

# **Pharmacophore Assessment Through 3-D QSAR: evaluation of the predictive ability on new derivatives by the application on a serie of antitubercular agents.**

Laura Friggeri,<sup>§,†</sup> Flavio Ballante,<sup>\*,§,‡</sup> Rino Ragno,<sup>\*,‡</sup> Ira Musmuca,<sup>‡</sup> Daniela De Vita,<sup>†</sup> Fabrizio Manetti,<sup>¤</sup> Mariangela Biava,<sup>†</sup> Luigi Scipione,<sup>†</sup> Roberto Di Santo,<sup>+,†</sup> Roberta Costi,<sup>+,†</sup> Marta Feroci,<sup>^</sup> and Silvano Tortorella.<sup>†</sup>

<sup>‡</sup>Rome Center for Molecular Design, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185 Roma, Italy.

<sup>†</sup>Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185 Roma, Italy.

<sup>\*</sup>Istituto Pasteur-Fondazione Cenci Bolognetti, Dipartimento di “Chimica e Tecnologie del Farmaco”, “Sapienza” Università di Roma, P.le A. Moro 5, 00185 Rome, Italy

<sup>¤</sup>Dipartimento di Biotecnologie, Chimica e Farmacia, Università degli Studi di Siena, Via Aldo Moro 2, I-53100 Siena, Italy.

<sup>^</sup>Dipartimento di Scienze di Base e Applicate per l’Ingegneria, Sapienza University of Rome, Via Castro Laurenziano 7, I-00161 Rome, Italy.

<sup>§</sup>*L.F. and F.B. contributed equally to this work*

## **SUPPORTING INFORMATION**

### **CONTENTS**

Training Set compounds references	S2-S3
Mono-probes models analysis	S4-S14
MPGRS model analysis	S15-S19
Mono-probes predictions	S20-S21
MPGRS predictions	S22
AutoGrid settings	S23
Chiral HPLC analysis	S24
Surflex and pharmacophoric alignment results comparison	S25-S29

Reference	Reference compd N°	New compd N°
(Bioorganic & Medicinal Chemistry 13 (2005) 1221–1230)		
“	6	1
“	7	2
“	8	3
“	9	4
“	10	5
“	11	6
“	12	7
“	13	8
“	14	9
“	17	10
“	18	11
“	19	12
“	21	13
“	23	14
“	26	15
“	27	16
“	28	17
“	29	18
“	30	19
“	31	20
“	32	21
“	33	22
“	34	23
“	35	24
“	36	25
“	38	26
“	39	27
“	40	28
“	41	29
“	42	30
“	44	31
(J. Med. Chem. 2006, 49, 4946-4952)		
“	1	32
“	2	33
“	4	34
“	5	35
“	8	36
“	9	37
“	10	38
“	11	39
“	14	40
“	15	41
“	16	42
“	18	43
“	19	44
“	21	45
“	22	46
“	23	47
“	24	48
“	25	49

“	26	50
“	27	51
“	28	52
(J. Med. Chem. 2008, 51, 3644–3648)	2	53
“	4	54
“	5	55
“	6	56
“	8	57
“	9	58
“	10	59
“	11	60
“	12	61
“	13	62
“	14	63
“	15	64
(European Journal of Medicinal Chemistry 44 (2009) 4734–4738)	2a	65
“	2b	66
“	2c	67
“	2d	68
“	2e	69
“	2g	70
“	2h	71

**Table S1.** Training set compounds' original and new enumeration.

$r^2$	model	P	PC1	PC2	PC3	PC4	PC5
1	1	A	0.35	0.80	0.89	0.92	0.94
$q^2_{\text{LOO}}$	1	A	0.48	0.83	0.86	0.85	0.85
$q^2_{\text{K5FCV}}$	1	A	0.28	0.72	0.79	0.80	0.80
$r^2$	2	C	0.35	0.80	0.89	0.92	0.94
$q^2_{\text{LOO}}$	2	C	0.28	0.73	0.80	0.81	0.81
$q^2_{\text{K5FCV}}$	2	C	0.28	0.72	0.79	0.80	0.80
$r^2$	3	HD	0.40	0.81	0.87	0.89	0.91
$q^2_{\text{LOO}}$	3	HD	0.31	0.70	0.75	0.75	0.75
$q^2_{\text{K5FCV}}$	3	HD	0.31	0.69	0.74	0.73	0.73
$r^2$	4	NA	0.35	0.81	0.89	0.93	0.94
$q^2_{\text{LOO}}$	4	NA	0.28	0.74	0.79	0.80	0.80
$q^2_{\text{K5FCV}}$	4	NA	0.28	0.73	0.78	0.79	0.80
$r^2$	5	N	0.35	0.81	0.89	0.93	0.94
$q^2_{\text{LOO}}$	5	N	0.28	0.74	0.79	0.80	0.80
$q^2_{\text{K5FCV}}$	5	N	0.28	0.73	0.78	0.79	0.79
$r^2$	6	OA	0.35	0.80	0.89	0.92	0.94
$q^2_{\text{LOO}}$	6	OA	0.28	0.74	0.79	0.79	0.79
$q^2_{\text{K5FCV}}$	6	OA	0.28	0.73	0.78	0.78	0.79
$r^2$	7	e	0.53	0.60	0.68	0.74	0.77
$q^2_{\text{LOO}}$	7	e	0.46	0.54	0.60	0.57	0.57
$q^2_{\text{K5FCV}}$	7	e	0.46	0.54	0.60	0.56	0.56
$r^2$	8	d	0.15	0.68	0.81	0.86	0.87
$q^2_{\text{LOO}}$	8	d	0.11	0.63	0.75	0.79	0.80
$q^2_{\text{K5FCV}}$	8	d	0.11	0.62	0.74	0.77	0.77

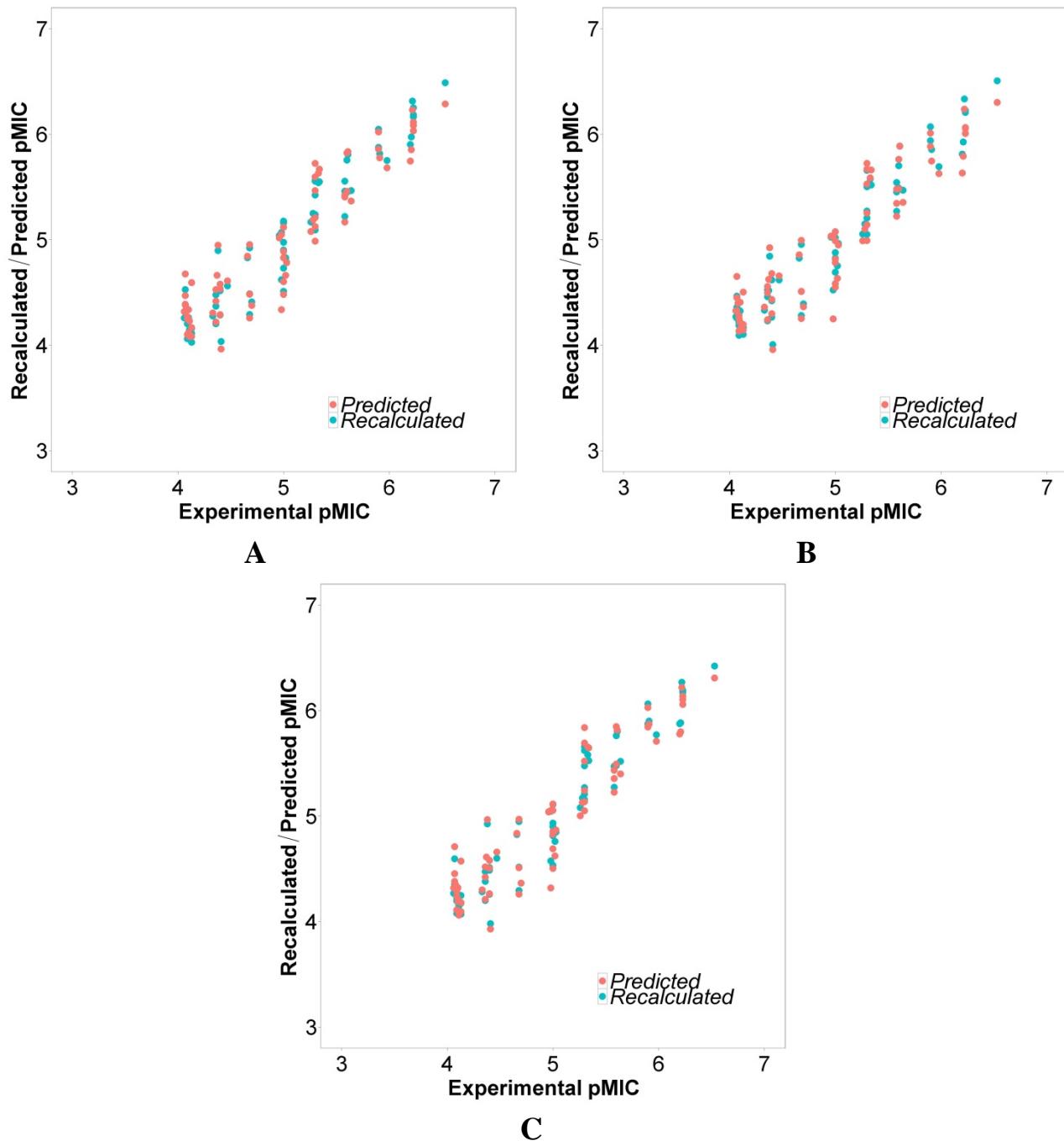
**Table S2.** AutoGrid/R PLS models raw models statistical results; P:AutoGrid Probe,  $r^2$ : conventional square-correlation coefficient;  $q^2_{\text{LOO}}$ : cross-validation correlation coefficient using the leave-one-out method;  $q^2_{\text{K5FCV}}$ : cross-validation correlation coefficient using the  $k$ -fold cross-validation with 5 random groups and 100 iterations.

	model	P	PC1	PC2	PC3	PC4	PC5
$r^2$	1	A	0.56	0.86	0.92	0.94	0.95
$q^2_{\text{LOO}}$	1	A	0.48	0.83	0.86	0.85	0.85
$q^2_{\text{K5FCV}}$	1	A	0.49	0.82	0.85	0.84	0.84
$r^2$	2	C	0.56	0.86	0.92	0.94	0.95
$q^2_{\text{LOO}}$	2	C	0.47	0.83	0.86	0.85	0.85
$q^2_{\text{K5FCV}}$	2	C	0.48	0.82	0.85	0.84	0.84
$r^2$	3	HD	0.46	0.86	0.91	0.93	0.95
$q^2_{\text{LOO}}$	3	HD	0.38	0.82	0.85	0.85	0.85
$q^2_{\text{K5FCV}}$	3	HD	0.39	0.81	0.84	0.84	0.84
$r^2$	4	NA	0.51	0.85	0.91	0.93	0.94
$q^2_{\text{LOO}}$	4	NA	0.43	0.82	0.86	0.86	0.85
$q^2_{\text{K5FCV}}$	4	NA	0.44	0.81	0.85	0.85	0.84
$r^2$	5	N	0.51	0.85	0.91	0.93	0.94
$q^2_{\text{LOO}}$	5	N	0.44	0.82	0.85	0.85	0.85
$q^2_{\text{K5FCV}}$	5	N	0.45	0.81	0.85	0.85	0.84
$r^2$	6	OA	0.51	0.85	0.91	0.93	0.95
$q^2_{\text{LOO}}$	6	OA	0.44	0.82	0.85	0.85	0.84
$q^2_{\text{K5FCV}}$	6	OA	0.45	0.81	0.85	0.84	0.83
$r^2$	7	e	0.45	0.78	0.83	0.88	0.91
$q^2_{\text{LOO}}$	7	e	0.36	0.65	0.73	0.78	0.80
$q^2_{\text{K5FCV}}$	7	e	0.35	0.62	0.72	0.76	0.79
$r^2$	8	d	0.17	0.76	0.88	0.91	0.92
$q^2_{\text{LOO}}$	8	d	0.13	0.71	0.83	0.85	0.85
$q^2_{\text{K5FCV}}$	8	d	0.14	0.71	0.82	0.84	0.83

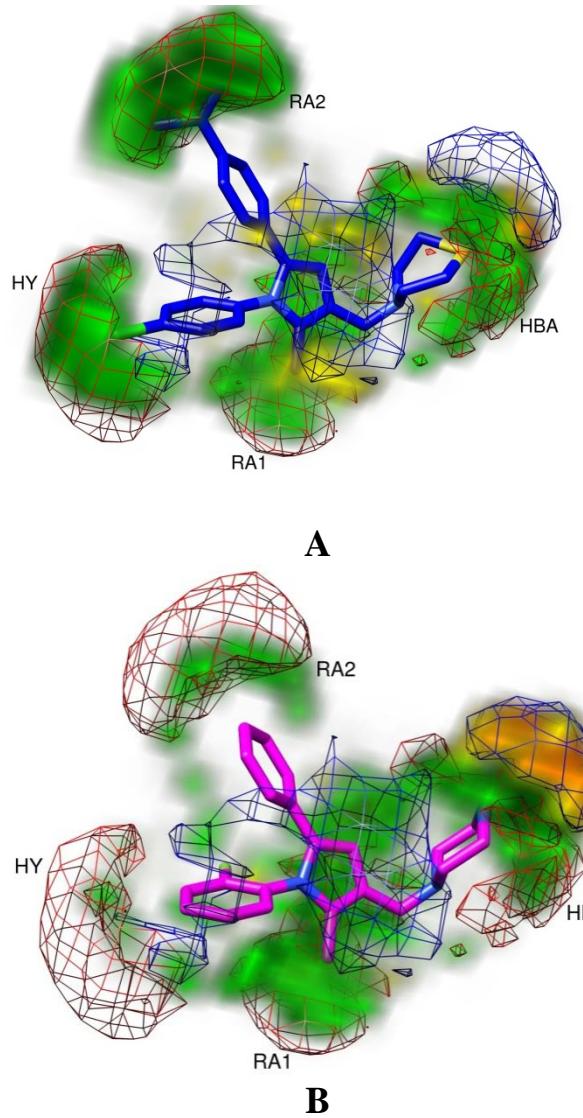
**Table S3.** AutoGrid/R PLS models pretreated models statistical results; P:AutoGrid Probe,  $r^2$ : conventional square-correlation coefficient;  $q^2_{\text{LOO}}$ : cross-validation correlation coefficient using the leave-one-out method;  $q^2_{\text{K5FCV}}$ : cross-validation correlation coefficient using the  $k$ -fold cross-validation with 5 random groups and 100 iterations.

Average $q^2_{\text{K5FCV}}$ value of Starting models:	Average $q^2_{\text{K5FCV}}$ value of Final models:	Average $q^2_{\text{K5FCV}}$ value increment
Raw models	Pretreated models with CAPP parameters	
0.65	0.74	14%

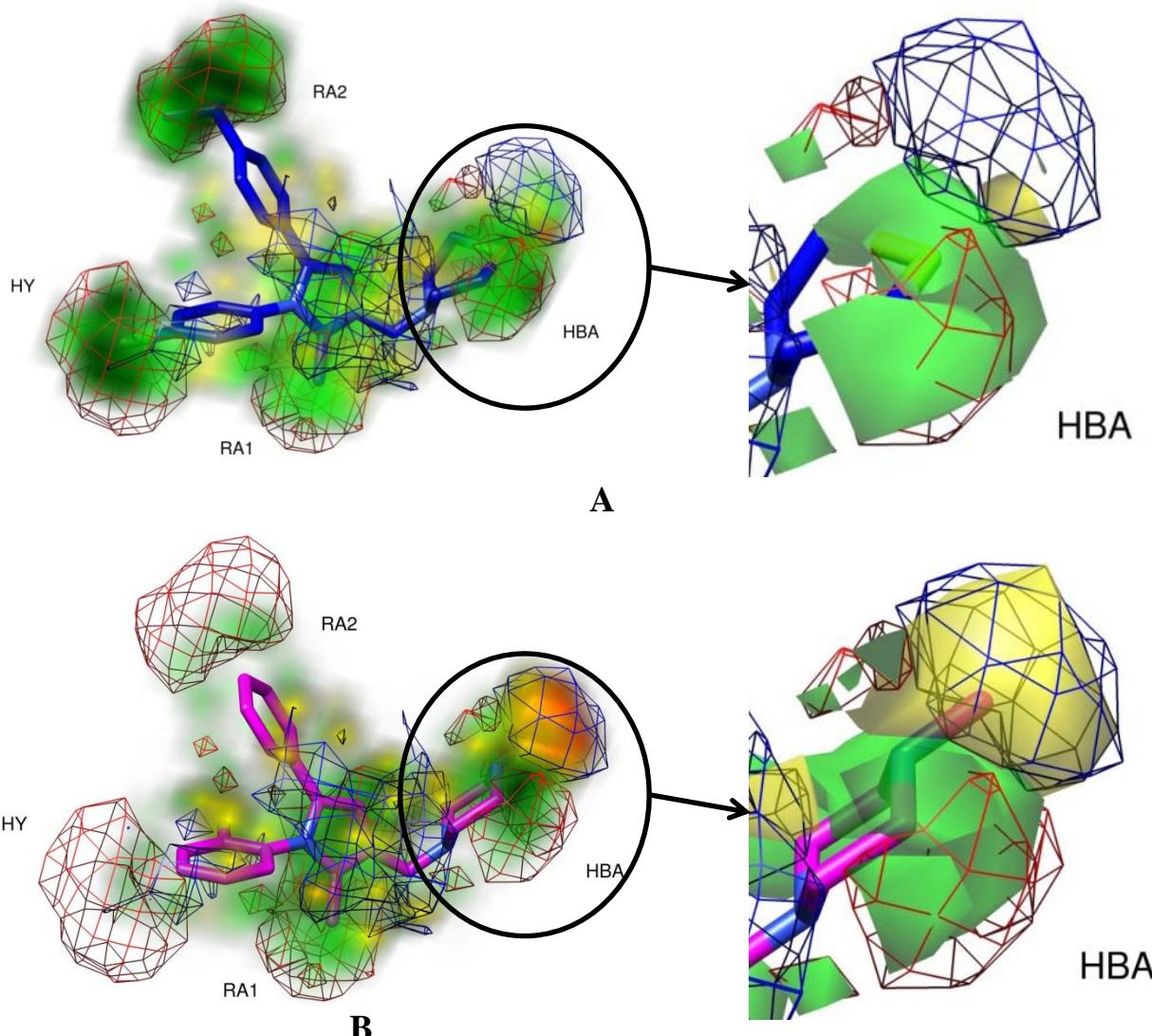
**Table S4.** Average  $q^2_{\text{K5FCV}}$  value increment after CAPP pretreatment procedure.



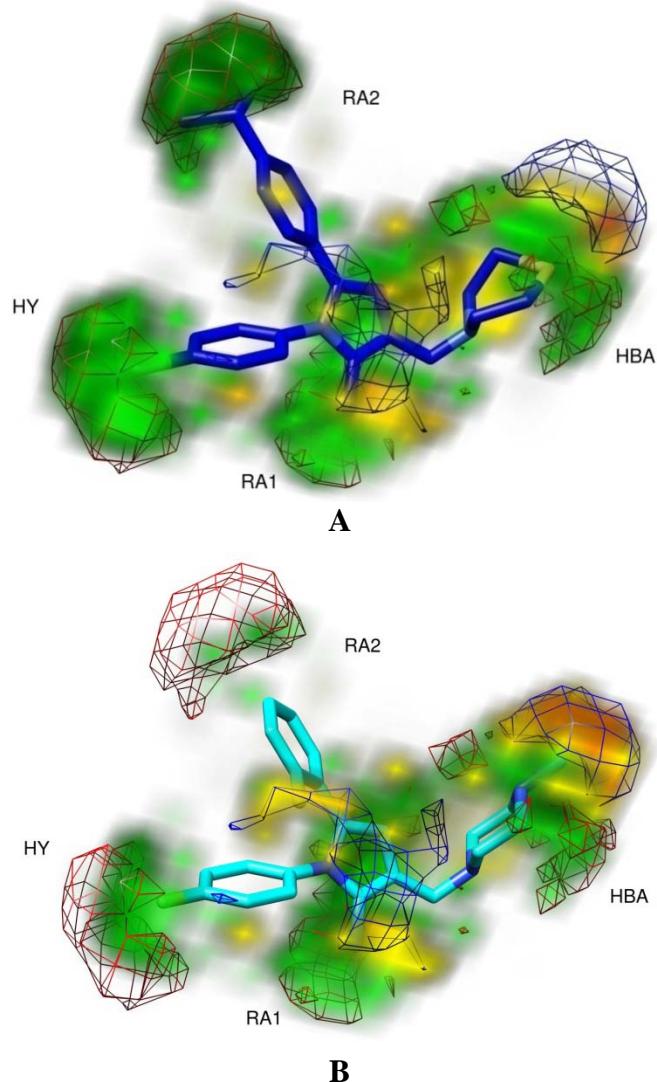
**Figure S1.** Fitting ( $r^2$ ) and Cross-Validation ( $q_2$  K-5-Fold) plots. A: from the A probe model at PC3; B:from HD probe model at PC3; C:from the NA probe model at PC3.



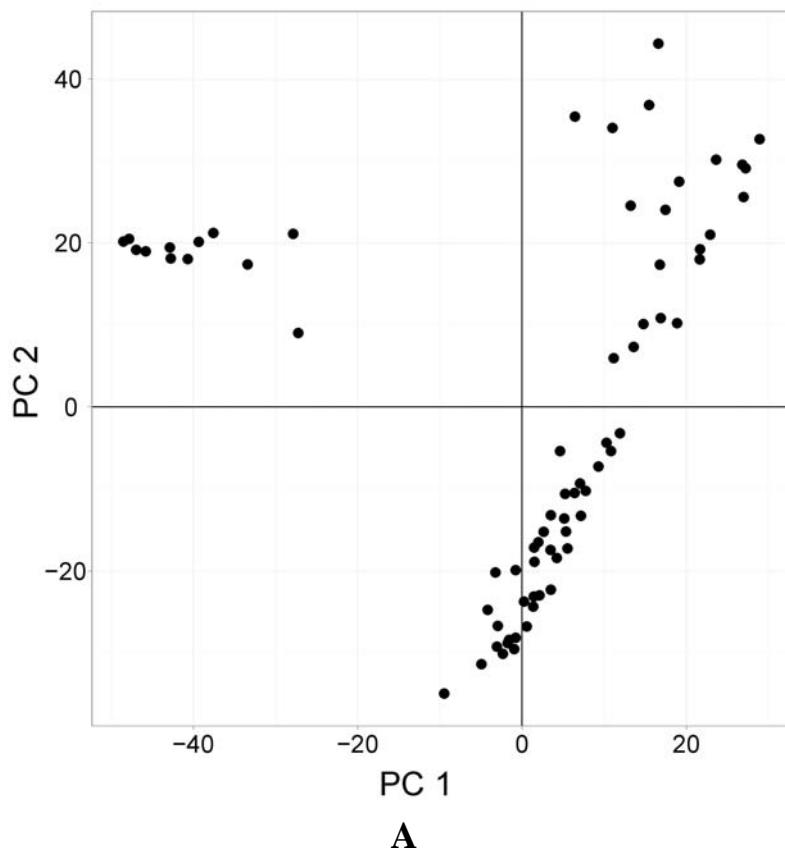
**Figure S2.** Probe A. A: PLS-coefficients (mesh: 80%, positive: red, negative: blue) with activity contribution (solid: 80%, positive: green, negative: yellow) for compound **60** (blue); B: PLS-coefficients (mesh: 80%, positive: red, negative: blue) with activity contribution (solid: 80%, positive: green, negative: yellow) for compound **21**(magenta). Activity contributions are shown in color gradient: for both green and yellow polygons, the darker areas (the most important) are characterized by the highest numerical coefficients, the lighter areas (the less important) are characterized by the lowest numerical coefficients.



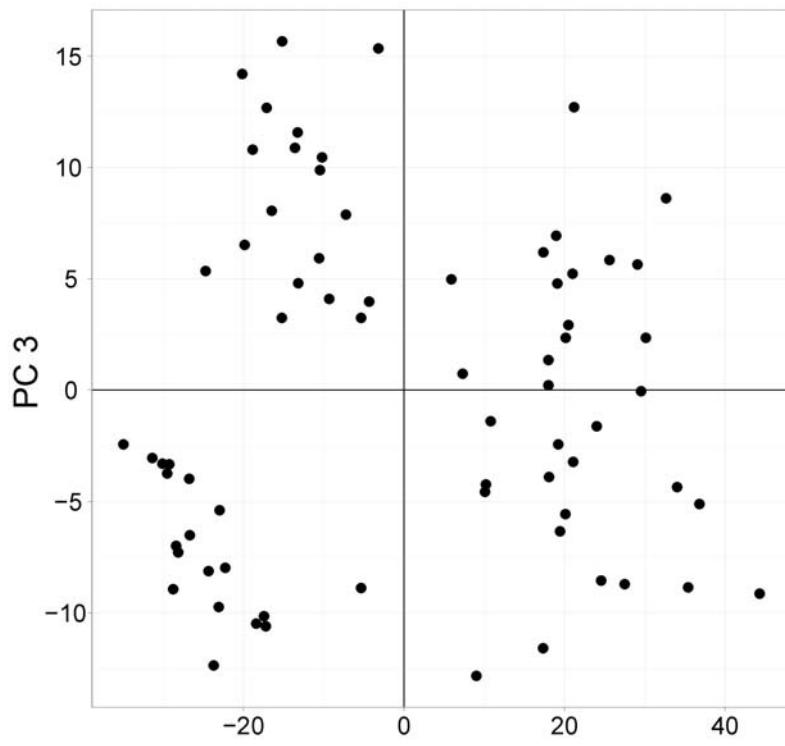
**Figure S3.** Probe HD. A: PLS-coefficients (mesh: 85%, positive: red, negative: blue) with activity contribution (solid: 90%, positive: green, negative: yellow) for compound **60** (blue); B: PLS-coefficients (mesh: 85%, positive: red, negative: blue) with activity contribution (solid: 90%, positive: green, negative: yellow) for compound **21**(magenta).Activity contributions in pictures on the left side are shown in color gradient: for both green and yellow polygons, the darker areas (the most important) are characterized by the highest numerical coefficients, the lighter areas (the less important) are characterized by the lowest numerical coefficients.



**Figure S4.** Probe NA. A: PLS-coefficients (mesh: 75%, positive: red, negative: blue) with activity contribution (solid: 85%, positive: green, negative: yellow) for compound **60** (blue); B: PLS-coefficients (mesh: 75%, positive: red, negative: blue) with activity contribution (solid: 85%, positive: green, negative: yellow) for compound **34**(cyan).Activity contributions are shown in color gradient: for both green and yellow polygons, the darker areas (the most important) are characterized by the highest numerical coefficients, the lighter areas (the less important) are characterized by the lowest numerical coefficients.

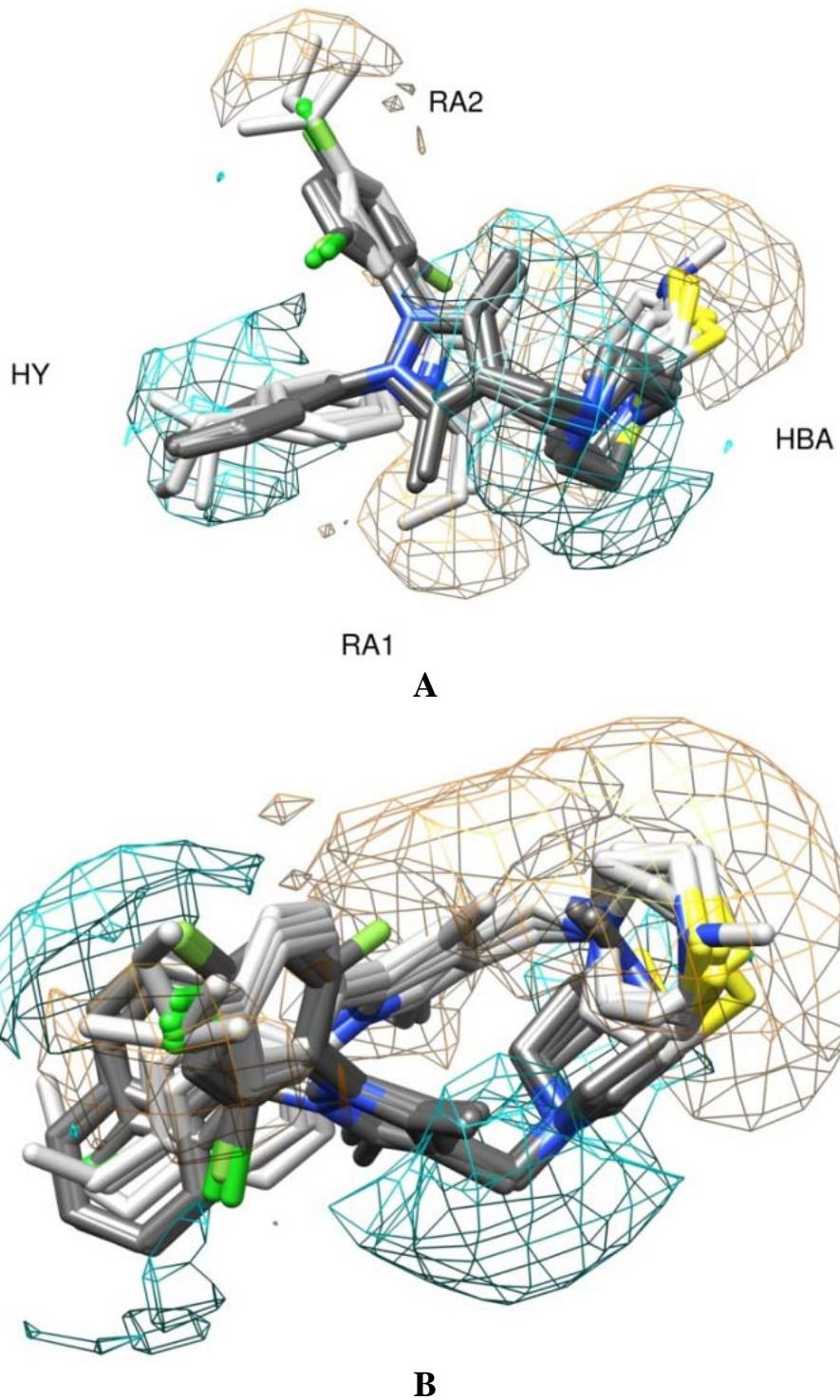


**A**

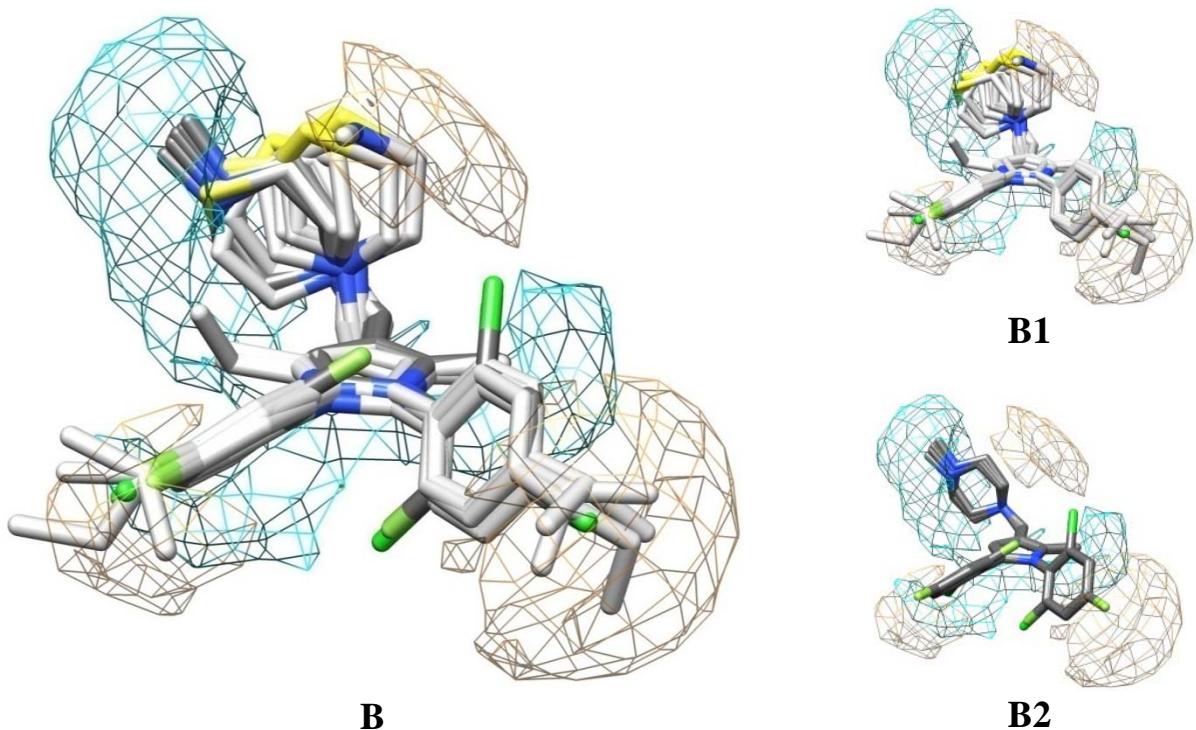
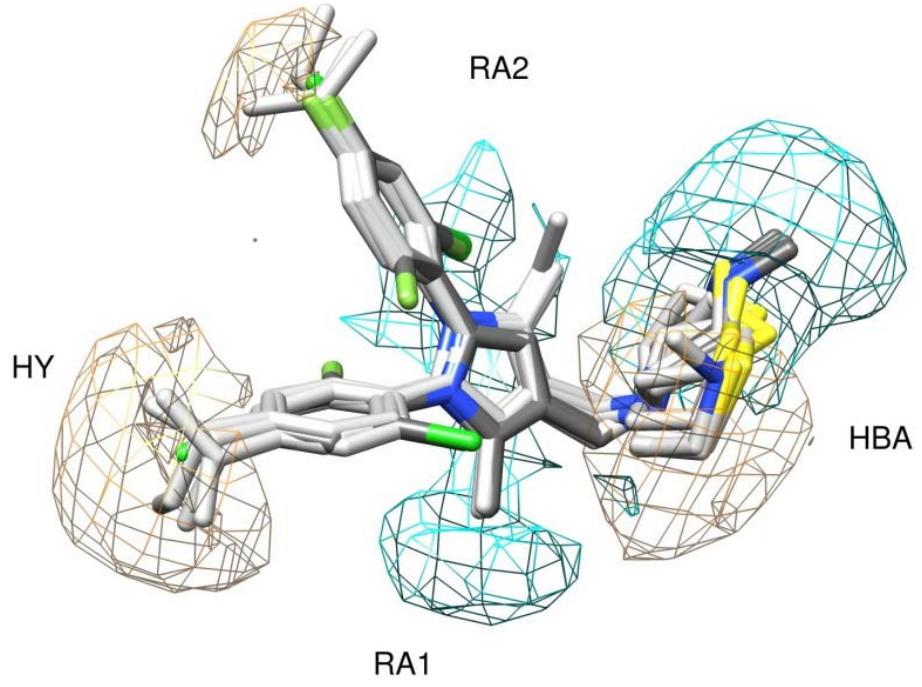


**B**

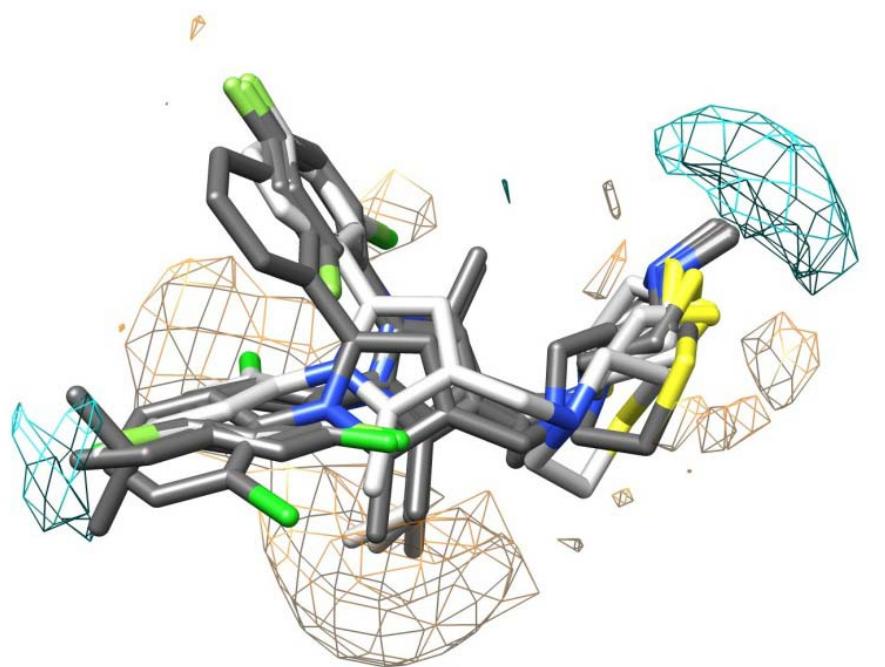
**Figure S5.** Probe A. A: PC1vsPC2 scores plot derived from A probe analysis; B: PC2vsPC3 scores plot derived from A probe analysis.



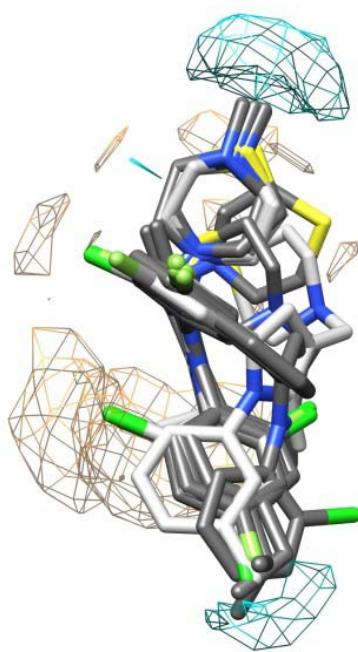
**Figure S6.** Probe A. PLS-loadings contour maps from the A probe analysis at PC1 (contour levels: 60%; positive: orange, negative: cyan). The ten most important molecules for each cluster are plotted and color coded according to the cluster membership (molecules in the negative field cluster: dark grey, molecules in the positive field cluster: light grey). A:side view; B: top view.



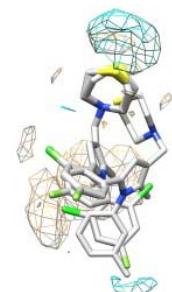
**Figure S7.** Probe A. PLS-loadings contour maps from the A probe analysis at PC2 (contour levels: 60%; positive: orange, negative: cyan). The ten most important molecules for each cluster are plotted and color coded (compounds in the positive loading field in light grey; compounds in the negative loading field in dark grey). A: side view; B: front view; B1: front view of only positive clustered molecules; B2: front view of only negative clustered molecules.



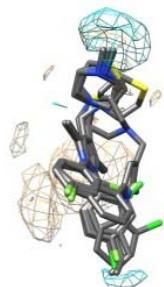
A



B



B1

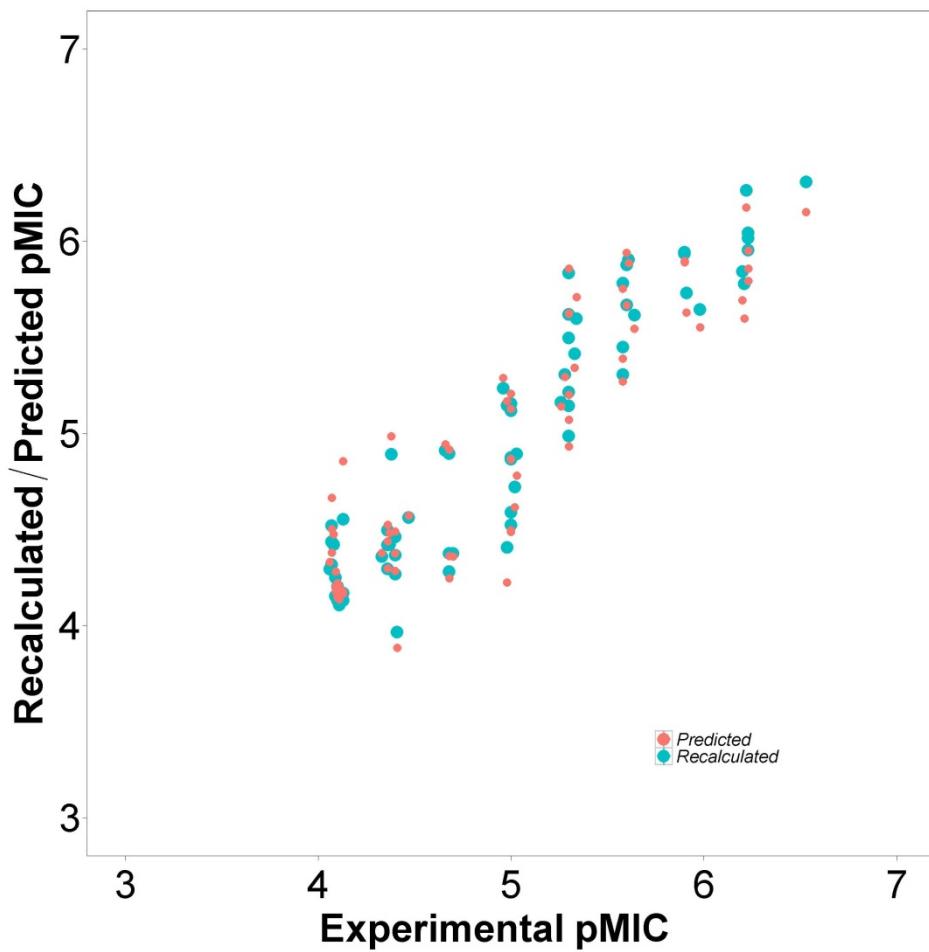


B2

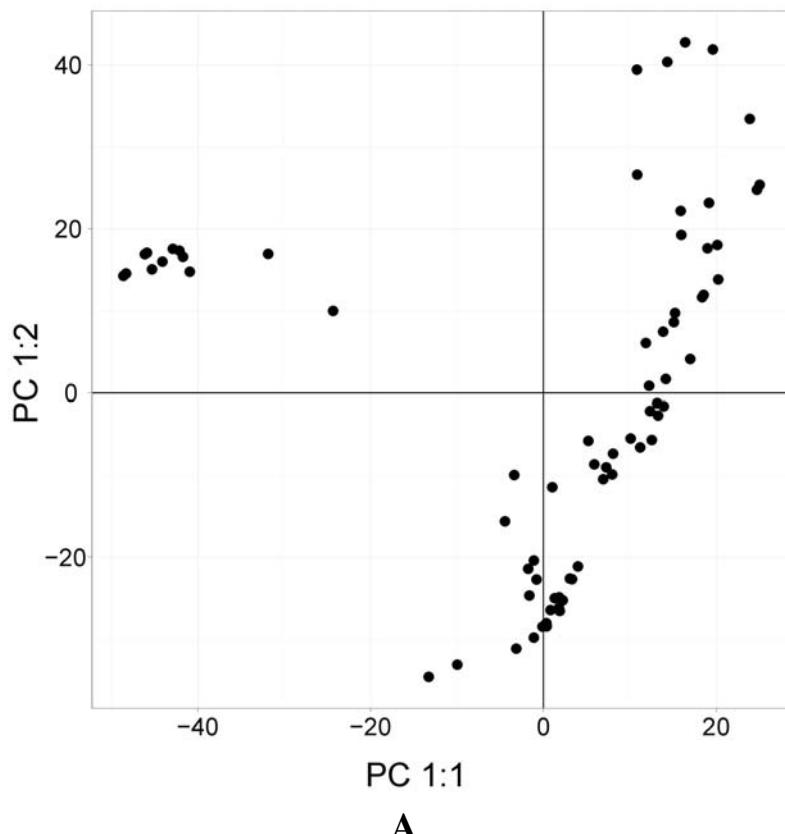
**FigureS8.** Probe A. PLS-loadings contour maps from the A probe analysis at PC3 (contour levels: 60%; positive: orange, negative: cyan). The ten most important molecules for each cluster are plotted and color coded (compounds in the positive loading field in light grey; compounds in the negative loading field in dark grey). A: side view; B: top view; B1: top view of only positive clustered molecules; B2: top view of only negative clustered molecules.

Probe	Description	MPGRS Colour
A	Aromatic Carbon	Gray
C	Aliphatic ( $sp^3$ ) Carbon	Dark Gray
HD	Hydrogen bonded to heteroatom	Green
NA	Hydrogen-bond-accepting amine nitrogen	Cyan
N	Amide nitrogen	Blue
OA	Hydrogen-bond-accepting oxygen	Red
e	Electrostatic	Orange
d	Desolvation	Yellow

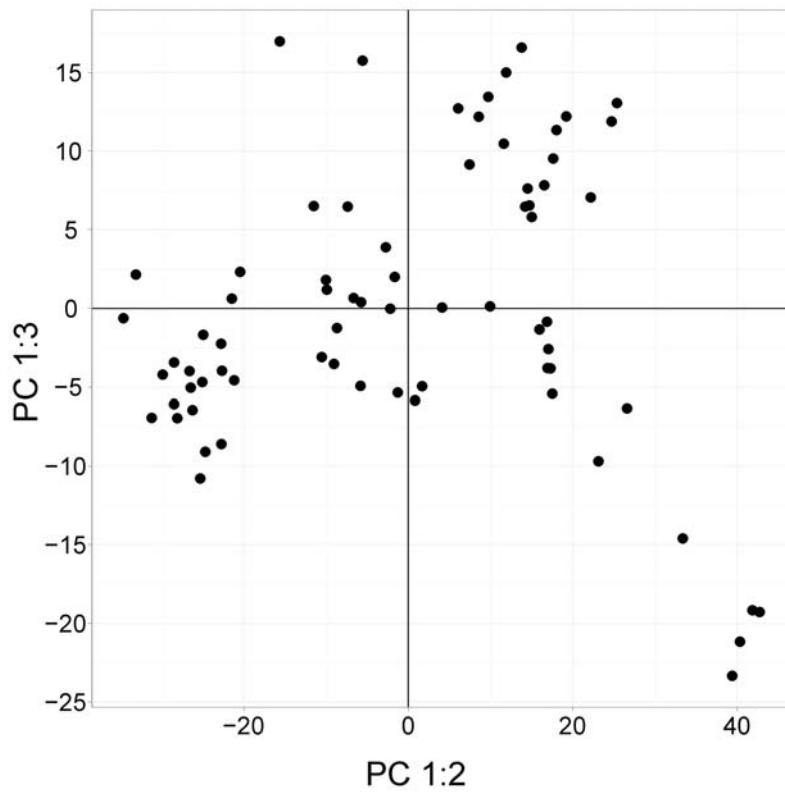
**Table S5.** List of the AutoGrid probes employed for MIF calculation and MPGRS Subregion color coding.



**Figure S9.** MPGRS. Fitting ( $r^2$ ) and Cross-Validation ( $q_2$  K-5-Fold) plot: from the multi probe(MP) model at PC<sub>1:3</sub>.

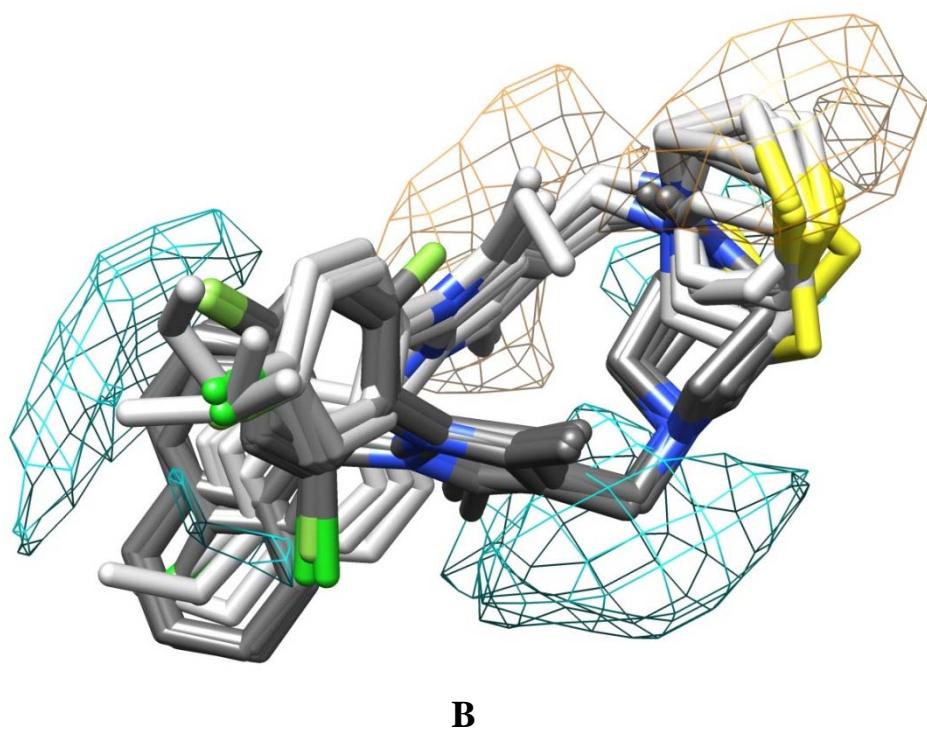
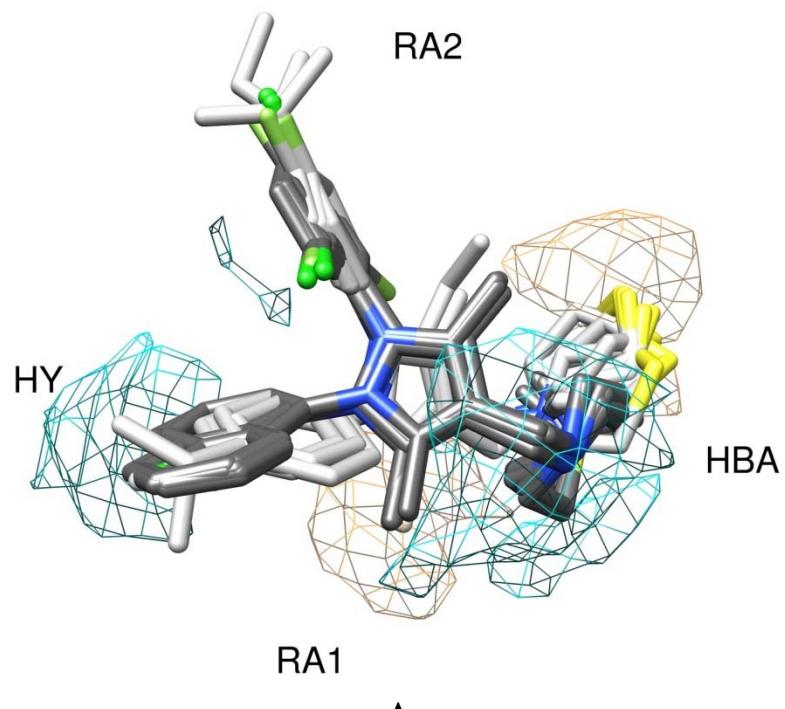


**A**

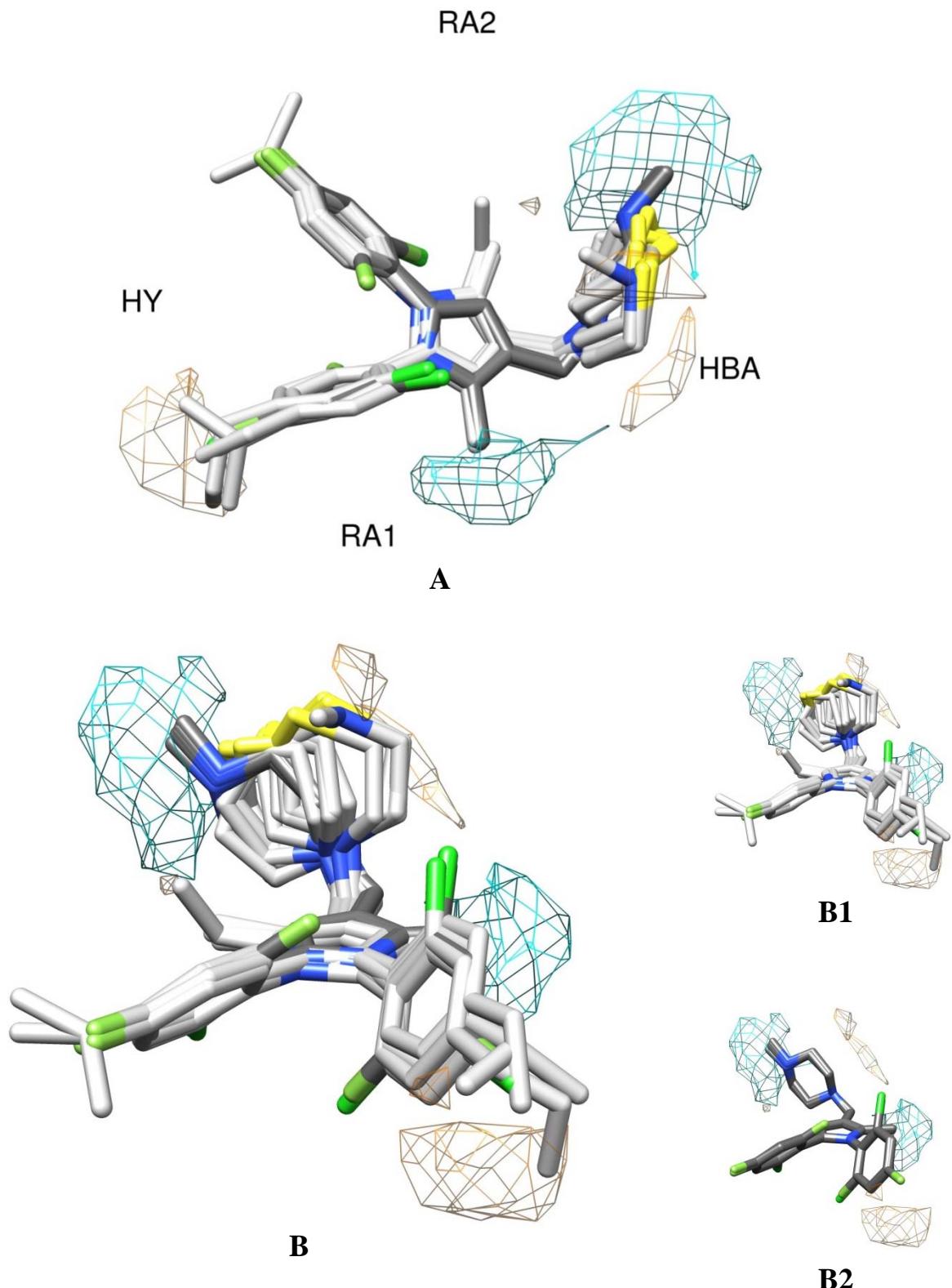


**B**

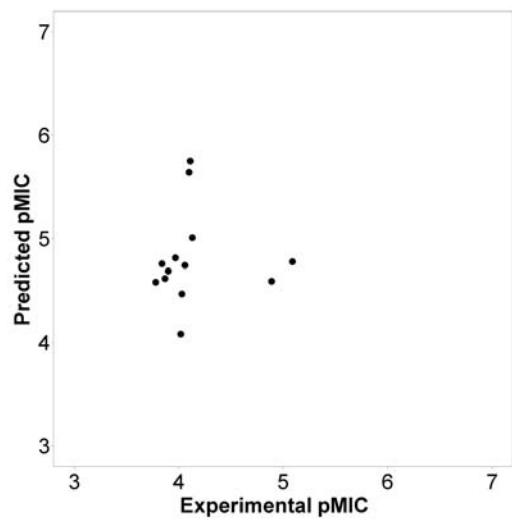
**Figure S10.MPGRS.A:** PC<sub>1:1</sub>vsPC<sub>1:2</sub> scores plot; **B:** PC<sub>1:2</sub>vsPC<sub>1:3</sub> scores plot.



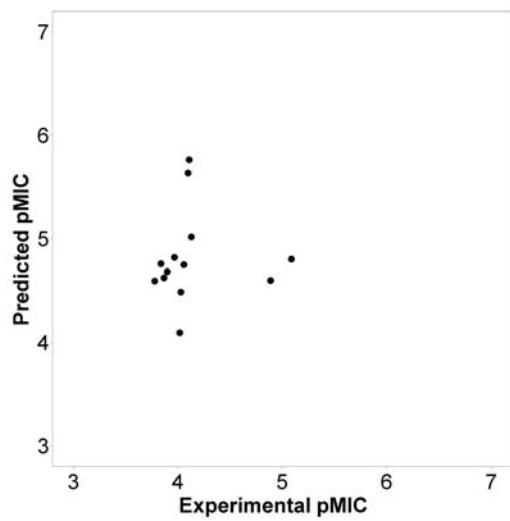
**Figure S11.** MPG.RS.PLS-loadings contour maps at PC<sub>1:1</sub> (contour levels: 60%; positive: orange, negative: cyan). The ten most important molecules for each cluster are plotted and color coded (compounds in the positive loading field in light grey; compounds in the negative loading field in dark grey). A: side view; B: top view.



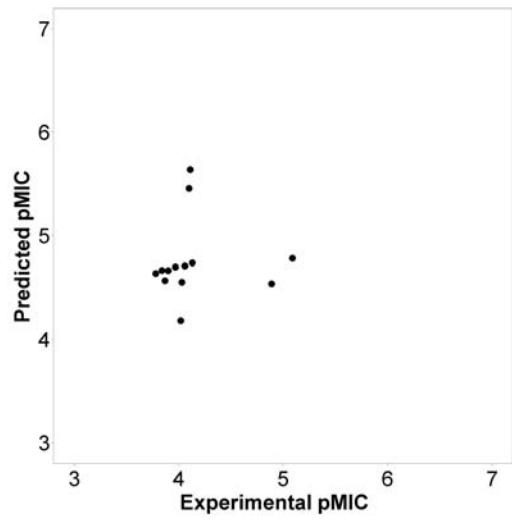
**Figure S12.MPGRS.** PLS-loadings contour maps at  $PC_{1:2}$  (contour levels: 60%; positive: orange, negative: cyan). The ten most important molecules for each cluster are plotted and color coded (compounds in the positive loading field in light grey; compounds in the negative loading field in dark grey). A: side view; B: front view; B1: front view of only positive clustered molecules; B2: front view of only negative clustered molecules.



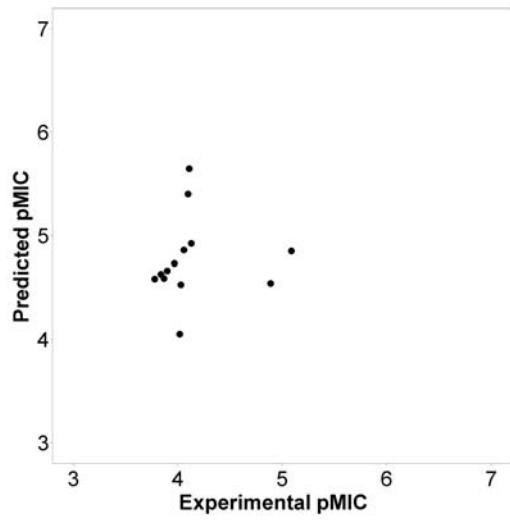
A



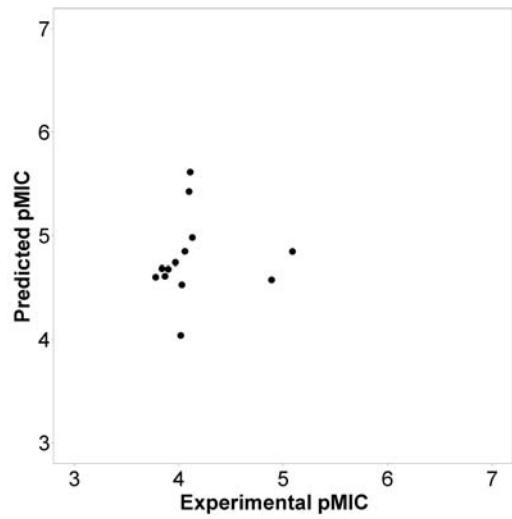
B



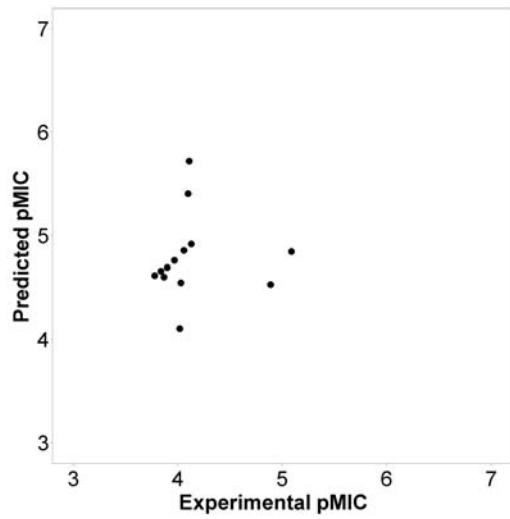
C



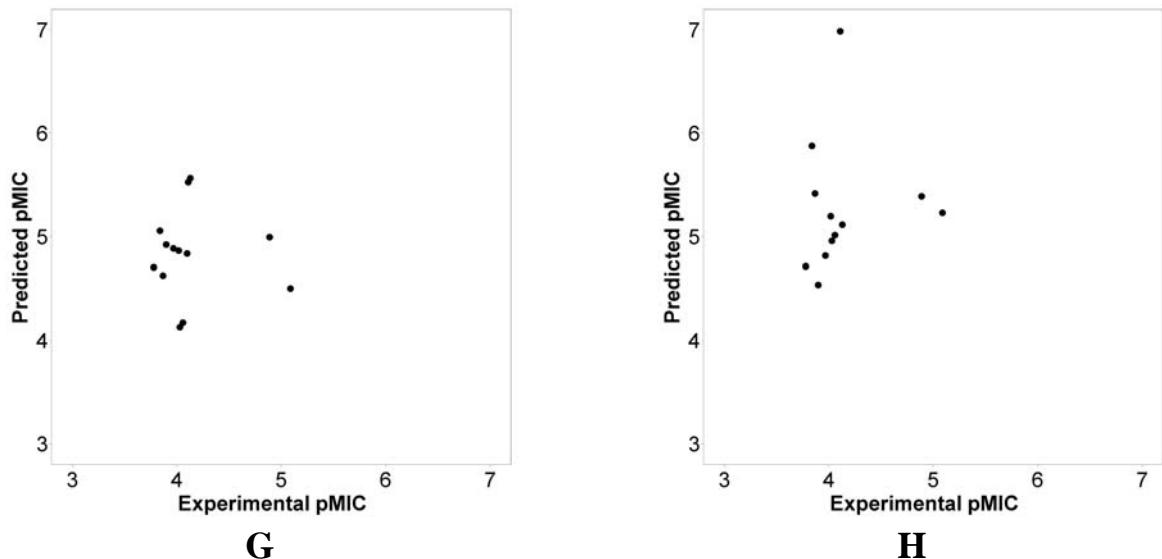
D



E



F

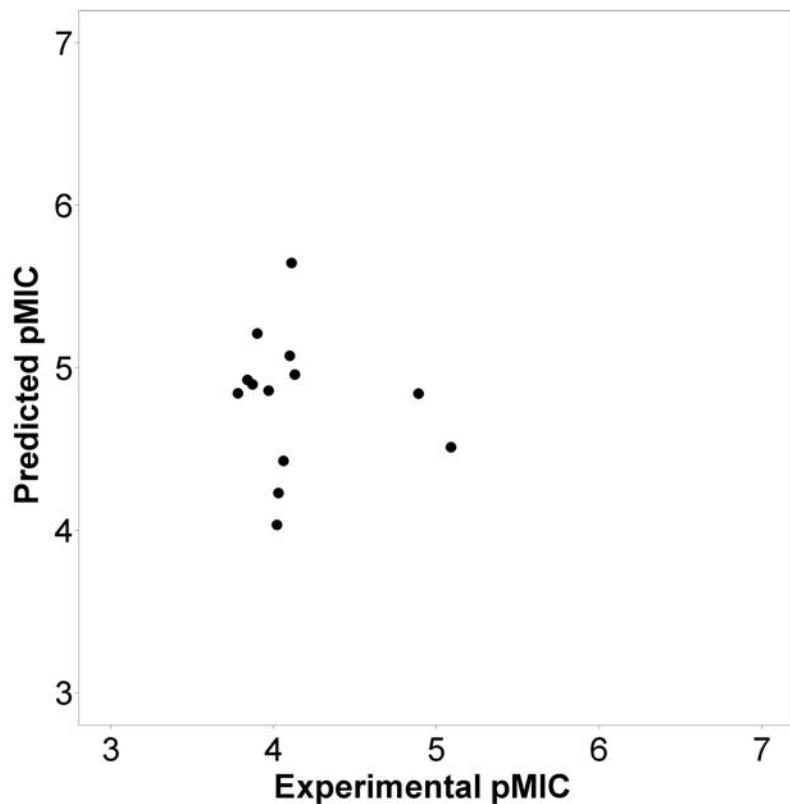


**Figure S13.** Experimental vs Predicted Test Set plots. A: from the A probe model at PC3; B: from the C probe model at PC3; C: from the HD probe model at PC3; D: from the NA probe model at PC3; E: from the N probe model at PC3; F: from the OA probe model at PC3; G: from the e probe model at PC4; H: from the d probe model at PC4.

Compd	Exp.	A	C	HD	NA	N	OA	e	d	AAEP <sup>a</sup>
1a	3.84	4.76	4.76	4.67	4.63	4.68	4.66	5.06	5.88	1.05
1b	3.87	4.62	4.63	4.57	4.59	4.61	4.60	4.63	5.42	0.84
1c	3.9	4.69	4.68	4.66	4.66	4.68	4.69	4.92	4.54	0.79
1d	4.89	4.59	4.60	4.54	4.54	4.58	4.53	4.99	5.39	0.32
1e	3.97	4.81	4.82	4.70	4.73	4.74	4.76	4.89	4.82	0.81
2a	4.03	4.47	4.49	4.55	4.53	4.53	4.55	4.13	4.96	0.50
2b	4.06	4.74	4.75	4.71	4.86	4.85	4.86	4.17	5.01	0.68
2c	4.1	5.64	5.63	5.46	5.40	5.43	5.40	4.84	7.02	1.50
2d	5.09	4.78	4.80	4.78	4.85	4.85	4.85	4.50	5.23	0.30
2e	4.11	5.75	5.76	5.64	5.65	5.61	5.72	5.53	6.99	1.72
2f	4.02	4.08	4.09	4.18	4.05	4.04	4.10	4.86	5.20	0.31
3h	3.78	4.58	4.59	4.64	4.58	4.60	4.62	4.70	4.71	0.85
3i	4.13	5.01	5.02	4.74	4.92	4.98	4.92	5.56	5.11	0.90

**Table S6.** Test Set prediction values for each mono probe model at the selected principal components (PCs, see Table 4);

$$^a\text{AAEP (Average Absolute Error of Prediction)} = \frac{\sum_{i=1}^{np} (|Y_{Exp} - Y_{Pred_i}|)}{np}$$



**Figure S14.** MPGRS. Experimental *vs* Predicted Test Set plots at PC<sub>1:3</sub>.

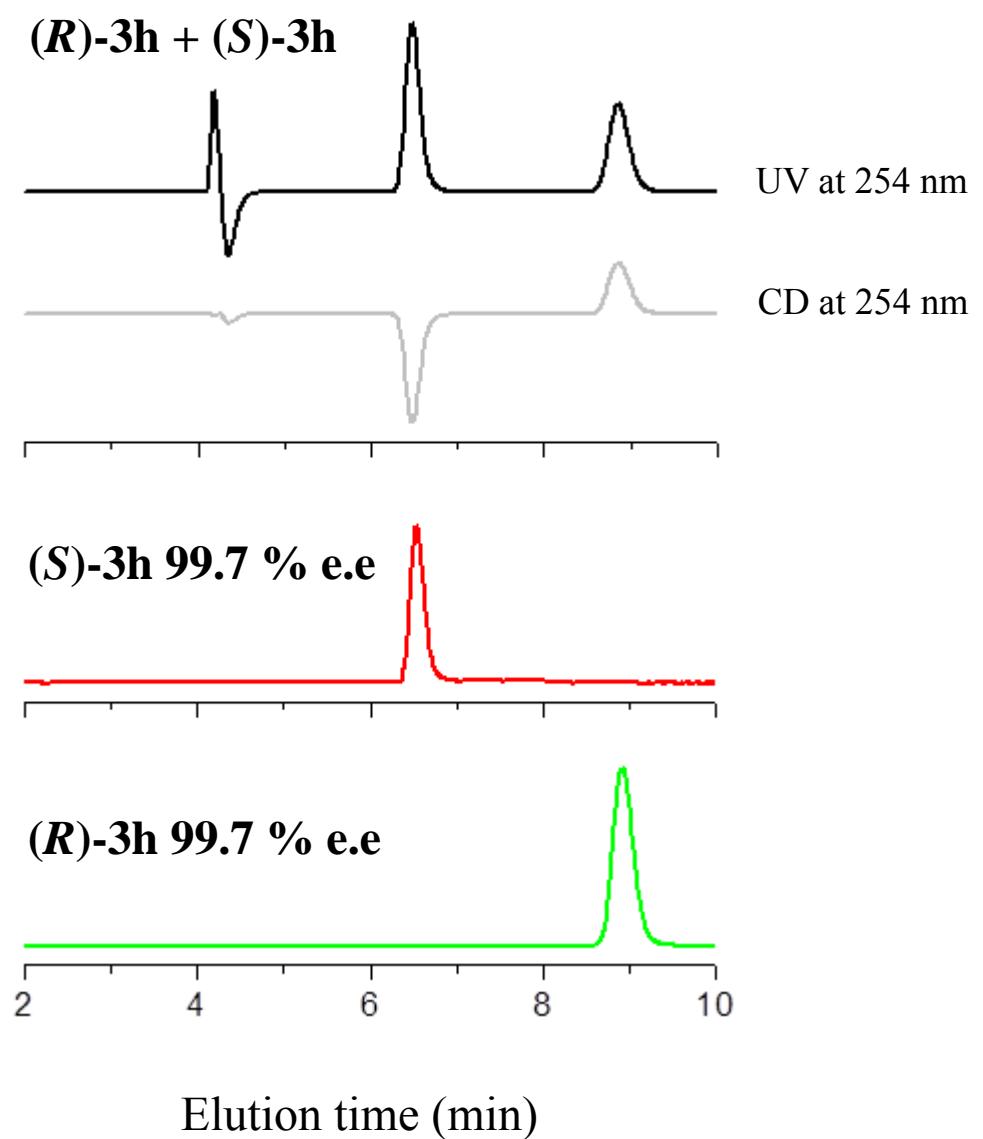
Compd	Exp.	MP	AEP <sup>a</sup>
1a	3.84	4.93	1.09
1b	3.87	4.90	1.03
1c	3.9	5.21	1.31
1d	4.89	4.84	0.05
1e	3.97	4.86	0.89
2a	4.03	4.23	0.20
2b	4.06	4.43	0.37
2c	4.1	5.07	0.97
2d	5.09	4.51	0.58
2e	4.11	5.65	1.54
2f	4.02	4.03	0.01
3h	3.78	4.84	1.06
3i	4.13	4.96	0.83

**Table S7.** MPGRS. Test Set prediction values for the Multi Probe model at the selected first and second principal components (PC<sub>FL:SL</sub>, see Table 5). MP: Multi Probe prediction results;

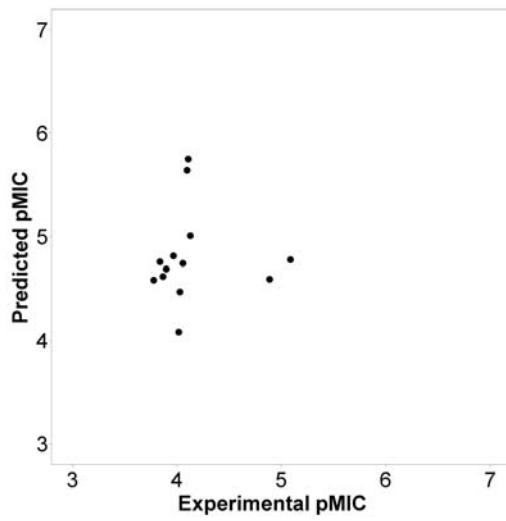
<sup>a</sup>AEP: Absolute Error of Prediction =  $|Y_{Exp} - Y_{Pred_i}|$

XYZ coordinates			Spacing (Å)	Number of grid points in xyz	Number of total grid Points
x	y	Z			
Min.Coord.	-9.828	-9.021	-10.481	1.000	23, 19, 21
Max.Coord.	12.172	8.979	9.519		

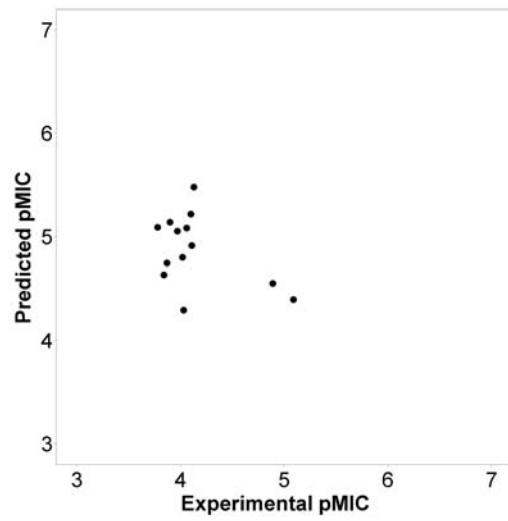
**Table S8.** AutoGrid settings for the MIF calculation. Spacing: spacing between the points in Angstrom. Number of grid points in xyz: number of grid points for each dimension of the used grid.



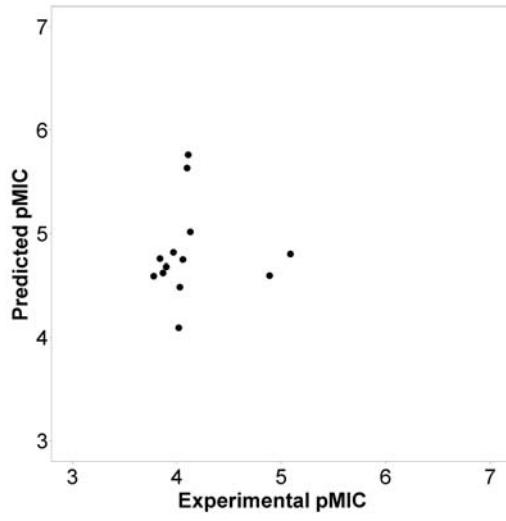
**Figure S15.** Chiral HPLC analysis of **3h**; Column: Chiraldak IC 250 mm x 4.6 mm I.D. Eluent: *n*-hexane-2-propanol 75/25 (v/v), Flow-rate: 1.0 mL/min, Temperature: 25°C.



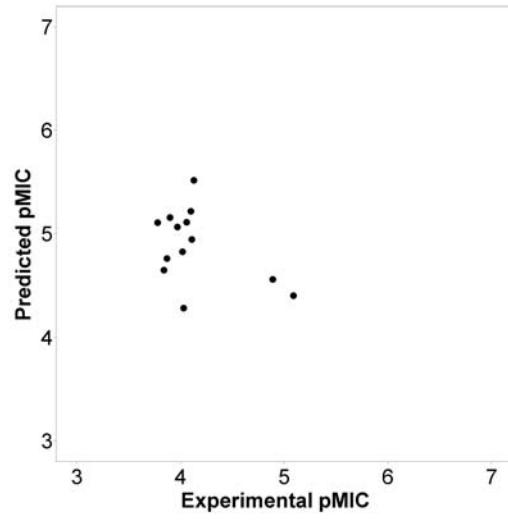
**As**



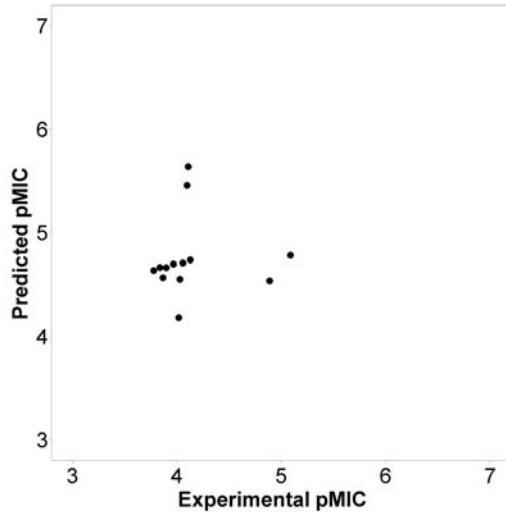
**Ap**



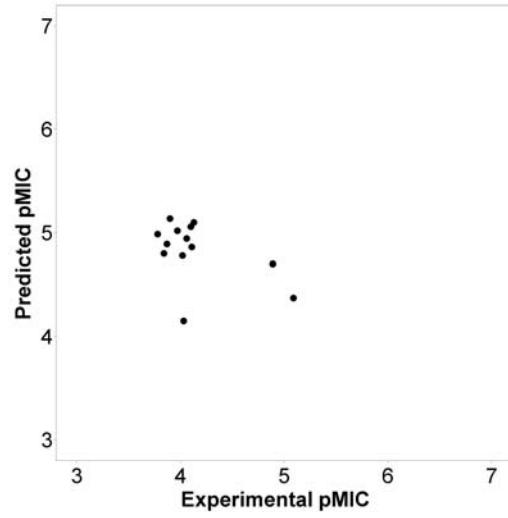
**Bs**



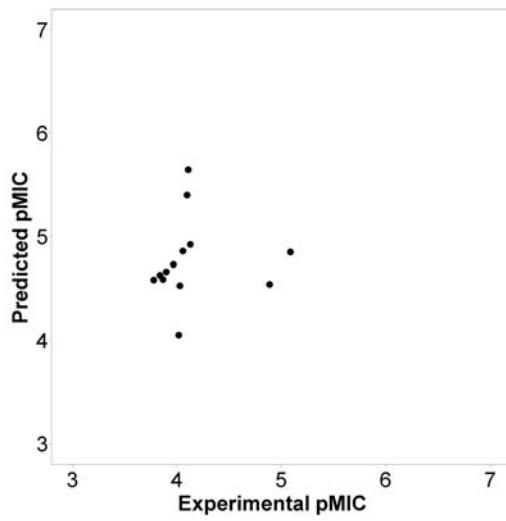
**Bp**



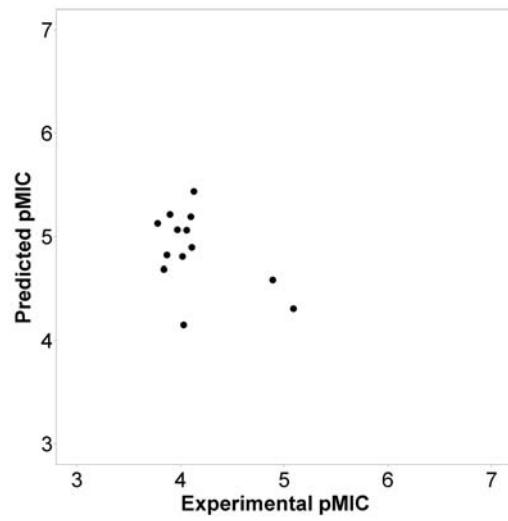
**Cs**



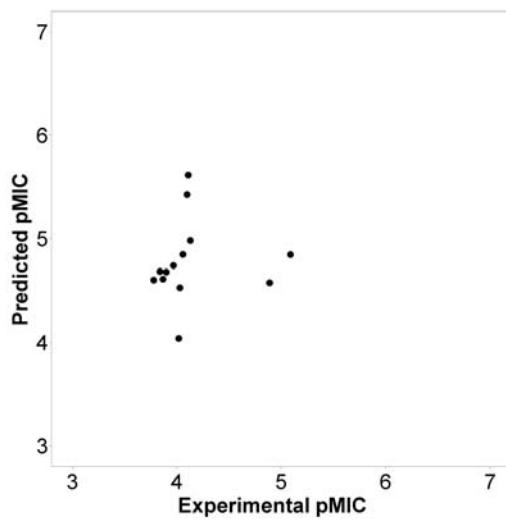
**Cp**



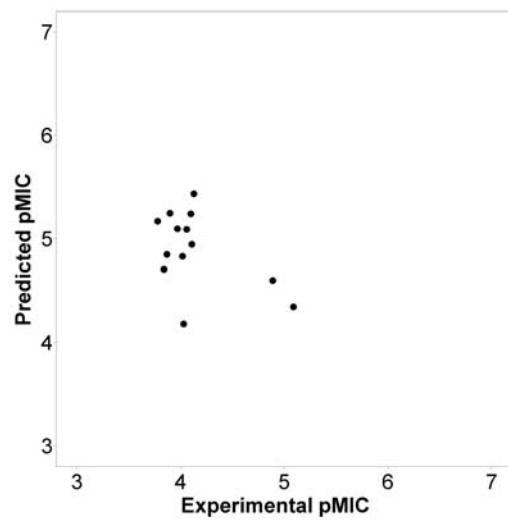
Ds



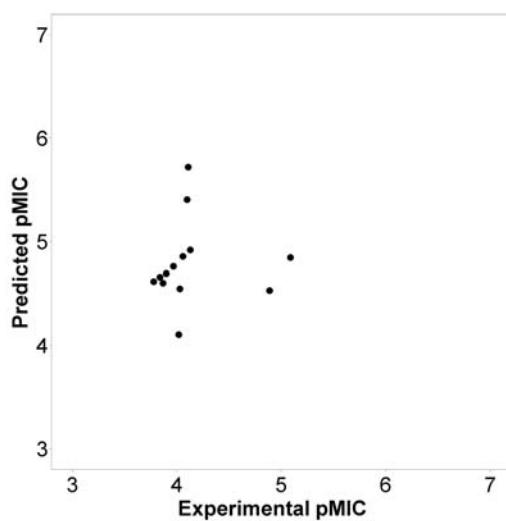
Dp



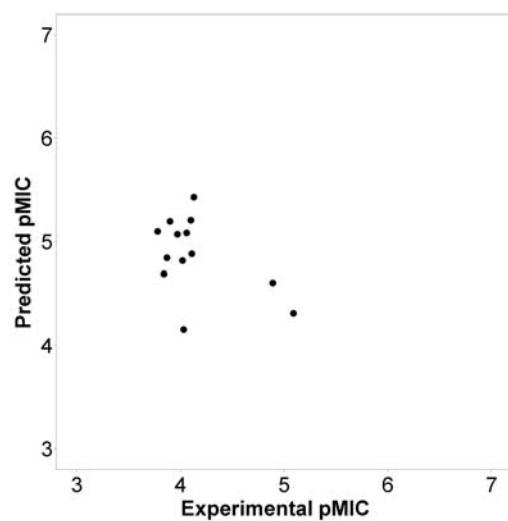
Es



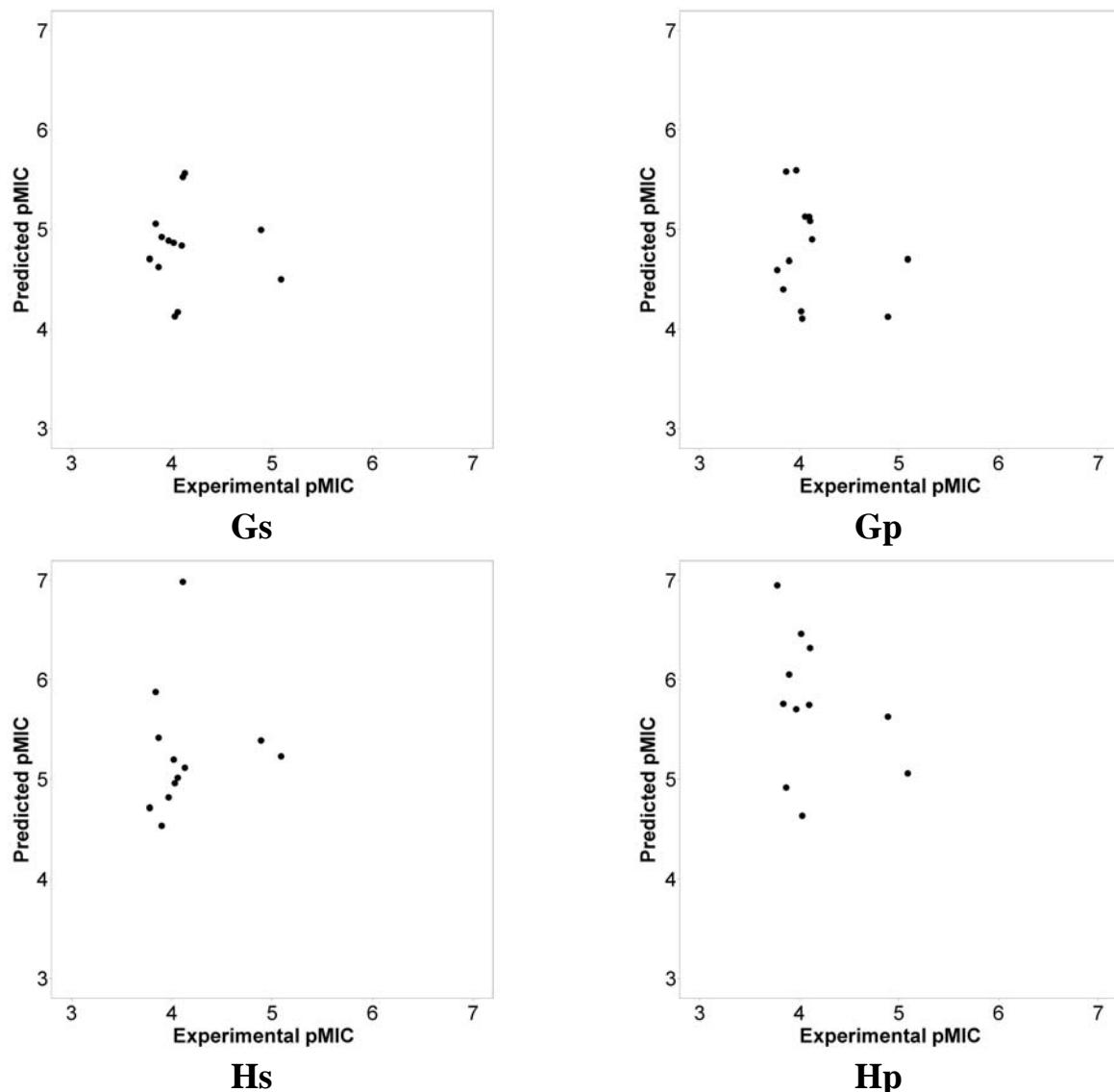
Ep



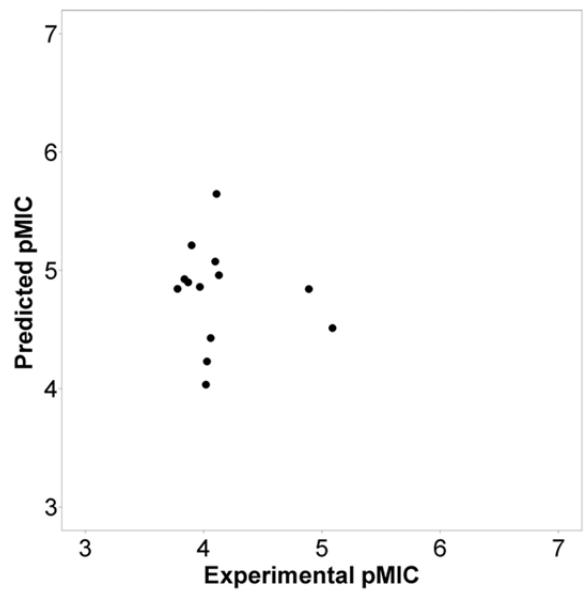
Fs



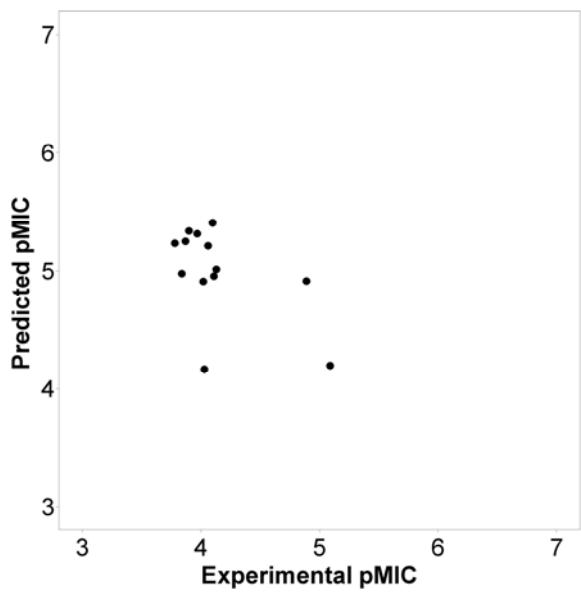
Fp



**Figure S16.** Experimental vs Predicted Test Set plots: Surfflex and pharmacophore alignment results. Left column, Surfflex derived results: As: from the A probe model at PC3; Bs: from the C probe model at PC3; Cs: from the HD probe model at PC3; Ds: from the NA probe model at PC3; Es: from the N probe model at PC3; Fs: from the OA probe model at PC3; Gs: from the e probe model at PC4; Hs: from the d probe model at PC4. Right column, pharmacophore derived results: Ap: from the A probe model at PC3; Bp: from the C probe model at PC3; Cp: from the HD probe model at PC3; Dp: from the NA probe model at PC3; Ep: from the N probe model at PC3; Fp: from the OA probe model at PC3; Gp: from the e probe model at PC4; Hp: from the d probe model at PC4.



**As**



**Ap**

**Figure S17.** MPGRS.Experimental vs Predicted Test Set plots at PC<sub>1:3</sub>. Surfflex and pharmacophore alignment results. Left column: Surfflex derived results (As). Right column: pharmacophore derived results (Ap).

P	SDEP <sub>S</sub>	SDEP <sub>P</sub>	AVERAGE SDEP <sub>S</sub>	AVERAGE SDEP <sub>P</sub>
A	0.88	0.95		
C	0.88	0.97		
HD	0.81	0.89		
NA	0.82	0.97		
N	0.83	0.99	0.93	1.11
OA	0.84	0.97		
e	0.90	0.95		
d	1.51	2.21		

**Table S9.** SDEP test set predictions values for each mono probe model at the selected principal components (PCs, see Table 4), and the relative average values, obtained from the Surflex (SDEP<sub>S</sub>) and pharmacophoric alignment (SDEP<sub>P</sub>). P: AutoGrid Probe.

P	SDEP <sub>S</sub>	SDEP <sub>P</sub>
AutoGrid MP	0.89	1.09

**Table S10.** MPGRS. SDEP test set prediction values for multi probe model (MP) at the selected first and second principal components (PC<sub>FL:SL</sub>, see Table 5) obtained from the Surflex (SDEP<sub>S</sub>) and pharmacophoric alignment (SDEP<sub>P</sub>). P:AutoGrid Multi-Probe.