



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

FABP4 inhibitors 3D-QSAR model and isosteric replacement of BMS309403 datasets



Giuseppe Floresta^{a,b,c,*}, Agostino Cilibrizzi^{c,d},
 Vincenzo Abbate^d, Ambra Spampinato^a, Chiara Zagni^a,
 Antonio Rescifina^{a,**}

^a Department of Drug Sciences, University of Catania, V.le A. Doria 6, 95125 Catania, Italy

^b Department of Chemical Sciences, University of Catania, V.le A. Doria, 95125 Catania, Italy

^c Institute of Pharmaceutical Science, King's College London, Stamford Street, London SE1 9NH, UK

^d King's Forensics, School of Population Health & Environmental Sciences, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

ARTICLE INFO

Article history:

Received 29 November 2018

Received in revised form

12 December 2018

Accepted 13 December 2018

Available online 19 December 2018

ABSTRACT

The data have been obtained from FABP4 inhibitor molecules previously published. The 120 compounds were used to build a 3D-QSAR model. The development of the QSAR model has been undertaken with the use of Forge software using the PM3 optimized structure and the experimental IC₅₀ of each compound. The QSAR model was also employed to predict the activity of 3000 new isosteric derivatives of BMS309403. The isosteric replacement was also validated by the synthesis and the biological screening of three new compounds reported in the related research article "3D-QSAR assisted identification of FABP4 inhibitors: An effective scaffold hopping analysis/QSAR evaluation" (Floresta et al., 2019).

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <https://doi.org/10.1016/j.bioorg.2018.11.045>

* Corresponding author at: Department of Drug Sciences, University of Catania, V.le A. Doria 6, 95125 Catania, Italy

** Corresponding author.

E-mail addresses: giuseppe.floresta@unict.it (G. Floresta), arescifina@unict.it (A. Rescifina).

<https://doi.org/10.1016/j.dib.2018.12.047>

2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications table

Subject area	Computational Chemistry
More specific subject area	Three-Dimensional Quantitative Structure-Activity Relationship (3D-QSAR) modeling
Type of data	Tables, figures
How data was acquired	Statistical modeling and online databases
Data format	Raw and analyzed
Experimental factors	The whole dataset consists of 120 FABP4 ligands and 3000 isosteric derivatives of BMS309403
Experimental features	The 3D-QSAR model has been developed using Forge as software. Chemical structure descriptors and pIC_{50} were used as variables. Spark was used for the isosteric replacement
Data source location	Department of Drug Sciences, University of Catania, Italy
Data accessibility	Data is with this article
Related research article	G. Floresta, A. Cilibrizzi, V. Abbate, A. Spampinato, C. Zagni, A. Rescina, 3D-QSAR assisted identification of FABP4 inhibitors: An effective scaffold hopping analysis/QSAR evaluation, <i>Bioorganic Chemistry</i> , 84 (2019) 276–284 [1].

Value of the data

- FABP4 recently demonstrated an interesting molecular target for the treatment of type 2 diabetes, other metabolic diseases and some type of cancers.
 - QSAR modeling data was generated to provide a method useful in finding or repurposing novel FABP4 ligands.
 - The model has also been used to predict the activity of 3000 isosteric derivatives of BMS309403.
 - The data can be used by others to build their own model.
 - The data can be used for the synthesis of some potent suggested compounds.
-

1. Data

FABP4 recently demonstrated an interesting molecular target for the treatment of type 2 diabetes, other metabolic diseases and some type of cancers [2–10]. Recently, a variety of effective FABP4 inhibitors have been developed [11], but unfortunately, none of them is currently in the clinical research phases (Table 1). CAMD (computer aided molecular design) shows a promising and effective tool for the identification of FABP4 inhibitors [12–15]. In line with our recent interest in the development of QSAR models and related applications [16–24], in order to identify novel hit compounds, herein we report the dataset and the parameter used to build a 3D-QSAR model for FABP4. This dataset is reported in Tables 2 and 3, were the molecules used in the training set (96) and in the test set (24) are reported, respectively. Information for the building of the 3D-QSAR model is reported in Figs. 1–9. Moreover, the 3D-QSAR model was also used to predict the biological activity of 3000 new isosteric derivatives of BMS309403 derived from a scaffold-hopping analysis, the analyzed areas of the selected compounds and the Spark's parameters used for the isosteric replacement are reported in Figs. 8 and 9. The results of the isosteric replacement of different portion of BMS309403 are reported in Tables S4–S9.

Table 1
PDB codes and molecules used as reference compounds for ligand-based alignment.

PDB code	2D structure	PDB code	2D structure
2NNQ		4NNT	
3FR5		4NNS	
3FR4		1TOW	
3FR2		1TOU	

Table 2
SMILES, experimental and predicted pIC₅₀ values of the molecules in the training set.

No	SMILES	pIC ₅₀	
		Exp.	Pred.
1	FC(F)(F)[C@H]1CCc2c(C1)c(c(c(n2)C3CCCC3)C=4[N-N]N=NN4)-c5ccnc(c5)C	8.0	8.0
2	CC1(CCCC1)c2c(c(c3c(n2)CCCC3)-c4ccnc(c4)C)C=5[N-N]N=NN5	8.0	8.0
3	Clc1c(F)cc2c(c(c(c(N(CC)CC)n2)C=3[N-N]N=NN3)-c4cccc4)c1	7.9	7.9
4	Clc1c(F)cc2c(c(c(c(n2)C(CC)CC)C=3[N-N]N=NN3)-c4cccc4)c1	7.8	7.8
5	OCC1(CCCC1)c2c(c(c3c(n2)CCCC3)-c4ccnc(c4)C)C=5[N-N]N=NN5	7.7	7.7
6	CCCC[C@H]1CCc2c(C1)c(c(c(n2)C3CCCC3)COC)C=4[N-N]N=NN4)-c5cccc5	7.7	7.7
7	FC(F)(F)c1ccc2c(c(c(c(N3CCCC3)n2)C=4[N-N]N=NN4)-c5cccc5)c1	7.5	7.5
8	Clc1ccc2c(c(c(c(n2)C3CC3)C([O-])=O)-c4cccc4)c1	7.4	7.4
9	Clc1ccc2c(c(c(c(N(CC)C)n2)C=3[N-N]N=NN3)-c4cccc4)c1	7.3	7.4
10	Clc1cc(Cl)cc(NC(=O)NC2(CCCC2)C([O-])=O)c1-c3cccc3	7.3	7.3
11	Clc1c(F)cc(c(NC(=O)NC2(CCCC2)C([O-])=O)c1)-c3cccc3	7.0	7.0
12	O=C(N)c1cccc1Cn2c3c(cccc3c4CCCCc42)C([O-])=O	7.0	7.0
13	n1c2c(CCCCC2)c(c(c1C3CCCC3)C=4[N-N]N=NN4)-c5ccnc5	7.0	6.9
14	Clc1ccc(c(NC(=O)NC2(CCCC2)C([O-])=O)c1)-c3ccc(F)cc3	6.9	6.9
15	FC(F)(F)c1cccc1Cn2c3c(cccc3c4CCCCc42)C([O-])=O	6.4	6.5
16	Fc1ccc(-c2c(c(n2)-c3cccc3-c4cccc(OCC([O-])=O)c4)CC)-c5cccc5)cc1	6.5	6.5
17	[O-]C(=O)c1ccc2c3CCCCc3n(c12)Cc4cccc4	6.2	6.3
18	Fc1cccc1Cn2c3c(cccc3c4CCCCc42)C([O-])=O	6.4	6.3
19	Fc1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	6.4	6.3
20	FC(F)(F)c1cccc1Cn2c3c(cccc3c4CCCCc42)C([O-])=O	6.2	6.3
21	[O-]C(=O)CCn1c2cccc2c3cccc31	6.2	6.3
22	FC(F)(F)c1ccc(c(NC(=O)NC2(CCCC2)C([O-])=O)c1)-c3cccc3	6.3	6.2
23	[O-]C(=O)c1ccc2c3CCCCc3n(c12)Cc4cccc(OC)c4	6.3	6.2
24	Fc1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	6.1	6.2
25	FC(F)(F)c1cc(O)nc(SCc2ccc(OC)cc2)n1	6.2	6.2
26	[O-]C(=O)c1ccc2c(n(c3CCCC23)Cc4cccc4)c1	6.1	6.1
27	[O-]C(=O)c1ccc2c3CCCc3n(c12)Cc4cccc4	6.1	6.1
28	[O-]C(=O)c1ccc2c3CCCCc3n(c12)Cc4cccc4OC	6.2	6.1
29	[O-]C(=O)c1ccc2c3CCCCc3n(c12)Cc4ccc(C)cc4	6.0	6.1
30	Fc1cccc1Cn2c3c(cccc3c4CCCCc42)C([O-])=O	6.2	6.1
31	Fc1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	6.1	6.1
32	[O-]C(=O)CCn1c2cccc2c3cccc31	6.1	6.1
33	FC(F)(F)c1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	6.0	6.0
34	FC(F)(F)c1cc(O)nc(SCC(=O)N2CCCC2)n1	6.0	6.0
35	O=S(=O)(n1ccc2ccc(cc21)C)c3ccsc3C([O-])=O	5.9	5.9
36	Brc1ccc2c(ccn2S(=O)(=O)c3ccsc3C([O-])=O)c1	5.9	5.9
37	FC(F)(F)c1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	5.8	5.7
38	FC(F)(F)c1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	5.6	5.7
39	FC(F)(F)c1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)cc1	5.7	5.7
40	O=S(=O)(n1cc(c2cccc21)C)c3ccsc3C([O-])=O	5.8	5.7
41	[O-]C(=O)c1ccc2c3CCCCc3n(c12)Cc4ccc(OC)cc4	5.6	5.6
42	[O-]C(=O)[C@H](Oc1ccc(-c2cccc2-n3c(c(c(n3)-c4cccc4)-c5cccc5)CC)c1)C	5.6	5.6
43	O=S(=O)(n1ccc2ccc(OC)c21)c3ccsc3C([O-])=O	5.6	5.6
44	O/N=C/1CCCc2c1c3ccc(c3n2Cc4cccc4)C([O-])=O	5.5	5.5
45	Clc1cccc(-n2c(-c3cccc3)cc(n2)-c4cccc4OCCCC([O-])=O)c1	5.6	5.5
46	[O-]C(=O)[C@H](Oc1ccc(-c2cccc2-n3c(c(c(n3)-c4cccc4)-c5cccc5)CC)c1)CC	5.5	5.5
47	Fc1ccc2c(ccn2S(=O)(=O)c3ccsc3C([O-])=O)c1	5.5	5.5
48	[O-]C(=O)c1ccc2c(c(n(c12)Cc3cccc3)C)C	5.4	5.4
49	Clc1ccc(-n2c(-c3cccc3)cc(n2)-c4cccc4OCCCC([O-])=O)cc1	5.4	5.4
50	Clc1cccc1-n2c(-c3cccc3)cc(n2)-c4cccc4OCCCC([O-])=O	5.4	5.4
51	[O-]C(=O)c1c(C(C)C)cc(C(C)C)c1C(C)C	5.4	5.4
52	O=S(=O)(n1c2cccc2c3cccc31)c4cccc4C([O-])=O	5.4	5.4
53	Fc1ccc2ccn(S(=O)(=O)c3ccsc3C([O-])=O)c2c1	5.4	5.4
54	FC(F)(F)c1cc(O)nc(NC2ccc(OC)cc2)n1	5.4	5.4
55	[O-]C(=O)CCOC1cccc1-c2cc(c(n2)-c3cccc3)-c4ccc(cc4)C	5.3	5.3
56	Brc1ccc(-n2c(-c3cccc3)cc(n2)-c4cccc4OCCCC([O-])=O)cc1	5.3	5.3
57	Fc1ccc(-c2c(nn(c2CC)-c3cccc3-c4cccc(OCC([O-])=O)c4)-c5cccc5)cc1	5.3	5.3
58	[O-]C(=O)CCCO1cccc1-c2cc(c(n2)-c3cccc3)-c4cccc4	5.2	5.2

Table 2 (continued)

No	SMILES	pIC ₅₀	
		Exp.	Pred.
59	O=S(=O)(n1ccc2cc(c2c1)C)c3ccsc3C([O-])=O	5.2	5.2
60	O=S(=O)(n1ccc2ccc(OC)c21)c3ccccc3C([O-])=O	5.2	5.2
61	Brc1ccc(-c2cc(nn2-c3ccccc3)-c4ccccc4OCCCC([O-])=O)cc1	5.0	5.0
62	Fc1ccc(-n2c(-c3ccccc3)cc(n2)-c4ccccc4OCCCC([O-])=O)cc1	5.0	5.0
63	[O-]C(=O)CCCOc1cccc1-c2cc(n(n2)-c3ccc(C(C)C)cc3)-c4ccccc4	5.0	5.0
64	[O-]C(=O)CCn1c2ccccc2c3ccccc31	5.0	5.0
65	O=S(=O)(n1ccc2c(c2c1)C)c3ccsc3C([O-])=O	5.1	5.0
66	O=S(=O)(n1ccc2cc(OC)c21)c3ccsc3C([O-])=O	5.1	5.0
67	O=S(=O)(n1cc(c2ccccc21)C)c3ccccc3C([O-])=O	5.1	5.0
68	O=S(=O)(n1ccc2c(c2c1)C)c3ccccc3C([O-])=O	4.9	4.9
69	Brc1ccc2c(ccn2S(=O)(=O)c3ccccc3C([O-])=O)c1	4.9	4.9
70	[O-]C(=O)CCCOc1cccc1-c2cc(n(n2)-c3ccc(OC)cc3)-c4ccccc4	4.9	4.8
71	[O-]C(=O)CCCOc1cccc1-c2cc(n(n2)C3CCCCC3)-c4ccccc4	4.8	4.8
72	Brc1ccc2c(n(S(=O)(=O)c3c(C)C)cc(C(C)C)cc3C(C)C)cn2)c1	4.8	4.8
73	Clc1ccc2c(nc(n2S(=O)(=O)c3c(C)C)cc(C(C)C)cc3C(C)C)C)c1	4.8	4.8
74	O=S(=O)(n1cncc1)c2c(C(C)C)cc(C(C)C)cc2C(C)C	4.7	4.8
75	Clc1cccc1CNe2nc(O)cc(n2)C(F)F	4.6	4.7
76	FC(F)F)c1cc(O)nc(n1)CCc2ccc(OC)cc2	4.6	4.7
77	O=C1CCCc2c1c3cccc(c3n2Cc4ccccc4)C([O-])=O	4.6	4.6
78	[O-]C(=O)CCCOc1cccc1-c2cc(n(n2)C3CCCCC3)-c4ccccc4	4.6	4.6
79	O=S(=O)(n1ccc2cc(c2c1)C)c3ccccc3C([O-])=O	4.5	4.6
80	FC(F)F)c1cc(O)nc(n1)N(Cc2ccccc2)C	4.6	4.6
81	Clc1ccc(-c2cc(nn2-c3ccccc3)-c4ccccc4OCCCCC([O-])=O)cc1	4.5	4.5
82	FC(F)F)c1cc(O)nc(NCC(=O)N2CCCC2)n1	4.4	4.4
83	Clc1cccc(CNc2nc(O)cc(n2)C(F)F)c1	4.5	4.4
84	FC(F)F)c1cc(O)nc(NCc2ccc(C)cc2)n1	4.5	4.4
85	Clc1ccc(-c2cc(nn2-c3ccccc3)-c4ccccc4OCCCCC([O-])=O)cc1	4.1	4.2
86	Brc1ccc(-c2cc(nn2-c3ccccc3)-c4ccccc4OCCCCC([O-])=O)cc1	4.1	4.1
87	O=S(=O)(n1ccc2c(OC)c21)c3ccccc3C([O-])=O	4.1	4.1
88	O=S(=O)(N)c1c(C(C)C)cc(C(C)C)cc1C(C)C	4.0	4.0
89	[O-]C(=O)Cn1c2ccccc2c3ccccc31	4.0	4.0
90	FC(F)F)c1cc(O)nc(n1)Nc2ccc(-c3ccccc3)cc2	4.0	4.0
91	FC(F)F)c1cc(O)nc(NCc2ccccc2)n1	4.0	4.0
92	FC(F)F)c1cc(O)nc(n1)CCc2ccccc2	4.0	4.0
93	FC(F)F)c1cc(O)nc(NCc2ccccc2)n1	4.0	3.9
94	[O-]C(=O)CCCOc1cccc1-c2cc(n(n2)-c3ccccc3)-c4ccc(cc4)C	3.6	3.6
95	Clc1ccc(CNc2nc(O)cc(n2)C(F)F)cc1	5.5	3.5
96	Clc1ccc(-c2cc(nn2-c3ccccc3)-c4ccccc4OCC([O-])=O)cc1	2.0	2.0

2. Experimental design, materials and methods

2.1. Compounds alignments

With the aim to generate a plausible and consistent set of alignment molecules, before running the regression analysis, we evaluated two different types of alignment (Fig. 1).

First, we evaluated a structure-based alignment, based on the docking of the different ligands on the active site of the protein. All 120 structures, optimized at the PM3 level of theory [25–27], have been converted into pdbqt format using Babel, [28] and subsequently docked in the active site of FABP4. Molecular docking was performed using the three-dimensional crystal structures of substrate-free fatty acid binding protein 4 in complex with BMS309403 (PDB ID: 2NNQ) obtained from the Protein Data Bank (PDB, <http://www.rcsb.org/pdb>). AutoDock Vina (version 1.1.2) [29], was used for all docking experiments. The default values of the docking parameters in AutoDock Vina were all maintained, except for “exhaustiveness” that was set to 15. A grid box of 18 Å × 18 Å × 18 Å was used, encompassing the inhibitor binding cavity of FABP4 and centered on the ligand. The binding modes were clustered through the root mean square deviation among the Cartesian coordinates of

the ligand atoms. The docking results were ranked based on the binding free energy. After the calculations with AutoDock Vina, all the generated structures were manually checked, in order to ensure a correct positioning within the binding site. Then the generated structures were imported into Forge [30] to build the Structure-based 3D-QSAR model. A classic ligand-based alignment is the second type of alignment that was evaluated. This was carried out with the same software used for the

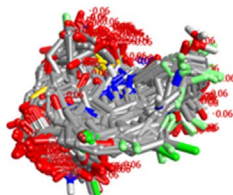
Table 3SMILES, experimental, and predicted pIC₅₀ values of the molecules in the test set.

No	SMILES	pIC ₅₀	
		Exp.	Pred.
1	FC(F)F)c1ccc2c(c(c(c(N(CC)CC)n2)C=3{N-]N=NN3)-c4cccc4)c1	7.6	7.8
2	Clc1c(F)cc2c(c(c(c(N3CCCC3)n2)C=4{N-]N=NN4)-c5cccc5)c1	7.9	7.3
3	Clc1ccc(c(NC(=O)NC2(CCCC2)C([O-])=O)c1)-c3cccc3	6.8	6.5
4	O=C(N)c1cccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)c1	7.2	6.2
5	[O-]C(=O)c1ccc2c(c3CCCCc3n2Cc4cccc4)c1	4.6	6.1
6	Fc1ccc(Cn2c3c(cccc3c4CCCCc42)C([O-])=O)cc1	6.1	6.1
7	[O-]C(=O)c1cccc2c3CCCCc3n(c12)Cc4cccc4	6.2	5.9
8	Fc1cccc(c1Cn2c3c(cccc3c4CCCCc42)C([O-])=O)C(F)(F)F	5.7	5.9
9	O=S(=O)(n1c2cccc2c3cccc31)c4ccsc4C([O-])=O	6.0	5.9
10	[O-]C(=O)c1cccc2c3CCCCc3n(CCC)c12	6.4	5.7
11	[O-]S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C	5.1	5.7
12	O=S(=O)(n1ccc2ccc(OC)cc21)c3ccsc3C([O-])=O	5.6	5.7
13	[O-]C(=O)c1cccc2c3CCCCc3n(CCC)c12	6.1	5.6
14	Fc1cccc2ccn(S(=O)(=O)c3ccsc3C([O-])=O)c12	5.4	5.4
15	[O-]C(=O)CCCOc1cccc1-c2cc(n2)-c3cccc3)-c4cccc4	5.5	5.3
16	Clc1ccc(-c2cc(nn2-c3cccc3)-c4cccc4OCCCC([O-])=O)cc1	5.2	5.2
17	Fc1cccc2c1ccn2S(=O)(=O)c3cccc3C([O-])=O	5.0	5.2
18	Clc1ccc(CN(c2nc(O)cc(n2)C(F)(F)F)C)cc1	5.4	5.1
19	FC(F)(F)c1cc(O)nc(Nc2cccc2)n1	4.0	4.8
20	Brc1ccc2c(n(S(=O)(=O)c3c(C(C)C)cc(C(C)C)cc3C(C)C)c(n2)C)c1	4.1	4.7
21	O=S(=O)(n1c(nc2cccc21)C)c3c(C(C)C)cc(C(C)C)cc3C(C)C	4.0	4.6
22	[O-]C(=O)CCCOc1cccc1-c2cc(n2)C3CCCC3)-c4cccc4	4.8	4.5
23	O=S(=O)(n1ccc2c(OC)cccc21)c3ccsc3C([O-])=O	4.9	4.3
24	FC(F)(F)c1cc(O)nc(n1)NC2CCCC2	4.5	4.2

Preferred alignment protocols

Structure-based Autodock Vina Docking

- AutoDock Vina 1.1.2
- Retain 8 poses per ligand
- Manually select preferred poses
- Tweak and refine inconsistent poses



Ligand-based Forge

- Substructure alignment, protein present, no constraints
- Use most accurate conformation hunt
- Tweak and refine inconsistent groups
- Use 8 different references to align ligands

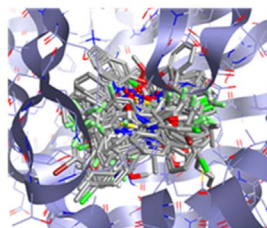


Fig. 1. Comparison of alignment methods.

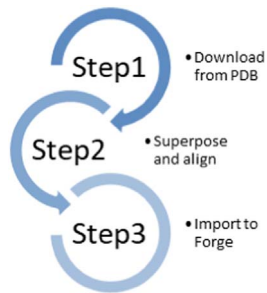


Fig. 2. Schematic representation of the process adopted to obtain the template compounds for the ligand-based alignment.

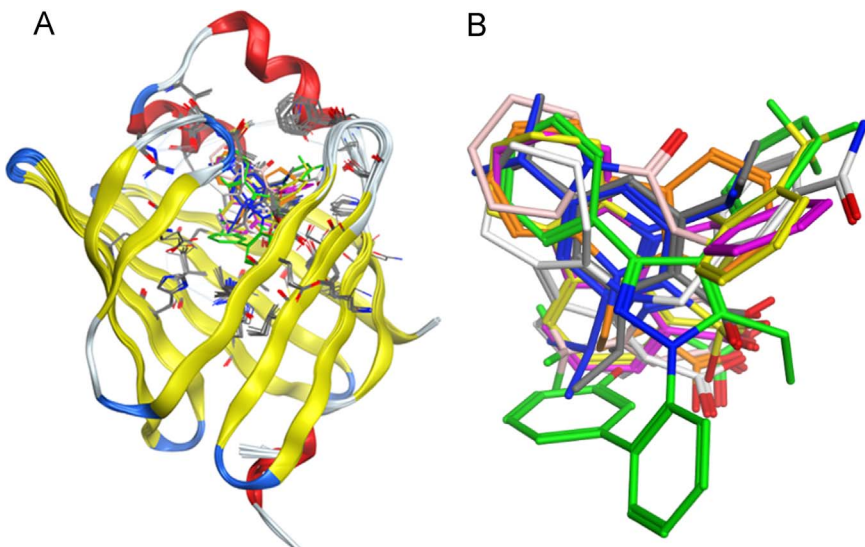


Fig. 3. A) Protein and inhibitors aligned. B) Aligned inhibitors imported to Forge for ligand-based alignment.

The screenshot shows the 'Conformation Hunt' window in the Forge software. The window has three tabs: 'Conformation Hunt', 'Alignment', and 'Build Model'. The 'Conformation Hunt' tab is active, showing the following settings:

- Calculation Method: [Custom] (dropdown menu)
- Buttons: Save As... (button), Delete (button)
- Delete existing conformations
- Perform Conformation Hunt
- Maximum number of conformations: 500 (spin box)
- No. of high-T dynamics runs for flexible rings: 20 (spin box)
- Gradient cutoff for conformer minimization: 0,100 kcal/mol/Å (spin box)
- Filter duplicate conformers at RMS: 0,50 Å (spin box)
- Energy window: 2,50 kcal/mol (spin box)
- Acyclic secondary amide handling: Use input amide geometry (dropdown menu)
- Turn off Coulombic and attractive vdW forces:
- Use external tool for conformation generation:

Fig. 4. Forge's parameters used for conformation hunt.

building of the model. All the optimized structures, together with their respective IC_{50} values, were imported into Forge (10.4.2, Cresset, Litlington, Cambridgeshire, UK, <http://www.cresset-group.com/forge>) [30–34] for setting-up the field-based 3D-QSAR model. Eight different molecules were

Conformation Hunt Alignment Build Model

Calculation Method: [Custom] Save As... Delete

Delete existing alignments

Perform Alignment

Invert achiral imported confs

Take shortcuts in alignments

Maximum-common-substructure conformers and alignment

Matching rules Normal (element + hybridisation)

Allow conformations to move

Perform Scoring

Score method for multiple references Weighted Average

Reference weights

Reference	1	2	3	4	5	6	7	8
Weight	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Weight%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%

Fraction of score from shape similarity 0.50

Reference into db fieldpoints weight 0.50

Hardness of protein excluded volume Soft

Add/remove field constraints Mark field points

Fig. 5. Forge's parameters used for alignment.

Conformation Hunt Alignment Build Model

Calculation Method: Field QSAR Normal Save As... Delete

Activity: IC50uM Activity Manager

Field QSAR model

Maximum number of components 20

Sample point minimum distance 1.0 A

Generate samples from references

Number of Y scrambles 50

Fields to use Electrostatic Volume

Weight molecules by similarity

Weight ramp type Linear

Minimum similarity 0.00

Maximum similarity 1.00

Cross-validation

Cross-validation type Leave-one-out

Training set to use as validation data 20%

Repeats 1000

Fig. 6. Forge's parameters used to build the QSAR model.

Model statistics:

=====

Comps	R ²	Q ²	Test R ²	RMSE	RMSEpred	Tau	Tau-pred
0	-0.044	-0.023	-0.001	1.093	1.096	0.494	-0.910
1	0.620	0.508	0.641	0.666	0.767	0.587	0.492
2	0.781	0.541	0.625	0.501	0.740	0.675	0.514
3	0.897	0.585	0.718	0.347	0.711	0.787	0.573
4	0.948	0.618	0.745	0.244	0.681	0.855	0.568
5	0.968	0.636	0.743	0.192	0.663	0.895	0.598
6	0.983	0.651	0.772	0.142	0.651	0.920	0.594
7	0.988	0.671	0.746	0.118	0.631	0.929	0.621
8	0.993	0.673	0.757	0.088	0.629	0.947	0.621
9	0.996	0.683	0.744	0.068	0.619	0.958	0.633
10	0.997	0.687	0.733	0.058	0.616	0.957	0.626
11*	0.998	0.688	0.731	0.044	0.615	0.969	0.626
12	0.999	0.683	0.729	0.037	0.619	0.969	0.618
13	0.999	0.680	0.726	0.030	0.622	0.976	0.616
14	1.000	0.679	0.731	0.022	0.623	0.983	0.613
15	1.000	0.679	0.730	0.018	0.623	0.985	0.615
16	1.000	0.678	0.728	0.015	0.624	0.988	0.616
17	1.000	0.678	0.728	0.011	0.625	0.988	0.615
18	1.000	0.677	0.730	0.008	0.625	0.991	0.615
19	1.000	0.678	0.728	0.006	0.625	0.993	0.616
20	1.000	0.677	0.728	0.005	0.625	0.995	0.616

Preferred number of components: 11

Fig. 7. Model statistics for FABP4 model.

chosen as a template for the calculations of field points and as a template for the alignment. These eight molecules were selected since they are present in crystallized forms with FABP4 (PDB IDs: 2NNQ, 3FR2, 3FR4, 3FR5, 4NNS, 4NNT, 1TOU and 1TOW, Table 1) [35–38]. The structures, small protein, and inhibitors, were first downloaded from the Protein Data Bank (PDB); the amino acid sequence was then superposed and aligned with YASARA (version 17.8.15) to get also the ligands in the binding site aligned and superposed, thus the eight molecules were imported on Forge (Figs. 2 and 3).

The XED (eXtended Electron Distribution) force field was used to generate the field point. The compounds in the training set were aligned to the reference compound by maximum common substructure using a customized set-up for the conformation hunt:

- Max number of conformations: 500.
- RMS cut-off for duplicate conformers: 0.5 Å.
- Gradient cut-off for conformer minimization: 0.1 kcal/mol.
- Energy window: 2.5 kcal/mol.

The RMS cut-off for duplicate conformers parameter controls the similarity threshold below which two conformers are assumed identical. Conformations that gave a minimized energy outside the energy window were discarded.

All the alignments were manually checked to ensure the best possible model. All the field points of the training set were used to derive a gauge invariant set of sampling points, which reduced the number of descriptors that needed to be taken into account, with a distance of 1 Å between the sample points. Sample values were calculated, ensuring that all areas around the molecule (and possibly contributing to the activity) are properly described.

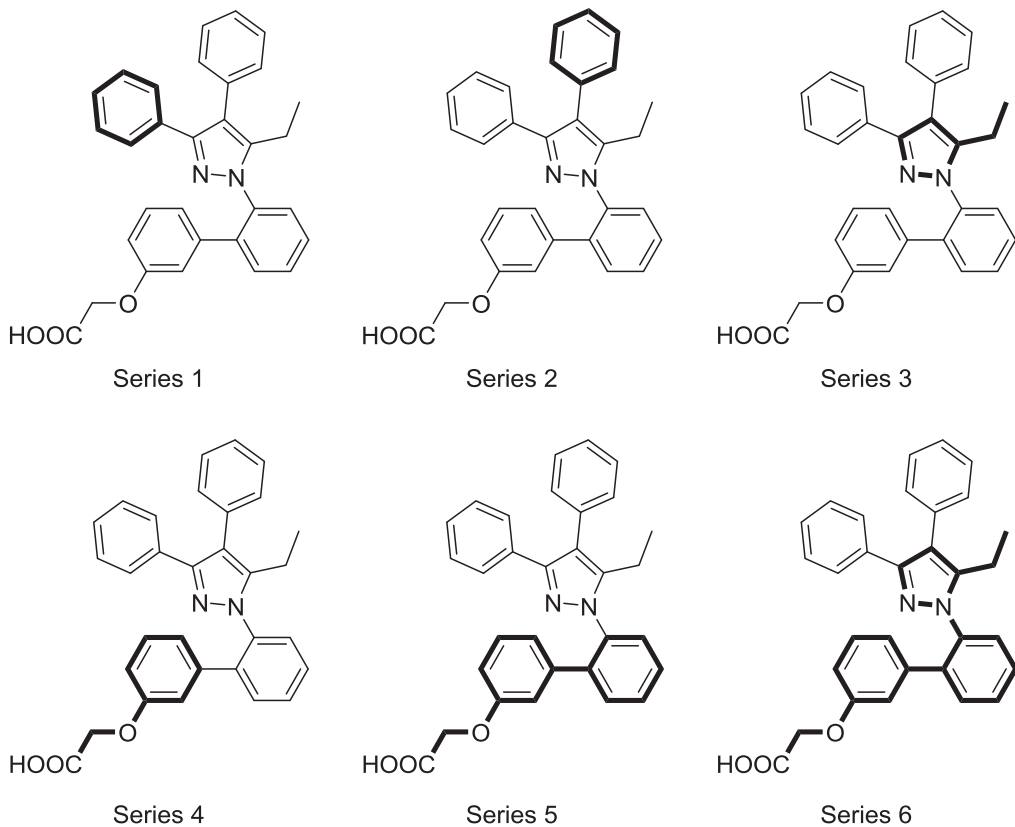


Fig. 8. The studied position for the bioisosteric replacement of BMS309403 are highlighted in bold.

Calculation Method: **Accurate but Slow**

Select one or more databases to search.

Name	Fragments	Description	Created On
<input checked="" type="checkbox"/> Creset			
<input checked="" type="checkbox"/> ChEMBL			
<input checked="" type="checkbox"/> ChEMBL_common 101156	101156	ChEMBL_21 (http://www.e...	2016-03-14 18:30:48
<input type="checkbox"/> ChEMBL_rare 156128	156128	ChEMBL_21 (http://www.e...	2016-03-14 18:41:25
<input checked="" type="checkbox"/> Zinc			
<input checked="" type="checkbox"/> VeryCommon 24894	24894	Zinc 15 very common frag...	2016-03-24 16:58:54
<input checked="" type="checkbox"/> Common 52508	52508	Zinc 15 common fragment...	2016-03-24 16:59:50
<input type="checkbox"/> LessCommon 115525	115525	Zinc 15 less common frag...	2016-03-24 17:00:55
<input type="checkbox"/> Creset Reagents			
<input type="checkbox"/> VEHICLE			

Fig. 9. Spark's parameters used for bio-isosteric replacement.

2.2. Statistical analysis

For the validation of the QSAR model, the leave-one-out method was used. 20 was the maximum number of components to extract from the PLS regression. 50 was the number of Y scrambles to use. The threshold of the sample point minimum distance was set to 1 Å. The Leave-one-out method was used during the validation of the QSAR model. The regression method used in Forge was PLS (SIMPLS algorithm) [39–43]. All the parameters for the QSAR model are resumed in Figs. 4–6.

The predictive ability of the generated QSAR model was confirmed by several statistical tests. The cross-validation regression coefficient (q^2) was calculated based on the PRESS (Prediction error sum of squares) and SSY (Sum of squares of deviation of the experimental values from their mean):

$$q^2 = 1 - \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^n (Y_{exp} - Y_{pred})^2}{\sum_{i=1}^n (Y_{exp} - Y_{mean})^2}$$

Y_{exp} = experimental activity of training set compound

Y_{pred} = predicted activity of training set compound

Y_{mean} = mean values of the activity of training set compound

The derived ligand-based approach results to be more reliable ($r^2 = 0.92$, $q^2 = 0.64$) than the structure-based alignment ($r^2 = 0.90$, $q^2 = 0.38$). The ligand-based 3D-QSAR align model was further validated with a set of external compounds (i.e. test set). Out of 120 molecules, we randomly choose 96 molecules (covering the whole range of activities of the compounds) as a training set to build the model, while the remaining 24 compounds served as a test set to evaluate the model.

The statistical reliability of this model was also validated by the determination of the r^2_{test} , using the following equation:

$$r^2_{test} = 1 - \frac{\sum_{i=1}^n (Y_{predtest} - Y_{test})^2}{\sum_{i=1}^n (Y_{test} - Y_{mean})^2}$$

$Y_{predtest}$ = predicted activity of test set compound by QSAR equation

Y_{test} = experimental activity of test set compound

Y_{mean} = mean values of the activity of training set compound

The 11-components model (Fig. 7) shows both good predictive and descriptive capability as it is shown by the good r^2 (0.99) and q^2 (0.69) [44] values for the training and the cross-validated training sets. The plot of experimental vs. predicted activity for the compounds, in both the training set and the cross-validated training set ($q^2 = 0.69$), shows a reasonable distribution of the values. The plot of experimental vs. predicted activity for the compounds in the test set is still reasonably good with only few outliers and a good cross-validated r^2 of 0.73.

2.3. Isosteric replacement

The isosteric replacement was performed using Spark as a software (10.4.0, Cresset, Litlington, Cambridgeshire, UK, <http://www.cresset-group.com/forge>) [30–34]. As reported in Fig. 8, Different portions of the BMS309403 were replaced. Then, the newly designed molecules were aligned with the 3D-QSAR model and evaluated. The replacement was performed through the same 178,558 fragments for each part, which derive from ChEMBL and Zinc databases (Fig. 9) [45,46]. Five hundred compounds were generated for each substitution producing 3000 hits (reported in Tables S4–S9). There of the suggested molecules were synthesized and tested as reported in the related research article [1].

Acknowledgements

Free academic licenses from Cresset and ChemAxon for their suites of programs are gratefully acknowledged.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.12.047>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.12.047>.

References

- [1] G. Floresta, A. Cilibrizzi, V. Abbate, A. Spampinato, C. Zagni, A. Rescifina, 3D-QSAR assisted identification of FABP4 inhibitors: an effective scaffold hopping analysis/QSAR evaluation, *Bioorg. Chem.* 84 (2019) 276–284.
- [2] Y. Adachi, S. Hiramatsu, N. Tokuda, K. Sharifi, M. Ebrahimi, A. Islam, Y. Kagawa, L. Koshy Vaidyan, T. Sawada, K. Hamano, Y. Owada, Fatty acid-binding protein 4 (FABP4) and FABP5 modulate cytokine production in the mouse thymic epithelial cells, *Histochem Cell Biol.* 138 (2012) 397–406.
- [3] I.M. Blecha, F. Siqueira, A.B. Ferreira, G.L. Feijo, R.A. Torres Junior, S.R. Medeiros, I.I. Sousa, G.G. Santiago, A.L. Ferraz, Identification and evaluation of polymorphisms in FABP3 and FABP4 in beef cattle, *Genet. Mol. Res.* 14 (2015) 16353–16363.
- [4] A. Bosquet, S. Guaita-Esteruelas, P. Saavedra, R. Rodriguez-Calvo, M. Heras, J. Girona, L. Masana, Exogenous FABP4 induces endoplasmic reticulum stress in HepG2 liver cells, *Atherosclerosis* 249 (2016) 191–199.
- [5] S. Guaita-Esteruelas, A. Bosquet, P. Saavedra, J. Guma, J. Girona, E.W. Lam, K. Amillano, J. Borrás, L. Masana, Exogenous FABP4 increases breast cancer cell proliferation and activates the expression of fatty acid transport proteins, *Mol. Carcinog.* 56 (2017) 208–217.
- [6] U. Harjes, E. Bridges, K.M. Gharpure, I. Roxanis, H. Sheldon, F. Miranda, L.S. Mangala, S. Pradeep, G. Lopez-Berestein, A. Ahmed, B. Fielding, A.K. Sood, A.L. Harris, Antiangiogenic and tumour inhibitory effects of downregulating tumour endothelial FABP4, *Oncogene* (2016).
- [7] Y. Luo, Z. Yang, D. Li, Z. Liu, L. Yang, Q. Zou, Y. Yuan, LDHB and FABP4 are associated with progression and poor prognosis of pancreatic ductal adenocarcinomas, *Appl. Immunohistochem. Mol. Morphol.* (2015).
- [8] D.M. Meng, L. Wang, J.R. Xu, S.L. Yan, L. Zhou, Q.S. Mi, Fabp4-Cre-mediated deletion of the miRNA-processing enzyme Dicer causes mouse embryonic lethality, *Acta Diabetol.* 50 (2013) 823–824.
- [9] M.R. Syamsunarno, T. Iso, H. Hanaoka, A. Yamaguchi, M. Obokata, N. Koitabashi, K. Goto, T. Hishiki, Y. Nagahata, H. Matsui, M. Sano, M. Kobayashi, O. Kikuchi, T. Sasaki, K. Maeda, M. Murakami, T. Kitamura, M. Suematsu, Y. Tsushima, K. Endo, G. S. Hotamisligil, M. Kurabayashi, A critical role of fatty acid binding protein 4 and 5 (FABP4/5) in the systemic response to fasting, *PLoS One* 8 (2013) e79386.
- [10] Z. Tang, Q. Shen, H. Xie, X. Zhou, J. Li, J. Feng, H. Liu, W. Wang, S. Zhang, S. Ni, Elevated expression of FABP3 and FABP4 cooperatively correlates with poor prognosis in non-small cell lung cancer (NSCLC), *Oncotarget* (2016).
- [11] G. Floresta, V. Pistrà, E. Amata, M. Dichiarà, A. Marrazzo, O. Prezzavento, A. Rescifina, Adipocyte fatty acid binding protein 4 (FABP4) inhibitors: A comprehensive systematic review, *Eur. J. Med. Chem.* 138 (2017) 854–873.
- [12] U. Tagami, K. Takahashi, S. Igarashi, C. Ejima, T. Yoshida, S. Takeshita, W. Miyanaga, M. Sugiki, M. Tokumasu, T. Hatanaka, T. Kashiwagi, K. Ishikawa, H. Miyano, T. Mizukoshi, Interaction analysis of FABP4 inhibitors by x-ray crystallography and fragment molecular orbital analysis, *ACS Med. Chem. Lett.* 7 (2016) 435–439.
- [13] H. Cai, T. Wang, Z. Yang, Z. Xu, G. Wang, H.Y. Wang, W. Zhu, K. Chen, Combined virtual screening and substructure search for discovery of novel FABP4 inhibitors, *J. Chem. Inf. Model* 57 (2017) 2329–2335.
- [14] H. Cai, Q. Liu, D. Gao, T. Wang, T. Chen, G. Yan, K. Chen, Y. Xu, H. Wang, Y. Li, W. Zhu, Novel fatty acid binding protein 4 (FABP4) inhibitors: virtual screening, synthesis and crystal structure determination, *Eur. J. Med. Chem.* 90 (2015) 241–250.
- [15] H. Cai, G. Yan, X. Zhang, O. Gorbenko, H. Wang, W. Zhu, Discovery of highly selective inhibitors of human fatty acid binding protein 4 (FABP4) by virtual screening, *Bioorg. Med. Chem. Lett.* 20 (2010) 3675–3679.
- [16] G. Floresta, O. Apirakkan, A. Rescifina, V. Abbate, Discovery of high-affinity cannabinoid receptors ligands through a 3D-QSAR ushered by scaffold-hopping analysis, *Molecules* 23 (2018).
- [17] G. Floresta, V. Pittalà, V. Sorrenti, G. Romeo, L. Salerno, A. Rescifina, Development of new HO-1 inhibitors by a thorough scaffold-hopping analysis, *Bioorg. Chem.* 81 (2018) 334–339.
- [18] G. Floresta, E. Amata, M. Dichiarà, A. Marrazzo, L. Salerno, G. Romeo, O. Prezzavento, V. Pittalà, A. Rescifina, Identification of potentially potent heme oxygenase 1 inhibitors through 3D-QSAR coupled to scaffold-hopping analysis, *Chemmedchem* 13 (2018) 1336–1342.

- [19] L. Salerno, E. Amata, G. Romeo, A. Marrazzo, O. Prezzavento, G. Floresta, V. Sorrenti, I. Barbagallo, A. Rescifina, V. Pittalà, Potholing of the hydrophobic heme oxygenase-1 western region for the search of potent and selective imidazole-based inhibitors, *Eur. J. Med. Chem.* 148 (2018) 54–62.
- [20] G. Floresta, A. Rescifina, A. Marrazzo, M. Dichiarà, V. Pistarà, V. Pittalà, O. Prezzavento, E. Amata, Hyphenated 3D-QSAR statistical model-scaffold hopping analysis for the identification of potentially potent and selective sigma-2 receptor ligands, *Eur. J. Med. Chem.* 139 (2017) 884–891.
- [21] A. Rescifina, G. Floresta, A. Marrazzo, C. Parenti, O. Prezzavento, G. Nastasi, M. Dichiarà, E. Amata, Sigma-2 receptor ligands QSAR model dataset, *Data Brief* 13 (2017) 514–535.
- [22] A. Rescifina, G. Floresta, A. Marrazzo, C. Parenti, O. Prezzavento, G. Nastasi, M. Dichiarà, E. Amata, Development of a sigma-2 receptor affinity filter through a Monte Carlo based QSAR analysis, *Eur. J. Pharm. Sci.* 106 (2017) 94–101.
- [23] K.F. Greish, L. Salerno, R. Al Zahrani, E. Amata, M.N. Modica, G. Romeo, A. Marrazzo, O. Prezzavento, V. Sorrenti, A. Rescifina, G. Floresta, S. Intagliata, V. Pittalà, Novel structural insight into inhibitors of heme oxygenase-1 (HO-1) by new imidazole-based compounds: biochemical and in vitro anticancer activity evaluation, *Molecules* 23 (2018).
- [24] G. Floresta, V. Pistarà, E. Amata, M. Dichiarà, A. Damigella, A. Marrazzo, O. Prezzavento, F. Punzo, A. Rescifina, Molecular modeling studies of pseudouridine isoxazolodinyl nucleoside analogues as potential inhibitors of the pseudouridine 5'-monophosphate glycosidase, *Chem. Biol. Drug Des.* 91 (2018) 519–525.
- [25] P. La Manna, C. Talotta, G. Floresta, M. De Rosa, A. Soriente, E. Rescifina, C. Gaeta, P. Neri, Mild Friedel-Crafts reactions inside a hexameric resorcinarene capsule: C-Cl bond activation through hydrogen bonding to bridging water molecules, *Angew. Chem. Int. Ed. Engl.* 57 (2018) 5423–5428.
- [26] P. La Manna, M. De Rosa, C. Talotta, C. Gaeta, A. Soriente, G. Floresta, A. Rescifina, P. Neri, The hexameric resorcinarene capsule as an artificial enzyme: ruling the regio and stereochemistry of a 1,3-dipolar cycloaddition between nitrones and unsaturated aldehydes, *Org. Chem. Front* 5 (2018) 827–837.
- [27] G. Floresta, A. Rescifina, Metyrapone- β -cyclodextrin supramolecular interactions inferred by complementary spectroscopic/spectrometric and computational studies, *J. Mol. Struct.* 1176 (2019) 815–824.
- [28] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, *J. Cheminform.* 3 (2011) 33.
- [29] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31 (2010) 455–461.
- [30] T. Cheeseright, M. Mackey, S. Rose, A. Vinter, Molecular field extrema as descriptors of biological activity: definition and validation, *J. Chem. Inf. Model* 46 (2006) 665–676.
- [31] T.J. Cheeseright, M.D. Mackey, R.A. Scoffin, High content pharmacophores from molecular fields: a biologically relevant method for comparing and understanding ligands, *Curr. Comput. Aided Drug Des.* 7 (2011) 190–205.
- [32] T.J. Cheeseright, M. Holm, F. Lehmann, S. Luik, M. Gottert, J.L. Melville, S. Laufer, Novel lead structures for p38 MAP kinase via FieldScreen virtual screening, *J. Med. Chem.* 52 (2009) 4200–4209.
- [33] T.J. Cheeseright, M.D. Mackey, J.L. Melville, J.G. Vinter, FieldScreen: virtual screening using molecular fields. Application to the DUD data set, *J. Chem. Inf. Model* 48 (2008) 2108–2117.
- [34] T. Cheeseright, M. Mackey, S. Rose, A. Vinter, Molecular field technology applied to virtual screening and finding the bioactive conformation, *Expert Opin. Drug Discov.* 2 (2007) 131–144.
- [35] R. Sulsky, D.R. Magnin, Y. Huang, L. Simpkins, P. Taunk, M. Patel, Y. Zhu, T.R. Stouch, D. Bassolino-Klimas, R. Parker, T. Harrity, R. Stoffel, D.S. Taylor, T.B. Lavoie, K. Kish, B.L. Jacobson, S. Sheriff, L.P. Adam, W.R. Ewing, J.A. Robl, Potent and selective biphenyl azole inhibitors of adipocyte fatty acid binding protein (aFABP), *Bioorg. Med. Chem. Lett.* 17 (2007) 3511–3515.
- [36] T. Barf, F. Lehmann, K. Hammer, S. Haile, E. Axen, C. Medina, J. Uppenberg, S. Svensson, L. Rondahl, T. Lundback, N-Benzylindole carboxylic acids: design and synthesis of potent and selective adipocyte fatty-acid binding protein (A-FABP) inhibitors, *Bioorg. Med. Chem. Lett.* 19 (2009) 1745–1748.
- [37] F. Lehmann, S. Haile, E. Axen, C. Medina, J. Uppenberg, S. Svensson, T. Lundback, L. Rondahl, T. Barf, Discovery of inhibitors of human adipocyte fatty acid-binding protein, a potential type 2 diabetes target, *Bioorg. Med. Chem. Lett.* 14 (2004) 4445–4448.
- [38] R. Ringom, E. Axen, J. Uppenberg, T. Lundback, L. Rondahl, T. Barf, Substituted benzylamino-6-(trifluoromethyl)pyrimidin-4(1H)-ones: a novel class of selective human A-FABP inhibitors, *Bioorg. Med. Chem. Lett.* 14 (2004) 4449–4452.
- [39] J.M. Luco, F.H. Ferretti, QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives, *J. Chem. Inf. Comput. Sci.* 37 (1997) 392–401.
- [40] A. Surribas, J.M. Amigo, J. Coello, J.L. Montesinos, F. Valero, S. Maspocho, Parallel factor analysis combined with PLS regression applied to the on-line monitoring of *Pichia pastoris* cultures, *Anal. Bioanal. Chem.* 385 (2006) 1281–1288.
- [41] K. Tuppurainen, S.P. Korhonen, J. Ruuskanen, Performance of multicomponent self-organizing regression (MCSOR) in QSAR, QSPR, and multivariate calibration: comparison with partial least-squares (PLS) and validation with large external data sets, *SAR QSAR Environ. Res.* 17 (2006) 549–561.
- [42] G. Palermo, P. Piraino, H.D. Zucht, Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data, *Adv. Appl. Bioinform. Chem.* 2 (2009) 57–70.
- [43] T. Asadollahi, S. Dadfarnia, A.M. Shabani, J.B. Ghasemi, M. Sarkhosh, QSAR models for CXCR2 receptor antagonists based on the genetic algorithm for data preprocessing prior to application of the PLS linear regression method and design of the new compounds using in silico virtual screening, *Molecules* 16 (2011) 1928–1955.
- [44] A. Golbraikh, A. Tropsha, Beware of q²! *J. Mol. Graph Model* 20 (2002) 269–276.
- [45] J.J. Irwin, B.K. Shoichet, ZINC—a free database of commercially available compounds for virtual screening, *J. Chem. Inf. Model* 45 (2005) 177–182.
- [46] A.P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Kruger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J.P. Overington, The ChEMBL bioactivity database: an update, *Nucleic Acids Res.* 42 (2014) D1083–1090.