking up the fluorescence is directly correlated with the concentration of the substrate solution used for incubation. This underlines that the observed differences in fluorescence activity are indeed due to different binding

affinities of the individual receptors and not to a selective fluorescent quenching within the complex (which would be concentration independent).

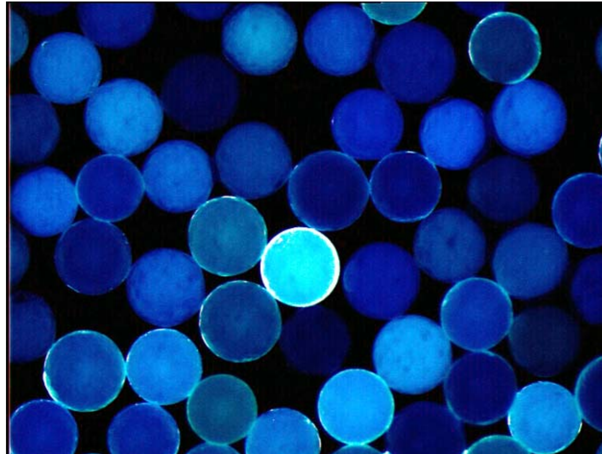


Figure 1: On bead binding assay in buffered water ([substrate] = 5 μM, 20 μM bis-tris buffer, pH = 6.0). The strong fluorescence activity indicates selective binding of tetrapeptide **2** (EKAA) to selected cationic receptors **1** (top), whereas the inverse tetrapeptide **6** (AAKE) is bound less efficiently at the same concentration (bottom).

Quantitative solid phase screening. To determine the binding affinities of each of these 512 receptors, the individual members of the library were screened in a quantitative fluorescence assay using a high trough put microtiter plate reader. The IRORI-tagging technology provides the individual library members locally separated and in amounts of ca. 20-30 mg bead each, which is enough material for a quantitative screening. For every receptor a precisely measured sample (15 – 25 mg resin with a 0.22 mmol/g loading as determined by a quantification of the Fmoc-piperidine adduct during receptor synthesis) was incubated with 2 mL of a 39 μM solution of the dansylated tetrapeptide 5 in 200 μM bis-tris-buffer with pH = 6.0 for 20 h. After equilibration 200 μL of the supernatant solution were taken off and their

fluorescence intensity was measured. From the fluorescence intensity before and after incubation and the loading of the resin the association constants K_{ass} for each receptor can be calculated.

$$K_{ass} = \frac{[R-S]}{[R]\cdot[S]}$$

- The remaining fluorescence intensity after incubation provides the concentration of the free substrate [S] in solution via a calibration plot.
- The concentration of the complex ([R-S]) can be calculated from the difference between the initial fluorescence intensity in solution before the incubation and the remaining fluorescence: $[R\text{-}S] = [S]_o - [S]$
- The remaining concentration of the free receptor [R] can be calculated from the amount of the receptor initially bound to the resin $[R]_o$ and the concentration of the receptor substrate complex [R-S]. The initial concentration of the receptor $[R]_o$ on the solid support is given by the quantity of resin used and its loading corrected for the different molecular masses of the individual receptors.

Complex formation in free solution. To determine the complex stabilities also quantitatively in free solution, UV- titration experiments were performed. Stock solutions of the tetrapeptide **2** ($c = 7.7\cdot10^{-4}$ M^{-1}) and the receptors **3** and **4** ($c = 5.4\cdot10^{-5}$ M^{-1}) were freshly prepared in buffered water (bis-tris-puffer, pH = 6.15, $c = 1.6\cdot10^{-3}$ M^{-1}). Aliquots of the tetrapeptide were then added to a solution of the receptor and the changes in the UV-spectra were recorded after each addition. The decrease of the absorbance of the pyrrole band at 300 nm was followed. A non-linear curve fitting using a 1:1-complexation model was used to calculate the binding constants from the isotherms. The tetrapeptide substrate **2** also has a small but significant absorption at this wavelength which has to be taken into account in the analysis of the data as well as the changes due to the dilution of the sample during titration. The absorption

coefficients of the receptors (ε = 25.300 M cm^{-1} for **3** and ε = 22.250 M cm^{-1} for **4**) as well as the tetrapeptide substrate **2** (ε = 200 M cm^{-1}) were determined from independent measurements and used as constants in the data analysis. The only fitting parameters were hence the absorption coefficient of the complex and the binding constant itself.

QSAR analysis. In the first step of the QSAR analysis a so-called structure descriptor was computed which translates important properties of the studied molecules into numbers. Next, a linear mathematical model relating the variables of the structure descriptor to their activity values was estimated. This model is of the general form:

$$y_i = b_0 + b_1 \cdot x_{i,1} + b_2 \cdot x_{i,2} + ... + b_j \cdot x_{i,j} + ... + b_m \cdot x_{i,m}$$

where y_i represents the activity value of the *i*th molecule (*i*: 1, …, *n*), $x_{i,j}$, is the value of the *j*th structure descriptor (*j*: 1, …, *m*) for the *i*th molecule, and b_0 is the intercept term, and b_j is the regression coefficient (weight) for the *j*th variable of the structure descriptor.

This model is commonly computed with a linear regression technique. Here, Partial Least Squares Regression (PLSR) was used, which is the quasi-standard in QSAR analysis (Lindberg, W.; Persson, J. A.; Wold, S. *Anal. Chem.* **1983**, *55*, 643-648). PLSR has the advantage that is can handle underdetermined and multicollinear systems of equations. The tuning parameter in PLSR is the model dimension (complexity). This was determined by leave-one-out cross-validation (LOO-CV).[Fehler! Textmarke nicht definiert.a] In leave-one-out cross-validation a single molecule is removed

from the data set. The remaining molecules are used to construct the model with different dimensions. For each dimension the removed molecule is predicted from the model and the respective error is stored (cross-validated prediction error). This is done in turn for each molecule in the data set. The optimal dimension of the PLSR model is taken as the one that minimizes the mean squared error of prediction for the entire data set. The quality of the resulting model was characterized by the cross-validated squared correlation coefficient ($R^2_{CV\text{-}1}$) and by the cross-validated root mean squared error of prediction ($RMSEP_{CV\text{-}1}$).[Fehler! Textmarke nicht definiert.b] To thoroughly characterize the predictive ability of the model, test set prediction was also used. Here, the data set is split into a training data set and a test data set. The training data set is used to determine the optimal dimension of the PLSR model with LOO-CV. After having selected this dimension, the test set is predicted from the selected model. 100 different random splits into test set (33%) and training set (67%) were computed and the results were averaged to yield the figures of merit $R^2_{Test,Avg}$ and $RMSEP_{Test,Avg}$.