Bioorganic & Medicinal Chemistry 19 (2011) 5903–5911

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

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

Modeling the molecular basis for α4β1 integrin antagonism

Oliver E. Hutt [a], Simon Saubern [a], David A. Winkler [a,b,*]

[a] Ian Wark Laboratories, CSIRO Materials Science and Engineering, Private Bag 10, Clayton South MDC 3169, Australia
[b] Monash Institute of Pharmaceutical Sciences, 381 Royal Parade, Parkville 3052, Australia

ARTICLE INFO

Article history:
Received 9 June 2011
Revised 29 July 2011
Accepted 5 August 2011
Available online 16 August 2011

Keywords:
Integrin α4β1
QSAR
Antagonism
Molecular field analysis

ABSTRACT

We report a 3D QSAR study of almost 300 structurally diverse small molecule antagonists of the integrin α4β1 whose biological activity spans six orders of magnitude. The alignment of the molecules was based on the conformation of a structurally related ligand bound to the αIIBβ3 and αvβ3 integrins in X-ray crystallographic studies. The molecular field method, CoMSIA, was used to generate the 3D QSAR models. The resulting models showed that the lipophilic properties were the most important, with hydrogen bond donor and steric properties less relevant. The models were highly significant (r^2 = 0.89, $q2$(LOO) = 0.67, r^2 (test set) = 0.76), and could make robust predictions of the data (SEE = 0.46, SEP = 0.78, SEP (test set) = 0.66). We predicted the antagonist activities of a further ten compounds with useful accuracy. The model appears capable of predicting α4β1 integrin antagonist activity to within a factor of five for compounds within its domain of applicability. The implications for design of improved integrin antagonists will be discussed.

1. Introduction

Integrins are a family of heterodimeric, transmembrane, cell adhesion receptors whose biological role is to connect cells to the extracellular matrix, and to mediate specialized cell–cell interactions. In mammals the family is comprised of 19 α subunits and nine β subunits, which combine to form at least 25 αβ heterodimeric members.[1] Integrins regulate a diverse array of cellular functions crucial to the initiation, progression, and metastasis of solid tumors, cell adhesion and migration, inflammation, and tissue remodeling. For example, the αvβ3 integrin is a regulator of disease pathology associated with cancer, osteoporosis, and rheumatoid arthritis. However, the αvβ3 integrin is also expressed in endothelial cells and mediates survival of angiogenic vessels. In osteoclasts αvβ3 integrin supports adhesion to bone matrix, while in macrophages αvβ3 integrins play a role in modulating pro-inflammatory cytokine expression.[2] These diverse functions are regulated in part through the interaction with an equally diverse range of integrin binding proteins.[3] Sequence analysis has identified a number of integrin binding motifs within these proteins. The most ubiquitous motif, RGD, is recognized by α3β1, α5β1, α8β1, αvβ1, αvβ3, αvβ5, αvβ6, αIIbβ3 integrins and accordingly has been the focus of significant scientific attention.[4] The LDV motif of the CS-1 repeat of fibrinogen, which binds to the α4β1 integrin (Very-Late-Antigen-4; VLA-4), has also been focus of significant attention due to its role in inflammation. As a consequence, many potent non-peptide small molecule antagonists have been reported and in some instances advanced to clinical trials.[5] Many of these compounds have been disclosed in the non-patent literature and recently have been compiled, along with their biological activities, in the ChEMBL database.[6] Published 3D QSAR models for integrin antagonism are few, involve relatively small data sets, and contain relatively low molecular diversity and small range of biological responses.[7–9] Given the importance of integrins in many disease states, it is surprising that no comprehensive modeling studies have been published on the very large number of small molecule and peptide antagonists of integrins now available. Here we describe a KNIME workflow[10] for analysis and retrieval of small molecule integrin antagonists from the ChEMBL database,[11] and the subsequent development of a large, chemically diverse, quantitative 3D QSAR model of integrin α4β1 antagonism. Our aims were:

1. To develop a more general, 3D QSAR model of α4β1 antagonism;
2. To assess whether molecular alignments similar to those observed in X-ray structures of related integrins were consistent with α4β1 antagonism;
3. To understand the molecular requirements for α4β1 antagonism in a larger, more chemically diverse set of molecules than have previously been studied.

2. Results and discussion

The PLS analysis showed that seven principal components were optimum for the molecular field QSAR models. In the CoMSIA models, various combinations of field were tried, but all five fields were

* Corresponding author. Tel.: +61 3 9545 2477; fax: +61 3 9545 2446.
E-mail address: dave.winkler@csiro.au (D.A. Winkler).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.08.011

found to make significant contributions to the model. The electrostatic and lipophilic fields appeared to be the most important. The results are summarized in Table 1. SEP is the standard error of prediction, a measure of the uncertainty of prediction. The SEE is the standard error of estimation, a measure of how well the model can predict the activities of compounds used to generate the model.

Both CoMFA and CoMSIA methods generated similar results, with estimated prediction errors being approximately 0.8 logs (i.e., within a factor of six in IC_{50} values). The CoMFA models were more difficult to interpret so are not discussed here. The training set activities could be predicted with approximately 0.5 logs (i.e., within a factor of three in IC_{50}). The models could explain about 80% of the variance in the training data and 65% of the variance in the predicted test data. Given the origin of the data, it is very likely experimental error is responsible for the remaining 30–40% of the variance. The predicted versus measured pIC_{50} ($-\log IC_{50}$) values for compounds in the *external test set* from the CoMSIA model are shown in Figure 1.

As an additional validation, we used the model to also predict activities of members of the larger 1893 data set extracted by the workflow that were not chosen for modeling. Because the large amount of conformational analysis, chirality checking, and alignment makes it infeasible to calculated pIC_{50} values for all compounds, we selected only 10 additional examples for prediction. These included chemotypes where the carboxylic acid was replaced by a tetrazole, the pyrrolidine was replaced by an acridine, and alkoxy chains replaced the terminal chains. The results are summarized in Table 2.

Table 1
Summary of statistics for CoMSIA models for α4β1 antagonists

SEP	0.78
LOO q^2	0.67
Number of PCs	7
SEE	0.46
r^2	0.89
F	293
Number of compounds	272
Contributions of steric, electrostatic, lipophilic, donor, acceptor fields%	13, 27, 34, 10, 16
Test set SEP	0.66
Test set r^2	0.76

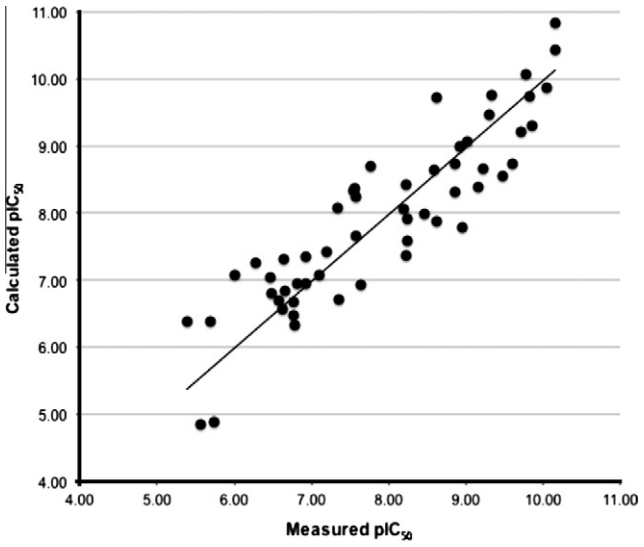
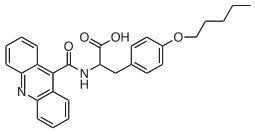
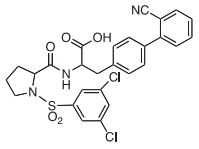
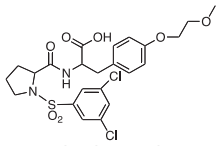
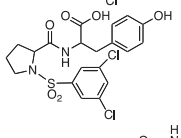
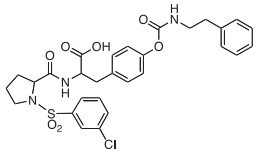
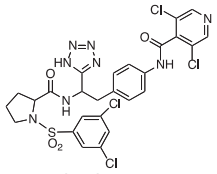
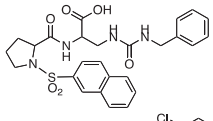
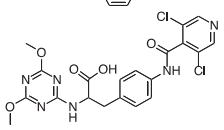
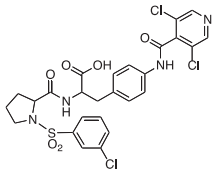
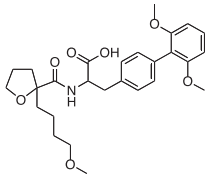


Figure 1. Calculated versus observed pIC_{50} for α4β1 antagonism for test set compounds not used in generating model.

The 3D QSAR models generate feature maps that summarize pictorially the requirements for antagonist activity. These are illustrated for a CoMSIA model in Figures 2–6.

Figure 2 shows two molecular steric fields. The field map shows that steric bulk attached to the short side chain (often an aromatic ring coupled via a sulfonyl moiety) enhances the antagonist effect of the ligands. The potent antagonist **1048** has a chloro-substituted

Table 2
Predicted antagonist pIC_{50} values for compounds from 1893 data set

Structure	Predicted α4β1 antagonist pIC_{50} (error)
	6.53 (−0.30)
	9.50 (−0.17)
	9.28 (+0.27)
	8.58 (+0.88)
	6.52 (+1.22)
	8.44 (−1.22)
	6.04 (−0.68)
	6.31 (−0.91)
	10.20 (−0.20)
	9.52 (−0.63)

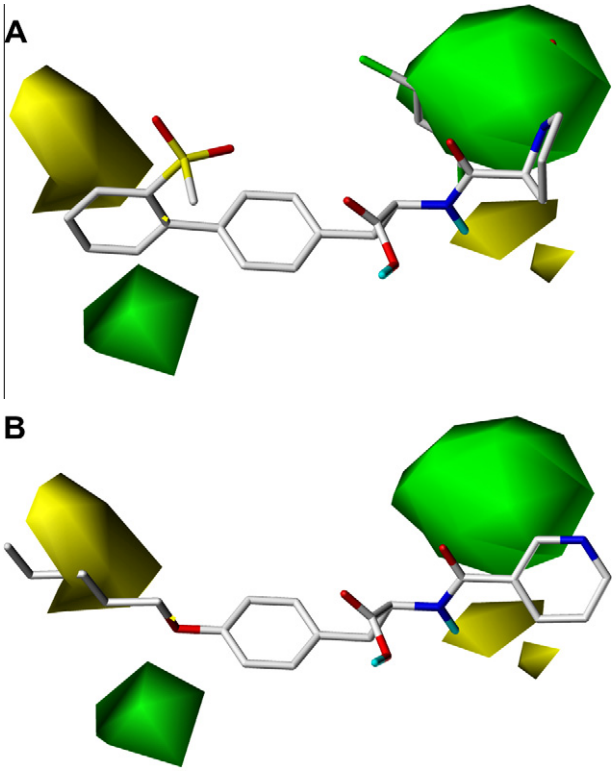


Figure 2. CoMSIA steric molecular field mapped onto a (A) potent integrin antagonist (**1048**, IC_{50} = 0.1 nM, pIC_{50} = 10.00) and (B) weak integrin antagonist (**1021**, IC_{50} = 38,000 nM, pIC_{50} = 5.42). The green regions denote where steric bulk increases the integrin antagonist activity, and the yellow regions denote where steric bulk decreases it.

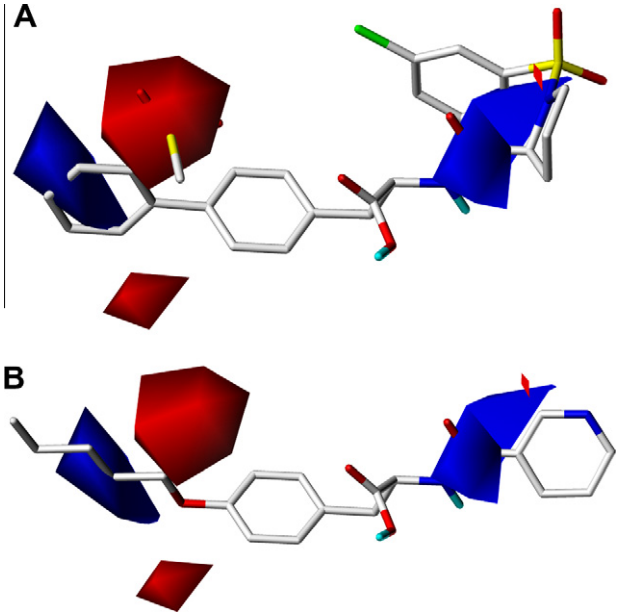


Figure 3. CoMSIA electrostatic molecular field mapped onto a (A) potent integrin antagonist (**1048**, IC_{50} = 0.1 nM, pIC_{50} = 10.00) and (B) weak integrin antagonist (**1021**, IC_{50} = 38,000 nM, pIC_{50} = 5.42). The blue regions denote where partial positive charge increases the integrin antagonism, and the red regions denote where partial negative charge increases integrin antagonism.

phenyl ring in this position, while the weakly antagonistic ligand **1021** lacks a substituent in this region. The field map also shows

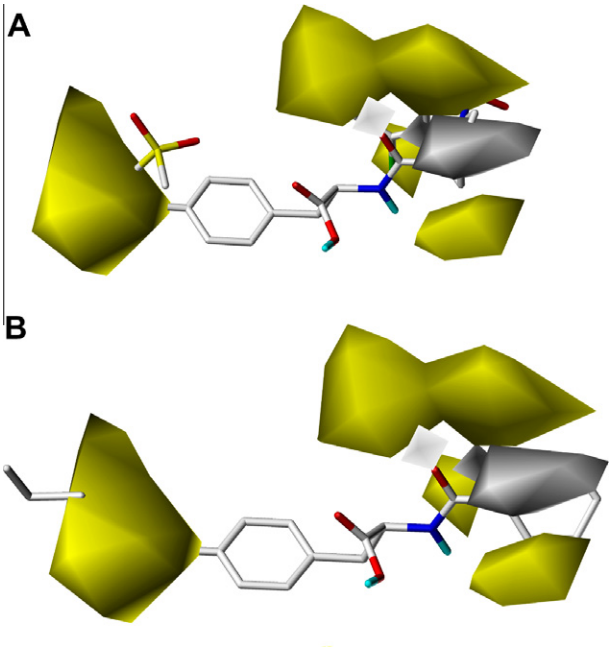


Figure 4. CoMSIA lipophilic molecular field mapped onto a (A) potent integrin antagonist (**1048**, IC_{50} = 0.1 nM, pIC_{50} = 10.00) and (B) weak integrin antagonist (**1021**, IC_{50} = 38,000 nM, pIC_{50} = 5.42). The yellow regions denote where lipophilic regions of the molecule increase integrin antagonism, and the white regions where lipophilic moieties in the molecule decrease integrin antagonism.

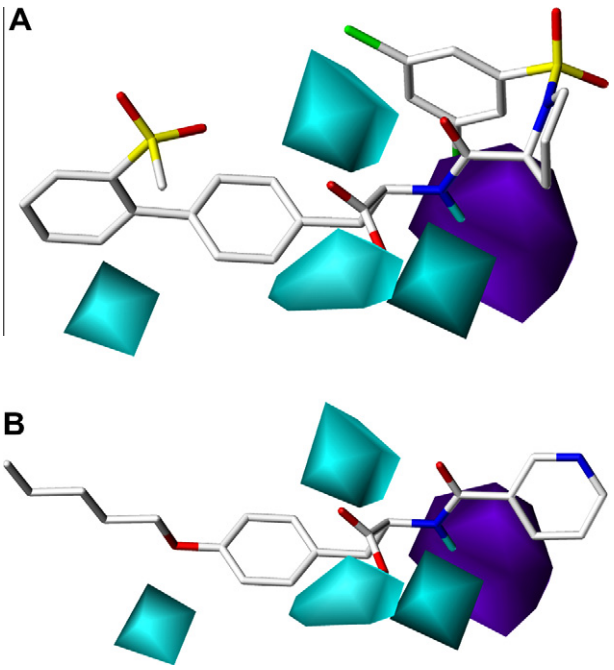


Figure 5. CoMSIA donor molecular field mapped onto a (A) potent integrin antagonist (**1048**, IC_{50} = 0.1 nM, pIC_{50} = 10.00) and (B) weak integrin antagonist (**1021**, IC_{50} = 38,000 nM, pIC_{50} = 5.42). The cyan regions denote where the presence of a hydrogen bond donor increases integrin antagonism and the purple region denote regions where donors decrease integrin antagonism.

that steric bulk corresponding to a large substituent on the 'tyrosine-like' aromatic ring is detrimental to antagonist activity. Weak antagonist **1021** contains an n-pentyl chain in this unfavorable region, while the potent **1048** antagonist does not.

Figure 3 shows the contribution of molecular electrostatic fields to the integrin antagonism. Electrostatic molecular field maps are

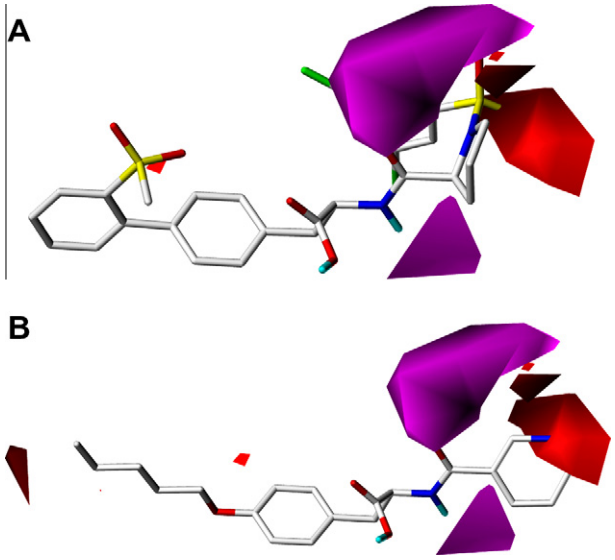


Figure 6. CoMSIA acceptor molecular field mapped onto a (A) potent integrin antagonist (**1048**, IC_{50} = 0.1 nM, pIC_{50} = 10.00) and (B) weak integrin antagonist (**1021**, IC_{50} = 38,000 nM, pIC_{50} = 5.42). The magenta regions denote where the presence of a hydrogen bond acceptor increases the integrin antagonism, and red regions denote where acceptors decrease the integrin antagonism.

generally difficult to interpret, and they represent properties that correlate with several others, such as lipophilic, donor, and acceptor. The blue field map shows that positive partial charge near the amide linker (largely associated with the amide carbon atom) enhances the antagonist activity of the ligands. The large red molecular field map, corresponding to regions where partial negative charge is favorable for antagonism, is occupied by the methanesulfonyl substituent in the potent antagonist **1048**, but is not occupied by substituents with similar properties on the weak ligand **1021**.

Figure 4 shows the lipophilic molecular fields, identifying how the hydrophilic/hydrophobic properties of substituents influence antagonist activity. The yellow region shows where lipophilic substituents favor activity, reflected by the high antagonist activity of arylsulfonyl substituents containing halogen substituents in Figure 4A. Strong antagonists often have electronegative atom in the heterocyclic ring that occupies the white region, showing that hydrophilic substituents enhance antagonist activity. The weak antagonist shown in Figure 7B does not contain such a heteroatom, so the ring occupying the white region is unfavorably lipophilic. The large yellow lipophilic field at the left-hand side of Figure 4A and B shows that lipophilic substituents on this 'tyrosine-like' ring favor antagonism.

Figure 5 shows the hydrogen bond donor molecular fields, identifying the substituents on the antagonists whose donor properties modulate antagonism. The main favorable regions, denoted by cyan field regions, occur near the amide NH, the adjacent COOH (modeled in the neutral form), and near the end of the short arm of the pharmacophore (see bold atoms in **1**). This molecular field map is difficult to interpret, as neither of the strong or weak ligands illustrated in Figure 5 make significantly different interactions with these field regions. As the 3D QSAR model involve a complex interplay between all five field maps, it is unsurprising that not all single feature maps can be interpreted easily.

The hydrogen bond acceptor properties of antagonists are summarized in Figure 6. The pink field region indicates that the presence of a hydrogen bond acceptor (such as the amide carbonyl group or a sulfonyl moiety) enhances antagonist activity. Although both weak and potent ligands shown in Figure 6 have amide linkers, the potent ligand contains a sulfonyl group with two hydrogen bond accepting oxygen atoms in this region. The red region, where hydrogen bond donor activity reduces antagonism, is occupied by a heterocyclic nitrogen atom (a good acceptor) in the weak ligand.

There are several relatively limited QSAR studies of small molecule antagonists of α4β1 integrins in the literature, none of which can be compared directly to our study. Singh et al.[12] used database screening to identify possible mimics of the LDV motif from fibronectin. Their binding conformation was based on the X-ray crystal structure of the integrin binding region of VCAM-1. They used a pharmacophore hypothesis generation method (CATALYST) to identify the important binding features in the resulting hits. While they could provide qualitative explanation of the SAR, they did not generate quantitative models. This group also published a 3D QSAR model of 29 small molecule LDV mimetics, spanning three orders of magnitude in activity, derived using the CoMFA molecular field-based QSAR modeling method.[9] Their model was again based on the LDV conformation alignment rule, and resulted in a cross-validated q^2 value of around 0.7. The steric field dominated. The SAR suggested that the Leu-mimetic moiety in the antagonists was relatively unimportant, as substitution by other amino acids with different physicochemical properties, or with other functional groups generally had a minor effect on activities. As the molecular alignment and chemotype of these antagonists was different those in our study, the molecular field property maps could not be compared. Subsequently, Khandelwal et al.[7] reported a 3D CoMSIA and CoMFA study of 25 tetrahydrofuroylphenylalanine analogs as α4β1 antagonists. They reported a relatively poor QSAR model, probably because the biological activity of the 25 analogs was very similar, ranging from 0.3 to 23 nM. The cross-validated q^2 values were in the range 0.3–0.5, marginally predictive, with the values near 0.5 requiring a relatively large number of PLS components relative to the number of compounds in the training set. The training set models had very high r^2 values between 0.95 and 0.99 and this, combined with the low q^2 values, suggests models with only modest performance. This was borne out by the inability of the best models to predict an 11 compound external test set. The r^2_{pred} for this set was essentially zero for most models, the best value being 0.19. More recently, Macchiarulo et al.[8] used 3D QSAR methods to model the α4β1 (VLA-4) antagonism of one hundred and 28 phenylalanine analogs spanning three orders of magnitude in activity. A molecular field grid, PLS, and a genetic algorithm were used to generate a 3D QSAR model. The molecular diversity of the data set was relatively limited, and most of the data involved substitutions on one or two rings in the compounds. The resulting QSAR model had an r^2 value of 0.63 and cross-validated q^2 value of 0.51 despite the relatively high similarity of the antagonists. Omitting large outliers improved the model slightly. Finally, a small 2D QSAR study was reported recently by Bhargava et al.[13] They used molecular descriptors to model the integrin inhibition of 31 pyrazinyl phenylalanine analogs. The activities ranged from 19 to 6800 nM and structural variations consisted of substitution at the terminal nitrogen of piperazine, and on the distal terminal amide moiety. They obtained a linear, four-descriptor model that had a training set r^2 value of 0.86, cross-validated q^2 value of 0.76 but only a modest test set r^2 value of 0.42. The most important descriptors were two electrotopological indices related to the number of chlorine atoms and the number of aromatic nitrogen atoms, another topological index related to the number of paths in the molecular graph of path length four, and the calculated log *P* values.

3. Conclusions

In summary, we report the first chemically diverse, large data set 3D QSAR model for α4β1 integrin antagonism. The modeling

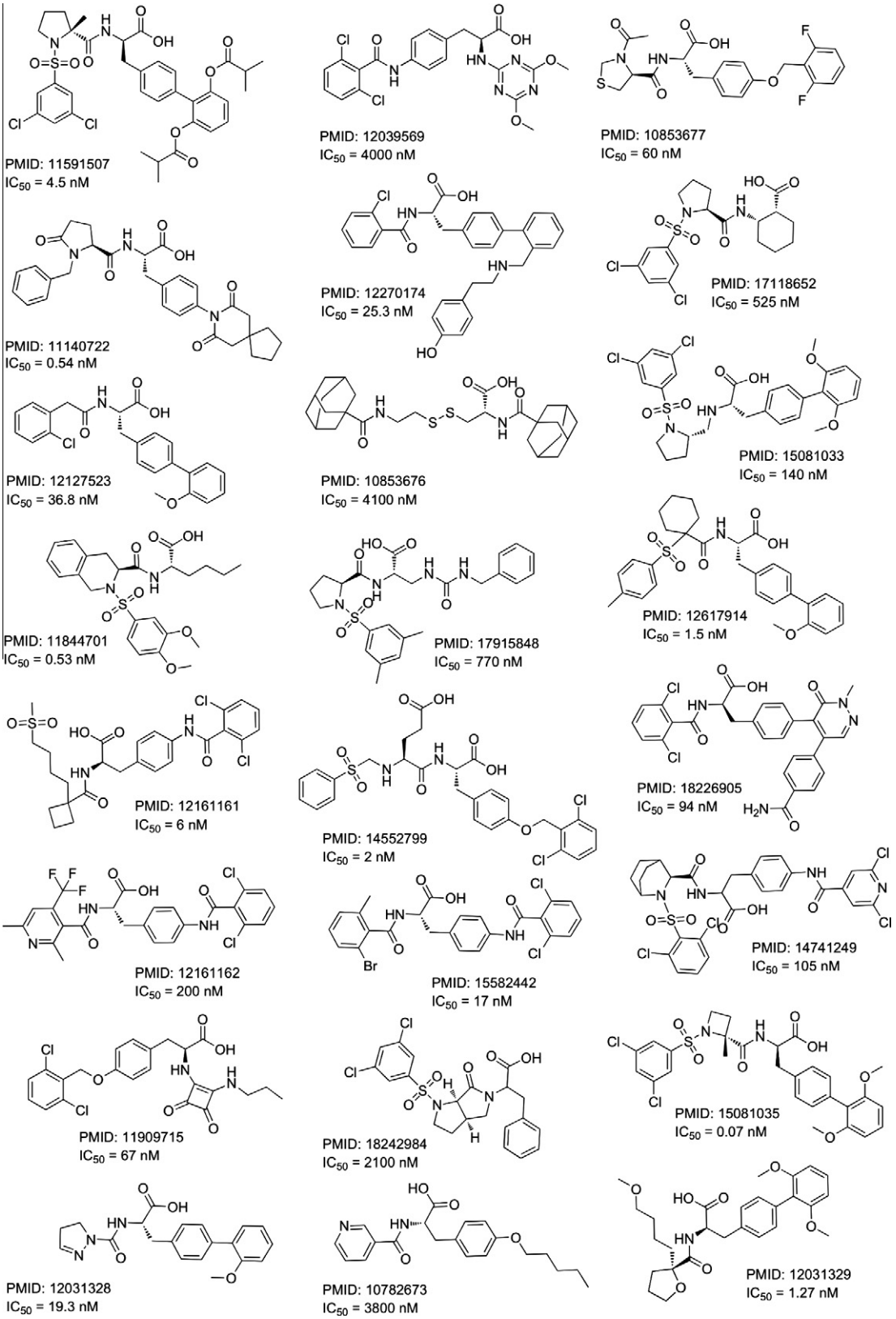


Figure 7. Representative chemical classes of compounds used in 3D QSAR model.

study shows that good, predictive QSAR models can be built from data on the α4β1 antagonist properties of small molecules. The data set size, diversity, and predictive power was substantially higher than those reported in previous classical QSAR models of this important biological response. The 3D QSAR model of integrin antagonists based on an alignment from the X-ray structure was quite robust, statistically significant, and predictive. This adds substantial support to the proposal that this conformation is

biologically relevant. This study also illustrates the value of using workflows to rapidly generate large, chemically diverse data sets for QSAR analysis. The model provide additional understanding of the relationship between molecular structure and integrin antagonist activity It will allow α4β1 antagonism to be estimated for a larger range of chemotypes than the previous QSAR models.

3.1. Method section

3D QSAR models were generated for a large, chemically diverse set of integrin antagonists reported in the literature. Given the very large number of compounds for which integrin antagonism has been reported, we employed the following strategy to develop a diverse and comprehensive QSAR model for α4β1 antagonism. We first used the publicly accessible ChEMBL database and a KNIME workflow to extract 1893 compounds that exhibited human α4β1 antagonism. Some of these compounds were cyclic peptides, and others appeared to contain a substantially different binding motif to the rest and were discarded.

Generating a 3D QSAR model for such a large number of compounds, given the requirements for chirality checking, conformational adjustment, and molecular alignment is not feasible or necessary. We selected a random subset of 272 compounds, which spanned a wide range of chemotypes and almost six orders of magnitude of biological activity, for the modeling study (Fig. 7). Our rationale was that this should capture the key 3D binding requirements for antagonism, and the addition of the remainder of the compounds from the workflow would only refine, not substantially change, the outcomes of the QSAR modeling.

3.2. KNIME workflow

Version 1 of the ChEMBL database[11] was indexed using the Bingo data cartridge[14] in an Oracle 10.2 database. The KNIME 2.1[10,15] workflow (Fig. 8), with the addition of the Oracle supplied jdbc database driver, was used to extract the data from the ChEMBL database using the following step-wise filtering criteria. Chemical structures were retrieved as SMILES strings and therefore the RDKit nodes present in the 'community contributions' collection were used for structure display and export.[16,17] Unfortunately, these conversions do not preserve the chirality of the molecules.

For this study, the following search criteria where used:

- Keywords containing the phrase 'integrin'.
- Organism = 'Homo sapiens'.
- Assay names containing the phrase 'Integrin alpha-4'.
- Assay values of type 'IC_{50}'.
- Removal of assay results with the phrase 'beta-7' in the assay name.

This search provided 1893 structures suitable for our modeling purposes. The workflow is available from Supplementary data.

3.3. Biological activity

The α4β1 antagonism was reported as IC_{50} across a range of 50 pM to 26 μM. The activity was converted to pIC_{50} values ($-\log IC_{50}$) that spanned the range 4.5–10.3. Several types of biological assays were used to measure human α4β1 antagonism. These assays are largely evaluating the ability of the antagonist to block the interaction between VCAM-1 and α4β1 integrins expressed in different cell line (Ramos, RPMI-8866, Jurkat, endothelial, HL-60) in different assay formats and in the presence of different divalent counter ions (Ca^{2+}, Mg^{2+}, Mn^{2+}) or fetal bovine serum. Two assays alternatively inhibited the fibronectin CS-1 peptide: α4β1 integrin interaction. Sufficient numbers of compounds in the data set had been measured in more than one assay to suggest that the results from different assay systems were comparable and the data could be aggregated.

3.4. 3D QSAR modeling

The 272 diverse structures and corresponding integrin antagonist activities were imported from the SD file generated by the workflow.[18–54] The chiralities of all stereo centers were checked using the modeling capabilities of Sybyl version 8.1 (Tripos Inc.). Structures were energy minimized using the Tripos force field and Gasteiger–Hückel atom charges. The compounds in the data were relatively flexible, and the question of the appropriate conformations and molecular alignments for the 3D QSAR modeling needed to be addressed. As *no X-ray structures of ligands bound to the human α4β1 integrin receptor* have been reported, this issue was problematic. Although a number of low energy conformations may be relevant to the binding of small molecules to the α4β1 integrin, we chose the conformation that most closely matched that described by Elliot et al. (Fig. 9) for the ligand Merck L739758

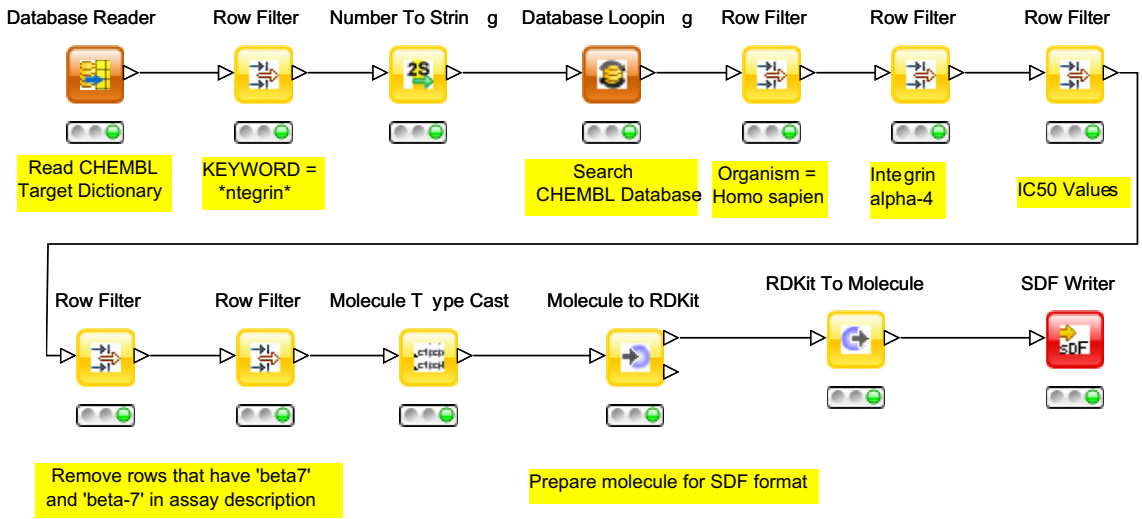


Figure 8. KNIME workflow developed to extract dataset from CHEMBL.

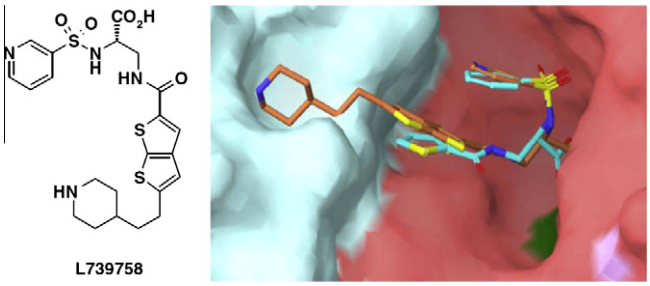
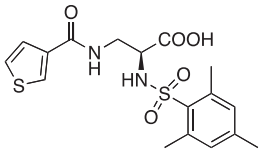


Figure 9. Overlay of Merck compound L739758 in orange (RCSB code: 1TY7) and potential binding mode of Elliot compound 18 (cyan) in αvβ3 crystal (RCSB code: 1L5G), showing molecular surface of the protein. Used with permission.

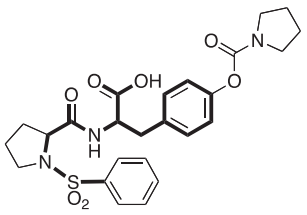
and their compound **18** ((*S*)-3-(thiophene-3-carboxamido)-2-(2,4,6-trimethylphenylsulfon-amido)propanoic acid).[55]



Elliot compound 18.

This putative bioactive conformation is also consistent with the X-ray structure of ligands with very similar chemotypes to those in our data set bound to the αvβ1 receptor (RCSB refcodes 1TY7, 1L5G).[56,57] While this conformation may not be the same as that adopted by the compounds reported in previous QSAR models of α4β1 antagonism, some of these models used binding chemotypes and conformations based on different binding motifs such as LDV. Given the consistency of our alignment within the more highly constrained chemical classes in our data set (e.g., the *N*-sulfonyl-pyrrolidines) and the published X-ray structures of related ligands, it is a reasonable starting point for 3D QSAR modeling. Clearly, the least flexible active antagonists could not adopt the same range of low energy conformations as the more flexible representatives. This resulted in essentially only one conformation and alignment being consistently available for the entire data set.

Assuming a relatively extended conformation for the amide arm of the molecules, and a conformation similar to that proposed by Elliot et al., we set the appropriate stereochemistry, adjusted the conformations to be in an energy minimum close to that of the ligand in the X-ray structure, and aligned the 272 compounds using the highly active and relatively simple antagonist **1**.



1

The alignment is critical for 3D QSAR modeling and is illustrated in Figure 10. We aligned the molecules in the data set to optimize the overlap between the amide backbone and attached carboxylic acid (or their bioisosteric equivalents in some molecules), the attached 'tyrosine-like' aromatic moiety (or equivalent), and the lipophilic group attached to the sulfonamide (or isosteric) equivalent. This is illustrated by the bold bonds in template structure **1**. Essentially, the backbone atoms shown in bold (or their topologically equivalents where relevant) were aligned. Where multiple conformations of the substituents attached to the 4 position of the bold phenyl ring in **1** were possible, we used low energy conformations that best matched the conformation of the pyrrolidone moiety. Likewise, low energy conformations for lipophilic functionality attached to the bolded sulfonamide group (or topological equivalent) were aligned with the phenyl group of **1**. Using this alignment rule, we could unambiguously align all compounds in the data set (Fig. 10). While it is possible that alternative conformations and alignment rules would also generate valid 3D QSAR models, we assume that only an alignment close to the biologically relevant one will yield a high quality model for a structurally diverse set of ligands.

The QSAR modeling approach adopted was molecular field analysis. Both the comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) methods were used.[58,59] These differ in the types of molecular property

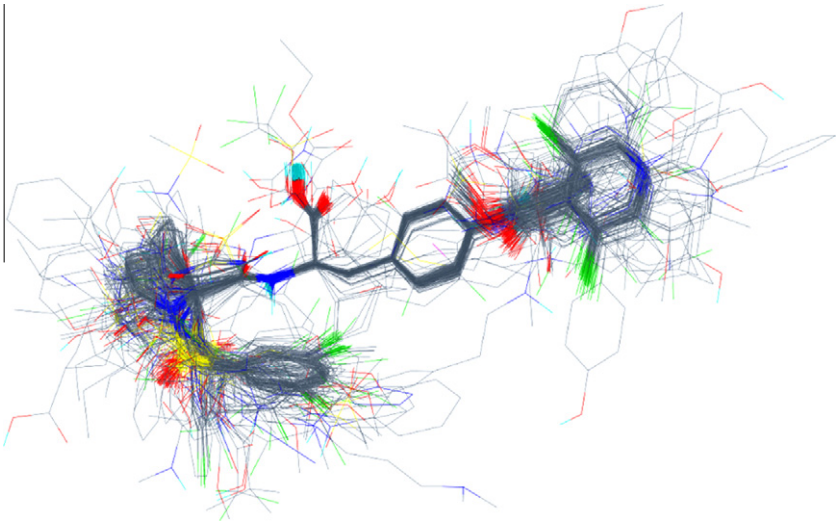


Figure 10. Alignment of 272 α4β1 antagonists used in the 3D QSAR studies. The structures are hydrogen suppressed for clarity.

fields used, and how the interactions between probe atoms and atoms in the molecules were treated. CoMFA employs two molecular fields (steric and electrostatic) and Lennard-Jones and Coulomb potentials for mapping interaction fields. CoMSIA employs five fields (steric, electrostatic, lipophilic, hydrogen bond donor and acceptor) and a Gaussian kernel to represent molecular interaction fields. They both employ a probe atom consisting of an sp^3 hybridized carbon atom with a +1 charge. Both methods use the partial least squares method (PLS), with leave-one-out (LOO) cross validation to determine the correct number of principal components to use in the model. LOO omits each molecule in turn from the data set, generates QSAR model using the remaining molecules, and then predicts the activity of the omitted molecule.

The models employed the five CoMSIA molecular fields (steric, electrostatic, lipophilic, donor, and acceptor) and were calculated with Sybyl8.1 on Dell Precision workstation running Red Hat Linux. The CoMFA models were similar to, but not quite as statistically significant as the CoMSIA models and more difficult to interpret so are not reported here. Both standard leave-one-out (LOO) cross validation and, somewhat unusually for a COMSIA study, an *independent test set* were used to validate models and assess their predictivity. A random selection of 20% of the compounds was withheld in a test set, and the remaining 80% used to generate the model subsequently used to predict the activities of the test set. Using a set of compounds that the model has not seen is a better gauge of the predictive ability of the model that LOO cross validation (which tends to give overly optimistic estimates of model predictivity).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.08.011.

References and notes

1. Humphries, M. J. *Biochem. Soc. Trans.* **2000**, *28*, 311.
2. Sheppard, D. *Physiol. Rev.* **2003**, *83*, 673.
3. Plow, E. F.; Haas, T. K.; Zhang, L.; Loftus, J.; Smith, J. W. *J. Biol. Chem.* **2000**, *275*, 21785.
4. Meyer, A.; Auemheimer, J.; Modlinger, A.; Kessler, H. *Curr. Pharm. Des.* **2006**, *12*, 2723.
5. Tilley, J. W. *Expert Opin. Ther. Pat.* **2008**, *18*, 841.
6. Warr, W. *J. Comput. Aided Mol. Des.* **2009**, *23*, 195.
7. Khandelwal, A.; Narayanan, R.; Gopalakrishnan, B. *Bioorg. Med. Chem.* **2003**, *11*, 4235.
8. Macchiarulo, A.; Costantino, G.; Meniconi, M.; Pleban, K.; Ecker, G.; Bellocchi, D.; Pellicciari, R. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1829.
9. Singh, J.; van Vlijmen, H.; Lee, W. C.; Liao, Y.; Lin, K. C.; Ateeq, H.; Cuervo, J.; Zimmerman, C.; Hammond, C.; Karpusas, M.; Palmer, R.; Chattopadhyay, T.; Adams, S. P. *J. Comput. Aided Mol. Des.* **2002**, *16*, 201.
10. Berthold, M. R.; Cebron, N.; Dill, F.; Di Fatta, G.; Gabriel, T. R.; Georg, F.; Meinl, T.; Ohl, P.; Sieb, C.; Wiswedel, B. *4th International Industrial Simulation Conference 2006*; 2006, 58.
11. ChEMBL data is from http://www.ebi.ac.uk/chembl - version 1 released 28 September 2009. ('CC Attribution-ShareAlike 3.0 Unported license').
12. Singh, J.; van Vlijmen, H.; Liao, Y. S.; Lee, W. C.; Cornebise, M.; Harris, M.; Shu, I. H.; Gill, A.; Cuervo, J. H.; Abraham, W. M.; Adams, S. P. *J. Med. Chem.* **2002**, *45*, 2988.
13. Bhargava, D.; Karthikeyan, C.; Moorthy, N. S. H. N.; Trivedi, P. *Med. Chem.* **2009**, *5*, 446.
14. SciTouch http://scitouch.net/bingo - version 1.4 ("GNU GPL version 3")
15. Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*: Pringer, 2007
16. KNIME Community Contributions' (see http://tech.knime.org/community/rdkit).
17. Landrum, G. RDKit, http://www.rdkit.org/
18. Archibald, S. C.; Head, J. C.; Gozzard, N.; Howat, D. W.; Parton, T. A.; Porter, J. R.; Robinson, M. K.; Shock, A.; Warrellow, G. J.; Abraham, W. M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 997.
19. Archibald, S. C.; Head, J. C.; Linsley, J. M.; Porter, J. R.; Robinson, M. K.; Shock, A.; Warrellow, G. J. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 993.
20. Castanedo, G. M.; Sailes, F. C.; Dubree, N. J. P.; Nicholas, J. B.; Caris, L.; Clark, K.; Keating, S. M.; Beresini, M. H.; Chiu, H.; Fong, S.; Marsters, J. C., Jr.; Jackson, D. Y.; Sutherlin, D. P. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2913.
21. Chang, L. L.; Truong, Q.; Doss, G. A.; MacCoss, M.; Lyons, K.; McCauley, E.; Mumford, R.; Forrest, G.; Vincent, S.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 597.
22. Chang, L. L.; Yang, G. X.; McCauley, E.; Mumford, R. A.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1688.
23. Chen, L.; Tilley, J.; Trilles, R. V.; Yun, W.; Fry, D.; Cook, C.; Rowan, K.; Schwinge, V.; Campbell, R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 137.
24. Chen, L.; Tilley, J. W.; Guthrie, R. W.; Mennona, F.; Huang, T. N.; Kaplan, G.; Trilles, R.; Miklowski, D.; Huby, N.; Schwinge, V.; Wolitzky, B.; Rowan, K. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 729.
25. Chen, L.; Tilley, J. W.; Huang, T. N.; Miklowski, D.; Trilles, R.; Guthrie, R. W.; Luk, K.; Hanglow, A.; Rowan, K.; Schwinge, V.; Wolitzky, B. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 725.
26. Chen, L.; Trilles, R.; Miklowski, D.; Huang, T.-N.; Fry, D.; Campbell, R.; Rowan, K.; Schwinge, V.; Tilley, J. W. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1679.
27. Choi, S.; Vilaire, G.; Marcinkiewicz, C.; Winkler, J. D.; Bennett, J. S.; DeGrado, W. F. *J. Med. Chem.* **2007**, *50*, 5457.
28. de Laszlo, S. E.; Li, B.; McCauley, E.; Van, R. G.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 685.
29. Doherty, G. A.; Kamenecka, T.; McCauley, E.; Van, R. G.; Mumford, R. A.; Tong, S.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 729.
30. Doherty, G. A.; Yang, G. X.; Borges, E.; Chang, L. L.; MacCoss, M.; Tong, S.; Kidambi, U.; Egger, L. A.; McCauley, E.; Van, R. G.; Mumford, R. A.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1501.
31. Doherty, G. A.; Yang, G. X.; Borges, E.; Tong, S.; McCauley, E. D.; Treonz, K. M.; Van, R. G.; Pacholok, S.; Si, Q.; Koo, G. C.; Shah, K.; Mumford, R. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1891.
32. Dyatkin, A. B.; Hoekstra, W. J.; Kinney, W. A.; Kontoyianni, M.; Santulli, R. J.; Kimball, E. S.; Fisher, M. C.; Carolyn, F. M.; Prouty, S. M.; Abraham, W. M.; de, G. L.; Andrade-Gordon, P.; Hlasta, D. J.; He, W.; Hornby, P. J.; Damiano, B. P.; Maryanoff, B. E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 591.
33. Gong, Y.; Barbay, J. K.; Dyatkin, A. B.; Miskowski, T. A.; Kimball, E. S.; Prouty, S. M.; Fisher, M. C.; Santulli, R. J.; Schneider, C. R.; Wallace, N. H.; Ballentine, S. A.; Hageman, W. E.; Masucci, J. A.; Maryanoff, B. E.; Damiano, B. P.; Andrade-Gordon, P.; Hlasta, D. J.; Hornby, P. J.; He, W. *J. Med. Chem.* **2006**, *49*, 3402.
34. Gong, Y.; Kent, B. J.; Kimball, E. S.; Santulli, R. J.; Carolyn, F. M.; Dyatkin, A. B.; Miskowski, T. A.; Hornby, P. J.; He, W. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1331.
35. Gutteridge, C. E.; de, L. S. E.; Kamenecka, T. M.; McCauley, E.; van, R. G.; Mumford, R. A.; Kidambi, U.; Egger, L. A.; Tong, S.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 885.
36. Hagmann, W. K.; Durette, P. L.; Lanza, T.; Kevin, N. J.; de, L. S. E.; Kopka, I. E.; Young, D.; Magriotis, P. A.; Li, B.; Lin, L. S.; Yang, G.; Kamenecka, T.; Chang, L. L.; Wilson, J.; MacCoss, M.; Mills, S. G.; Van, R. G.; McCauley, E.; Egger, L. A.; Kidambi, U.; Lyons, K.; Vincent, S.; Stearns, R.; Colletti, A.; Teffera, J.; Tong, S.; Fenyk-Melody, J.; Owens, K.; Levorse, D.; Kim, P.; Schmidt, J. A.; Mumford, R. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2709.
37. Hoshina, Y.; Ikegami, S.; Okuyama, A.; Fukui, H.; Inoguchi, K.; Maruyama, T.; Fujimoto, K.; Matsumura, Y.; Aoyama, A.; Harada, T.; Tanaka, H.; Nakamura, T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 217.
38. Kamenecka, T. M.; Lanza, T., Jr.; de, L. S. E.; Li, B.; McCauley, E. D.; Van, R. G.; Egger, L. A.; Kidambi, U.; Mumford, R. A.; Tong, S.; MacCoss, M.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2205.
39. Kamenecka, T. M.; Park, Y.-J.; Lin, L. S.; de, L. S.; McCauley, E. D.; Van, R. G.; Egger, L.; Kidambi, U.; Mumford, R. A.; Tong, S.; Tang, W.; Colletti, A.; Teffera, Y.; Stearns, R.; MacCoss, M.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2323.
40. Kopka, I. E.; Young, D. N.; Lin, L. S.; Mumford, R. A.; Magriotis, P. A.; MacCoss, M.; Mills, S. G.; Van, R. G.; McCauley, E.; Egger, L. E.; Kidambi, U.; Schmidt, J. A.; Lyons, K.; Stearns, R.; Vincent, S.; Colletti, A.; Wang, Z.; Tong, S.; Wang, J.; Zheng, S.; Owens, K.; Levorse, D.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 637.
41. Li, B.; de, L. S. E.; Kamenecka, T. M.; Kopka, I. E.; Durette, P. L.; Lanza, T., Jr.; MacCoss, M.; Tong, S.; Mumford, R. A.; McCauley, E. D.; Van, R. G.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2141.
42. Lin, L. S.; Kopka, I. E.; Mumford, R. A.; Magriotis, P. A.; Lanza, T., Jr.; Durette, P. L.; Kamenecka, T.; Young, D. N.; de, L. S. E.; McCauley, E.; Riper, G. V.; Kidambi, U.; Egger, L. A.; Tong, X.; Lyons, K.; Vincent, S.; Stearns, R.; Colletti, A.; Teffera, Y.; Fenyk-Melody, J.; Schmidt, J. A.; MacCoss, M.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 611.
43. Lin, L. S.; Lanza, T., Jr.; McCauley, E.; Van, R. G.; Kidambi, U.; Cao, J.; Egger, L. A.; Mumford, R. A.; Schmidt, J. A.; MacCoss, M.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 133.
44. Lin, L. S.; Lanza, T. J., Jr.; Castonguay, L. A.; Kamenecka, T.; McCauley, E.; Van, R. G.; Egger, L. A.; Mumford, R. A.; Tong, X.; MacCoss, M.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2331.
45. Porter, J. R.; Archibald, S. C.; Brown, J. A.; Childs, K.; Critchley, D.; Head, J. C.; Hutchinson, B.; Parton, T. A. H.; Robinson, M. K.; Shock, A.; Warrellow, G. J.; Zomaya, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1595.
46. Porter, J. R.; Archibald, S. C.; Brown, J. A.; Childs, K.; Critchley, D.; Head, J. C.; Hutchinson, B.; Parton, T. A. H.; Robinson, M. K.; Shock, A.; Warrellow, G. J.; Zomaya, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1591.

47. Porter, J. R.; Archibald, S. C.; Childs, K.; Critchley, D.; Head, J. C.; Linsley, J. M.; Parton, T. A. H.; Robinson, M. K.; Shock, A.; Taylor, R. J.; Warrellow, G. J.; Alexander, R. P.; Langham, B. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1051.
48. Saku, O.; Ohta, K.; Arai, E.; Nomoto, Y.; Miura, H.; Nakamura, H.; Fuse, E.; Nakasato, Y. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1053.
49. Sidduri, A.; Tilley, J. W.; Hull, K.; Lou, J. P.; Kaplan, G.; Sheffron, A.; Chen, L.; Campbell, R.; Guthrie, R.; Huang, T.-N.; Huby, N.; Rowan, K.; Schwinge, V.; Renzetti, L. M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2475.
50. Sidduri, A.; Tilley, J. W.; Lou, J. P.; Chen, L.; Kaplan, G.; Mennona, F.; Campbell, R.; Guthrie, R.; Huang, T.-N.; Rowan, K.; Schwinge, V.; Renzetti, L. M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2479.
51. Stasiak, M.; Mehlin, C.; Boni, E.; Vaisar, T.; Little, T.; Kim, H.-O.; Qabar, M. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3875.
52. Tilley, J. W.; Kaplan, G.; Rowan, K.; Schwinge, V.; Wolitzky, B. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1.
53. Venkatraman, S.; Lim, J.; Cramer, M.; Gardner, M. F.; James, J.; Alves, K.; Lingham, R. B.; Mumford, R. A.; Munoz, B. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4053.
54. Yang, G. X.; Chang, L. L.; Truong, Q.; Doherty, G. A.; Magriotis, P. A.; de, L. S. E.; Li, B.; MacCoss, M.; Kidambi, U.; Egger, L. A.; McCauley, E.; Van, R. G.; Mumford, R. A.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1497.
55. Elliot, D.; Henshaw, E.; MacFaul, P. A.; Morley, A. D.; Newham, P.; Oldham, K.; Page, K.; Rankine, N.; Sharpe, P.; Ting, A.; Wood, C. M. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4832.
56. Xiao, T.; Takagi, J.; Coller, B. S.; Wang, J. H.; Springer, T. A. *Nature* **2004**, *432*, 59.
57. Xiong, J. P.; Stehle, T.; Zhang, R. G.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Aranout, M. A. *Science* **2002**, *296*, 151.
58. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
59. Klebe, G. *Perspect. Drug Discovery Des.* **1998**, *12*, 87.