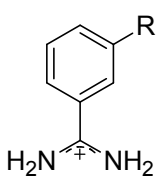
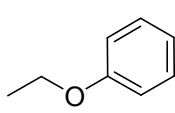
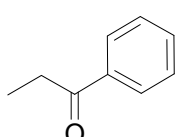
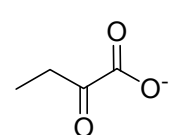
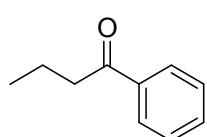


Supporting Information

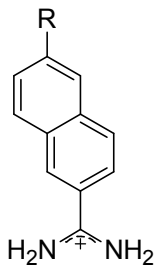
Table 1: Thrombin inhibitors used as training and test set compounds

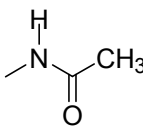
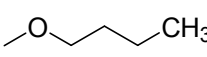
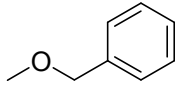
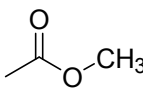
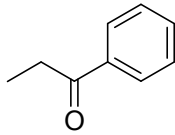
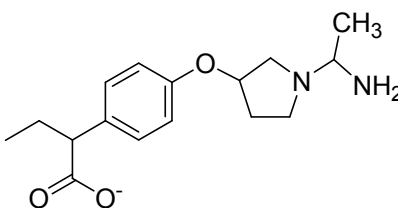
Benzamidine Derivatives



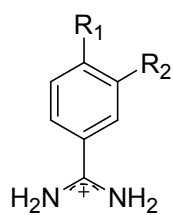
No.	P / T / N ^{a)}	R	<i>pK_i</i>
1	P	H	3.658
2	T	NH ₂	3.000
3	T		4.180
4	T		4.161
5	T		3.102
6	T		3.538

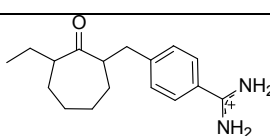
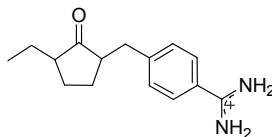
Naphthamidine Derivatives

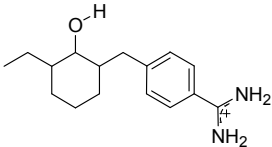
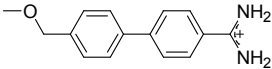
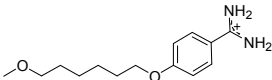
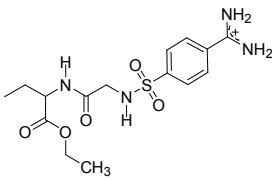
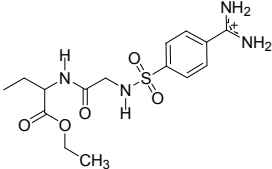
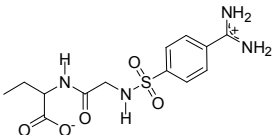
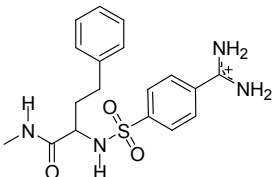


No.	P / T	R	pK_i
7	T	H	4.071
8	T	NH ₂	4.180
9	T		3.921
10	T		4.921
11	P		5.469
12	T		4.201
13	T		4.745
14	T		4.155

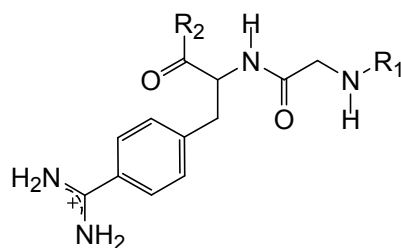
Bisbenzimidine Derivatives

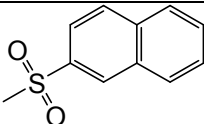
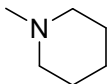


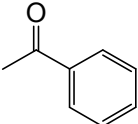
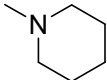
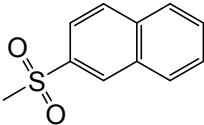
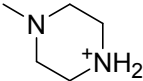
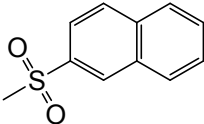
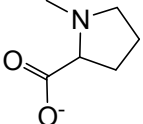
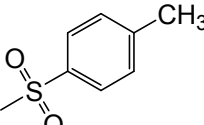
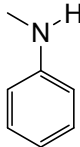
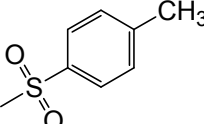
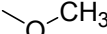
No.	P / T	R ₁	R ₂	pK_i
15	P		H	6.495
16	T	H		5.276

17	T		H	5.569
18	T		H	3.854
19	T		H	5.319
20	T	H		5.495
21	T		H	3.000
22	P		H	4.222
23	T		H	3.886

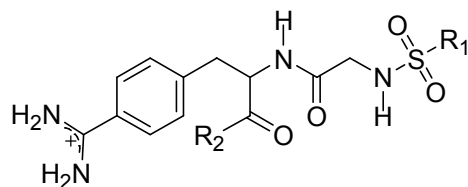
4NAPAP Derivatives

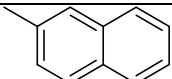
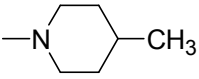
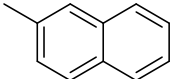
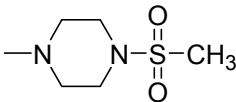
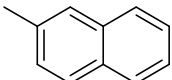
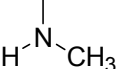
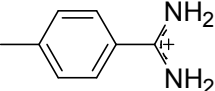

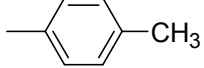
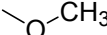


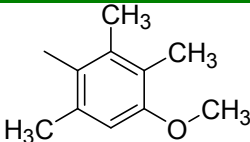
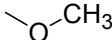
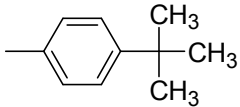
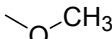
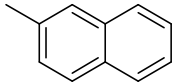
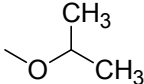
No.	P / T	R ₁	R ₂	<i>pK_i</i>
24	P			8.222

25	T			5.699
26	T			6.260
27	T			6.292
28	T			4.824
29	T			6.131

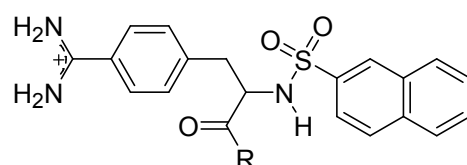
800 Derivatives

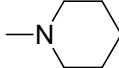
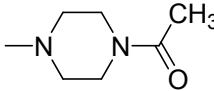

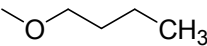


No.	P / T	R ₁	R ₂	<i>pK_i</i>
30	T			5.444
31	T			5.051
32	T			5.018
33	T			4.432
34	P			5.409

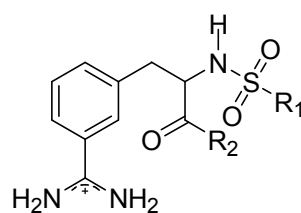
35	P			6.377
36	T			5.854
37	T			5.921

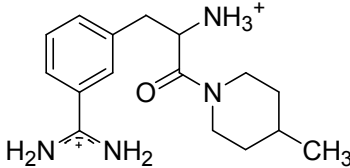
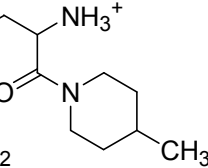
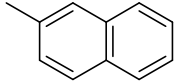
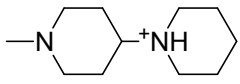
4TAPAP Derivatives

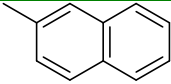
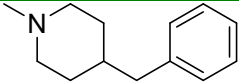
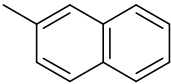
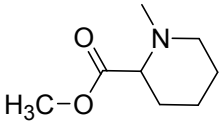
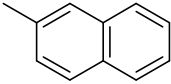
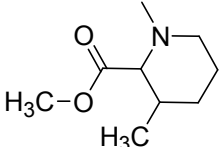
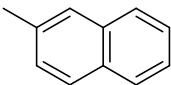
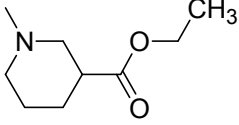
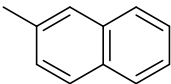
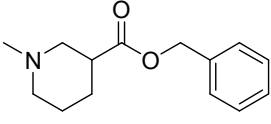
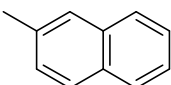
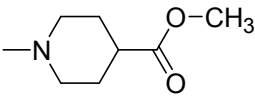
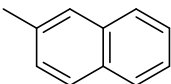
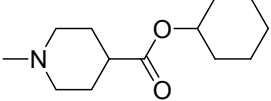
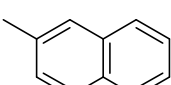
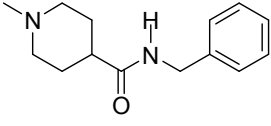
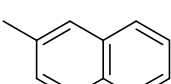
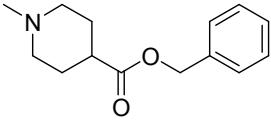
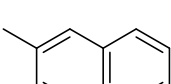
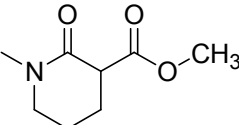
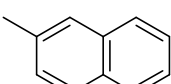
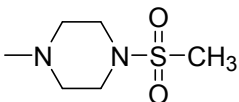


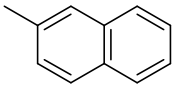
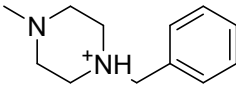
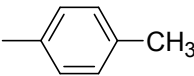
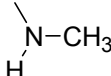
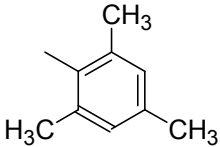
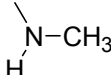
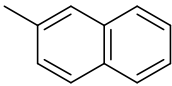
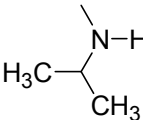
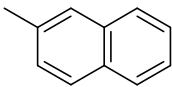
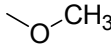
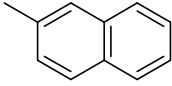
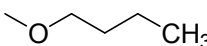
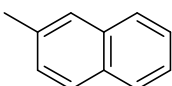
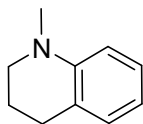
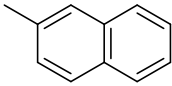
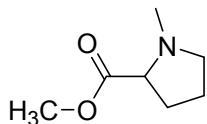
No.	P / T	R	<i>pK_i</i>
38	P		6.377
39	T		4.886
40	T		3.638
41	T		6.432

3TAPAP Derivatives

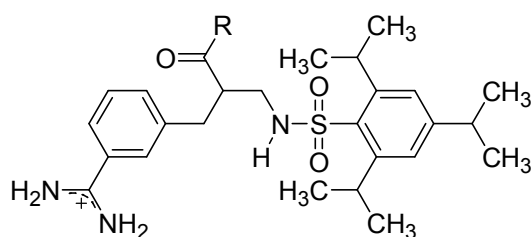


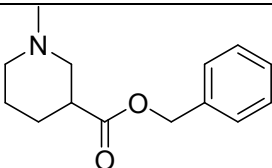
No.	P / T	R ₁	R ₂	<i>pK_i</i>
42 ^{b)}	T			4.569
43	P			4.824

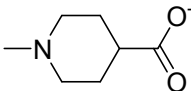
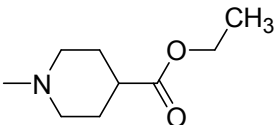
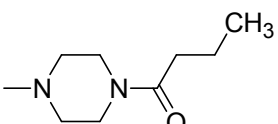
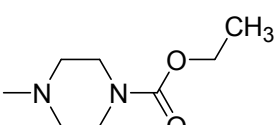
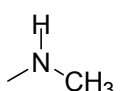
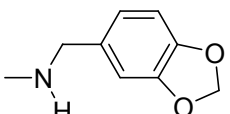
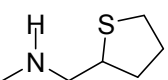
44	T			5.108
45	T			7.155
46	T			6.284
47	T			6.444
48	T			6.523
49	P			7.770
50	T			5.678
51	T			5.509
52	T			6.678
53	T			6.921
54	T			8.481

55	T			5.699
56	T			4.745
57	T			4.357
58	T			5.208
59	T			6.553
60	T			7.886
61	T			6.921
62	T			4.553

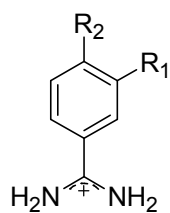
TIPPS Derivatives

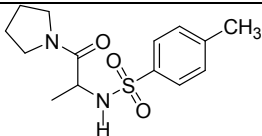
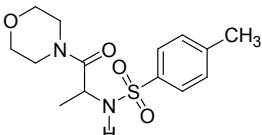


No.	P / T	R	pK_i
63	P		6.046

64	T		5.056
65	T		6.678
66	T		5.921
67	T		6.174
68	P		5.420
69	T		5.987
70	T		7.081

Glycinic Acid Derivatives

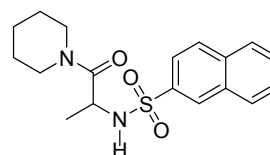


No.	P / T	R ₁	R ₂	<i>pK_i</i>
71	T		H	5.387
72	P		H	5.149

73

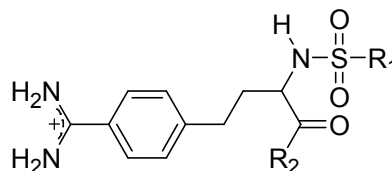
T

H



4.310

Butyric Acid Derivatives



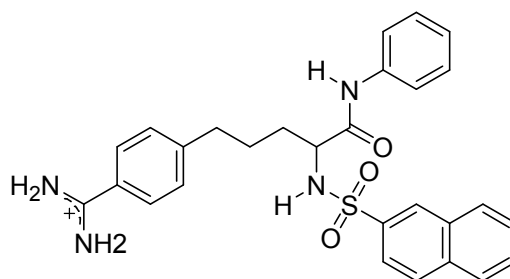
No.	P / T	R ₁	R ₂	<i>pK_i</i>
74	P			4.398
75	T			4.469
76	T			3.886

Valerianic Acid Derivatives

No.	P / T	<i>pK_i</i>
77 ^{b)}	T	7.022
78 ^{b)}	P	6.481

79^{b)}

T

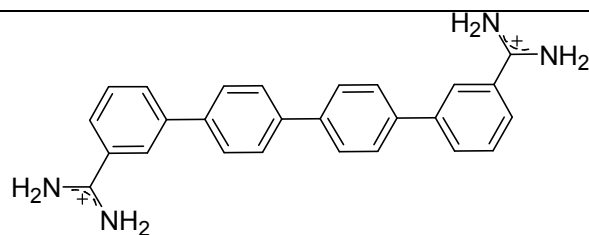


4.796

Other

80

N



5.538

a) T: Training set compound; P: External test set compound; N: Not considered. b) The whole ligand is depicted.

Table 2: q^2 values obtained by AFMoC using different combinations of interaction field^{a)}

Combination of fields ^{b)}	q^2 ^{c)}
C.3/C.ar/O.3/O.2/O.co2/N.am	0.57 (0.54)
C.3/C.ar/O.3/O.2/O.co2/S.3	0.54 (0.52)
C.3/C.ar/O.3/O.co2/N.am	0.49 (0.52)
C.3/C.ar/O.3/O.2/S.3	0.53 (0.51)
C.3/C.ar/O.3/O.co2	0.49 (0.55)
C.3/C.ar/O.3/N.am	0.51 (0.55)
C.3/C.ar/O.3/O.2	0.55 (0.55)
C.3/C.ar/O.2/N.am	0.53 (0.51)
C.3/C.ar/O.2/O.co2	0.55 (0.56)

a) Models are based on PLS calculations using the experimental thrombin structure, a σ value of 0.7 Å, and a grid spacing of 1.0 Å. b) The denotation of the combination of interaction fields follows the atom type convention of SYBYL¹, further details are given in reference². c) Values are given considering only pK_i^{PLS} or considering pK_i^{total} (values in parentheses).

Table 3: Statistical results of leave-10-out cross-validation runs^{a)}

Model	$q^{2\text{b)}$	$S_{press}^{\text{c)}$	Components
1ets ^{d)}	0.56 (0.53)	0.82	2
HM 1	0.56 (0.53)	0.78	3
HM 2	0.40 (0.39)	0.94	3
HM 3	0.39 (0.38)	0.94	1
HM 1+2 ^{e)}	0.67 (0.65)	0.76	3
HM 1+3 ^{e)}	0.63 (0.60)	0.81	3
HM 2+3 ^{e)}	0.49 (0.46)	0.93	1
HM 1+2+3 ^{e)}	0.62 (0.59)	0.82	3

a) Models are based on PLS calculations using a σ value of 0.7 Å and a grid spacing of 1.0 Å.

In the case of 1ets, HM1, HM2, and HM3, the results are averaged over 100 repetitions of the leave-multiple out procedures. In the case of the consensus models, only one leave-multiple out procedure was performed due to the computational burden.

b) Values are given considering only pK_i^{PLS} or considering pK_i^{total} (values in parentheses).

c) In logarithmic units.

d) The protein structure of PDB entry 1ets was used.

e) Results are given for models obtained after a VINFM-based variable selection was performed.

Table 4: q^2 values for ten AFMoC models obtained with randomly scrambled affinity data^{a)}

Model	$q^{2b)}$	Components
1	-0.15 (-0.22)	1
2	-0.29 (-0.37)	2
3	-0.31 (-0.38)	1
4	-0.38 (-0.46)	1
5	-0.19 (-0.26)	1
6	-0.09 (-0.15)	1
7	-0.11 (-0.17)	2
8	-0.29 (-0.36)	1
9	-0.08 (-0.14)	1
10	-0.24 (-0.31)	2

a) Models are based on PLS calculations using the experimental thrombin structure, a σ value of 0.7 Å, and a grid spacing of 1.0 Å. b) Values are given considering only pK_i^{PLS} or considering pK_i^{total} (values in parentheses).

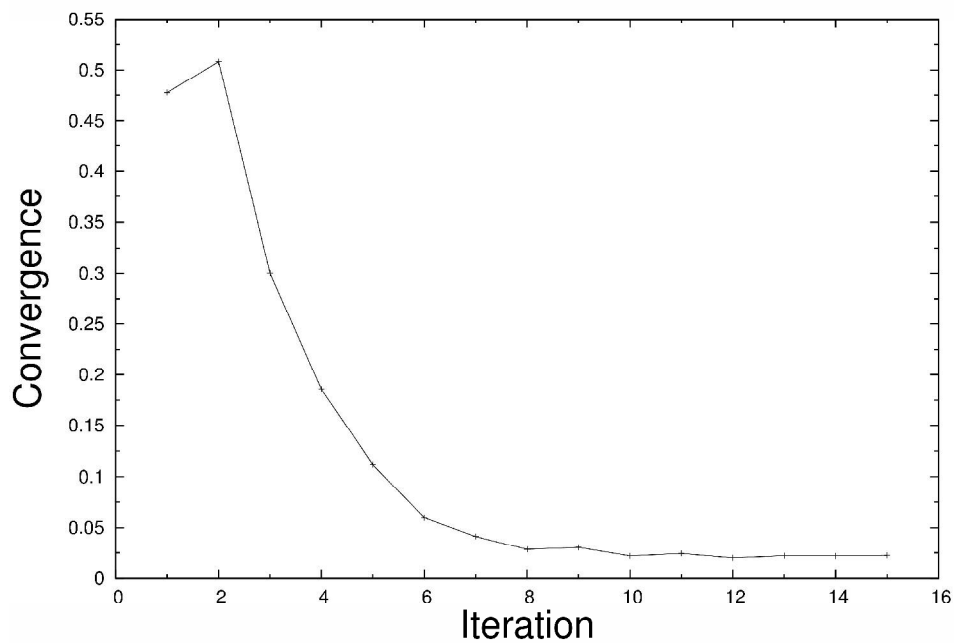


Figure 1: Convergence (in log units) as defined in eq. 7 as a function of the number of iterations performed in the MMB PLS calculations applied to HM 1+2+3.

References

1. SYBYL *Molecular Modeling Software*, Tripos Inc., St. Louis, MO.
2. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based Scoring Function to Predict Protein-Ligand Interactions. *J Mol Biol* **2000**, 295, (2), 337-356.