Bioorganic & Medicinal Chemistry 16 (2008) 7749–7758

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

Development of a receptor-based 3D-QSAR study for the analysis of MMP2, MMP3, and MMP9 inhibitors

Tiziano Tuccinardi, Elisa Nuti, Gabriella Ortore, Armando Rossello, Stanislava I. Avramova, Adriano Martinelli *

Dipartimento di Scienze Farmaceutiche, Università di Pisa, via Bonanno 6, 56126 Pisa, Italy

ARTICLE INFO

Article history:
Received 18 February 2008
Revised 23 June 2008
Accepted 2 July 2008
Available online 8 July 2008

Keywords:
MMP2
MMP9
MMP3
3D-QSAR
Docking

ABSTRACT

The ability of Gold software to predict the binding disposition of matrix metalloproteinase (MMP) inhibitors was evaluated using MMP3 and MMP8. The best procedure was subsequently employed to dock into MMP2, MMP3 and MMP9 nearly 70 compounds that were tested for their inhibitory activity against the three MMP subtypes. The best binding poses were used as an alignment tool for the development of 3D-QSAR studies. Evaluation of the three resulting 3D-QSAR models allowed us to indicate the ligand properties and residues important for activity and selectivity. MMP2 is an important anticancer drug target, while MMP3 and MMP9 are considered to be anti-targets for tumor pathologies. As such, our results could predict the binding affinities of new MMP2 inhibitors, providing additional information regarding the selectivity against MMP3 and MMP9. Furthermore, this strategy may be used also for the investigation of other MMPs.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Matrix metalloproteinases (MMPs) are a group of zinc-dependent endopeptidases which degrade all extracellular matrix components, including collagen, gelatin, fibronectin, elastin and aggrecan. Under normal physiological conditions, they play critical roles in development, tissue morphogenesis, wound healing, and apoptosis.[1,2] To date, 26 members of this enzyme family have been reported, all sharing significant sequence homologies. They have been grouped into five classes based on their substrate specificity. These include gelatinases (MMP2 and MMP9), collagenases (MMP1, MMP8, and MMP13), stromelysins (MMP3, MMP10), membrane type (MMP14, MMP16, and MMP17), and other enzymes such as matrilysin (MMP7).[3]

The production and activities of MMPs are regulated at the level of transcription, activation of the precursor zymogens, inhibition by endogenous inhibitors, and endocytosis. Abnormal MMP expression and activity due to an imbalance in these control mechanisms have been linked to several degenerative diseases such as cancer, arthritis, cardiac injury, periodontal disease and multiple sclerosis.[4]

Their involvement in all stages of cancer progression is well documented, and intensive efforts have been made to design MMP inhibitors (MMPIs) as cancer therapeutic agents.[5,6] Almost all MMPIs bear chelating moieties that interact directly with the catalytic zinc cation and protrude into the hydrophobic S1′ subsite, a deep pocket situated in proximity to the catalytic zinc ion.[7] These compounds behave as competitive inhibitors since the zinc binding group (ZBG) binding mode mimics one of the transition states occurring during substrate hydrolysis.[8] Widely utilized ZBGs include hydroxamic acids, carboxylic acids, and thiols. The hydroxamate group is most commonly used due to its optimal chelating properties. Many small-molecule MMPIs that have been synthesized in the past 20 years by both pharmaceutical companies and academicians have entered Phase III clinical studies in patients with advanced cancer. Unfortunately, clinical results have been disappointing.[9] Such pharmacological failures likely have several possible origins. The primary reason is that the majority of these clinical studies have involved advanced-stage cancer, while it has recently been established that MMPIs are most effective in the early phases of cancer progression. Moreover, the first clinical studies were started when little was known about the roles of various MMPs in cancer. As such, broad-spectrum inhibitors of MMPs were used.[10] Recent studies on both MMP-knockout mice and mice that transgenically overexpress MMPs have revealed many host-protective functions of MMPs, including repression of angiogenesis in mouse tumor models. Therefore, a broad MMP inhibition strategy aimed at stopping basement-membrane penetration and subsequent metastasis has failed in many patients due to the loss of these anti-tumor actions of some MMPs.

MMPs have recently been classified as targets and anti-targets in tumors, and a successful MMPI should spare MMP anti-targets

* Corresponding author. Tel.: +39 050 2219556; fax: +39 050 2219605.
E-mail address: marti@farm.unipi.it (A. Martinelli).

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.07.004

while potently inhibiting the targets.[11] A major challenge lies in distinguishing among specific MMPs due to their high active site sequence homologies that make selective inhibitors very difficult to develop.

For many years, MMP2 has been considered a therapeutic target based on a high level of expression in many human tumor samples. Further, MMP2 has been identified in association with highly invasive cells, and has the ability to degrade native type IV collagen. Recent studies have confirmed the role of MMP2 in tumors, and validated it as an anticancer drug target. In experimental models, MMP2-null mice have a 50% reduction in angiogenesis induction, melanoma growth, and lung carcinoma metastasis.[12] Moreover, MMP2 overexpression in MDA-MB-231 breast carcinoma cells has been shown to increase orthotopic tumor growth and metastasis in nude mice.[13]

MMP3, on the other hand, has been validated as an anti-target in cancer since in a carcinogen-induced skin cancer model, tumors that develop in MMP3-null mice were more advanced and associated with pulmonary metastasis than in wild-type littermates.[14] The increased tumor aggressiveness appears to be related to a function of the enzyme in regulating the immune system. Moreover, transgenic mice overexpressing MMP3 in their mammary glands developed fewer 7,12-dimethylbenz[a]anthracene (DMBA)-induced tumors than control mice, underscoring the protective role of this MMP in certain mouse models.[15]

MMP9 has both pro- and anti-tumorigenic effects, depending on the tumor stage. MMP9 inhibition might be useful in treating patients with early-stage cancers, but it is an anti-target in patients with advanced disease. For example, the human papilloma virus (HPV) 16-induced carcinomas that arise in an MMP9-null background are more aggressive and of a higher grade than those in wild-type mice, indicating a suppressive effect for this protease during tumor progression.[16] The protective effects of MMP9 have been proposed to arise from its capacity to generate angiogenesis inhibitors such as angiostatin and tumstatin, a peptide fragment from the NC1 domain of the α3 chain of type IV collagen.[17]

In all, MMP2 is considered an important anticancer drug target, while MMP3 and MMP9 are anti-targets for tumor pathologies. With the aim of directing the activity of synthetic MMPIs toward MMP2 with a good overall selectivity against MMP3 and MMP9, this work employs the Gold program[18] to evaluate the reliability of various docking procedures to predict the binding dispositions of MMPIs. Previously, we have verified that this software can be successfully used for docking studies with metalloenzymes.[19,20] By employing ligands with inhibitory activities that were tested at the same time on MMP2, MMP3, and MMP9, receptor-based 3D-QSAR models for the three MMPs were developed, and their abilities to explain ligand affinities and selectivities were verified.

2. Results and discussion

2.1. Cross-docking

In order to verify the reliability of our docking protocol, we applied a cross-docking approach to the MMP3 and MMP8 complexes by docking eight MMP3 ligands and six MMP8 ligands inside the X-ray proteins, and comparing the binding dispositions with those observed in the X-ray crystal structures. The 14 inhibitors were docked into their crystal structures using the Gold program[18] after an extensive conformational search (CS) using both the available GoldScore and ChemScore fitness functions.

The flexibility of MMPs was evaluated by aligning all X-ray proteins (using 1A85 and 1B3D as a reference protein for MMP8 and MMP3, respectively). For this purpose, the binding site was defined as all residues within 10 Å from the reference inhibitor. As shown in Table 1, MMPs appear to possess a low degree of motion upon the ligand binding. The root mean square deviation (RMSD) of the protein backbone never exceeded 0.5 Å, and the binding site RMSD value was consistently below 1.7 Å.

Figure 1 shows the cross-docking results obtained through GoldScore and ChemScore for the MMP3 and MMP8 complexes. Using the GoldScore function for MMP3, the Gold program predicted the binding disposition of the ligands with an average RMSD (ARMSD) of 3.9 Å. Fourteen ligands possessed a predicted disposition that, compared with their X-ray structures, showed an RMSD greater than 5.0 Å. The ChemScore fitness function provided docking results that were more promising than those obtained by means of the GoldScore, since they showed an ARMSD of 2.8 Å with 10 ligands that possessed an RMSD value greater than 5.0 Å.

The analysis of the ZBG disposition for the MMP3 complexes revealed that the ChemScore was better able to predict the ZBG interaction with the protein, since it showed a lower ARMSD (1.2 Å), and produced only one compound with an RMSD value greater than 5.0 Å. The GoldScore, on the other hand, produced docking results that showed an ARMSD of the ZBG disposition of 2.0 Å with three compounds that showed an RMSD greater than 5.0 Å.

The analysis of the results obtained for the MMP8 complexes was not in agreement with those obtained for MMP3. As shown in Figure 1, using the GoldScore function, the docking results for the binding disposition of the ligands showed an ARMSD of 3.1 Å, with six ligands that had an RMSD value greater than 5.0 Å. The ChemScore function, however, was able to predict the binding disposition of the ligands with an ARMSD of 3.9 Å, and fifteen ligands possessed an RMSD value greater than 5.0 Å.

The GoldScore function showed an ARMSD of 1.5 Å, with two ligands that had an RMSD value greater than 5.0 Å for analysis of the ZBG disposition. The docking results obtained using the ChemScore function showed an ARMSD of 2.4 Å, and six ligands possessed an RMSD value greater than 5.0 Å. Overall, the results obtained using the two fitness functions did not clearly indicate the optimal procedure.

The alignment of all the available experimental structures of MMPs complexed with ligands containing the hydroxamate function as ZBG suggested that the hydroxamate interacts with the zinc ion in the same binding position for all the MMPs. This allowed for the use of the scaffold match constraint (SMC) function of Gold, applicable for both the ChemScore and the GoldScore fitness functions. This option attempts to estimate the degree of possible fixed

Table 1
Backbone superimposition of the whole proteins (WP RMSD) and heavy atom superimposition of the binding sites (BS RMSD) using the X-ray structure 1A85 and 1B3D as reference proteins

PDB	WP RMSD	BS RMSD
MMP3		
1B3D	0	0
1BIW	0.40	0.62
1BQO	0.38	1.63
1D5J	0.37	0.89
1D7X	0.29	1.39
1D8F	0.37	0.98
1D8M	0.33	0.81
1G49	0.39	0.83
MMP8		
1A85	0	0
1A86	0.15	0.31
1JAQ	0.16	0.32
1KBC	0.40	0.49
1MMB	0.23	0.41
1MNC	0.42	0.91

All the reported values are in Å.

RMSD < 2.0 | 2.0 < RMSD < 3.5 | 3.5 < RMSD < 5.0 | RMSD > 5.0

GoldScore MMP3

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	1.0	1.3	2.2	3.7	2.0	9.2	3.5	2.9	3.2
1BIW	0.8	1.6	4.2	1.4	1.2	8.8	2.6	1.8	2.8
1BQO	5.1	3.0	4.7	5.5	3.3	9.0	3.7	2.9	4.7
1D5J	1.0	3.2	4.3	5.3	2.5	9.0	2.7	2.8	3.9
1D7X	1.2	3.1	4.2	1.4	4.2	2.9	3.4	4.1	3.1
1D8F	5.6	2.5	3.6	1.6	3.6	9.0	7.4	4.0	4.7
1D8M	5.2	1.5	4.5	3.4	4.1	9.0	3.5	4.3	4.4
1G49	5.1	3.3	4.7	3.9	3.4	8.9	3.6	4.4	4.7
								ARMSD	3.9

ChemScore MMP3

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	1.0	1.3	3.7	6.0	1.0	4.1	1.8	1.5	2.6
1BIW	4.5	1.4	3.8	7.1	0.9	3.5	1.4	1.5	3.0
1BQO	1.1	1.1	4.6	6.8	0.9	3.8	1.5	1.6	2.7
1D5J	2.1	8.4	4.9	3.4	1.5	1.6	1.4	1.4	3.1
1D7X	4.9	8.5	5.6	2.3	1.3	4.4	2.1	1.4	3.8
1D8F	1.0	1.1	3.5	6.8	0.9	3.7	1.8	1.2	2.5
1D8M	3.3	1.3	5.7	2.4	1.1	1.4	1.6	1.4	2.3
1G49	0.8	2.3	5.7	6.0	0.9	2.2	2.1	1.5	2.7
								ARMSD	2.8

ZBG

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	0.4	1.0	1.0	1.1	1.0	5.0	1.5	5.0	2.0
1BIW	0.6	0.5	0.9	2.2	0.6	1.1	0.7	0.7	0.9
1BQO	2.3	1.6	1.2	2.7	1.4	2.4	1.8	1.4	1.9
1D5J	1.0	3.4	1.2	2.6	1.3	3.3	1.3	1.4	1.9
1D7X	1.4	4.5	1.1	2.2	1.9	1.2	1.6	1.6	1.9
1D8F	1.2	2.7	2.6	2.2	1.4	1.8	4.3	4.1	2.5
1D8M	3.4	1.0	1.1	4.2	1.6	2.3	1.8	1.6	2.1
1G49	3.3	3.8	1.2	5.0	1.4	1.8	1.7	1.7	2.5
								ARMSD	2.0

ZBG

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	0.7	0.6	0.9	5.0	0.7	0.8	0.7	0.7	1.2
1BIW	0.5	0.5	0.4	2.5	1.1	1.4	1.0	1.0	1.0
1BQO	0.6	0.8	1.3	2.5	0.8	1.1	0.8	0.7	1.1
1D5J	0.9	3.0	1.6	2.4	1.1	0.9	1.3	1.1	1.5
1D7X	0.8	2.8	2.5	1.5	1.3	0.8	0.8	0.6	1.4
1D8F	0.7	0.8	0.7	2.4	0.5	1.0	0.7	0.7	0.9
1D8M	2.8	0.9	1.2	1.7	0.8	0.8	1.1	0.8	1.3
1G49	0.8	2.6	2.6	2.8	0.6	0.7	1.0	0.7	1.5
								ARMSD	1.2

GoldScore MMP8

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	2.1	6.2	3.8	2.0	6.8	2.1	3.8
1A86	2.1	1.9	3.7	1.9	2.6	1.0	2.2
1JAQ	1.8	1.7	3.6	6.9	6.2	5.9	4.4
1KBC	2.9	1.7	3.6	1.8	3.4	2.3	2.6
1MMB	2.1	2.0	3.8	1.4	6.1	2.2	2.9
1MNC	2.1	2.2	3.5	1.6	3.4	2.2	2.5
						ARMSD	3.1

ChemScore MMP8

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	2.0	2.3	5.2	3.8	7.0	1.2	3.6
1A86	1.9	6.2	5.2	3.4	1.8	1.2	3.3
1JAQ	4.8	5.3	3.5	3.8	2.4	6.4	4.4
1KBC	6.4	4.7	5.4	3.8	2.0	6.1	4.7
1MMB	6.8	5.3	5.9	2.6	2.0	1.4	4.0
1MNC	6.0	5.2	5.6	2.0	1.8	0.6	3.5
						ARMSD	3.9

ZBG

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	0.9	8.0	1.4	2.0	2.3	2.4	2.8
1A86	1.0	0.9	1.3	1.2	1.4	1.1	1.1
1JAQ	0.6	0.9	1.2	5.0	2.7	0.5	1.8
1KBC	2.9	0.7	0.6	0.9	0.6	0.6	1.0
1MMB	1.0	1.2	0.8	1.0	1.3	0.9	1.1
1MNC	0.7	0.8	0.7	2.0	0.5	0.7	0.9
						ARMSD	1.5

ZBG

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	0.9	0.7	1.6	1.8	2.8	1.5	1.5
1A86	0.7	1.7	1.7	1.1	0.6	1.2	1.2
1JAQ	5.0	5.0	2.6	1.6	0.7	2.8	2.9
1KBC	1.1	5.0	1.5	1.1	0.6	3.1	2.1
1MMB	10.6	7.8	2.9	2.4	1.8	1.8	4.6
1MNC	1.3	5.0	1.6	2.5	0.3	0.4	1.9
						ARMSD	2.4

Figure 1. Matrix of RMSDs obtained by the cross-docking study using the GoldScore and ChemScore for MMP3 and MMP8.

geometries of the substructure key for every ligand during the entire docking process. In this study, the substructure key for each ligand was the hydroxamate group, using the ZBG of 1A85 and 1B3D as reference substructures for MMP8 and MMP3, respectively.

As shown in Figure 2, the use of the SMC function provided a general improvement in the obtained RMSD for both procedures. All ARMSD for the ZBG showed values below 1.0 Å. As regards the ligand disposition, the docking results using the GoldScore fitness function showed an ARMSD of 2.2 Å for MMP3 and 2.3 for MMP8 with a total of eight ligands that showed a RMSD greater than 5.0 Å. The ChemScore function showed an ARMSD of 1.8 Å for MMP3 and 3.5 Å for MMP8 with eleven ligands that showed a RMSD greater than 5.0 Å.

2.2. Automated docking

Among the four procedures described above, the use of the GoldScore fitness function together with the Scaffold Match Constraint was optimal for determining binding geometries of the ZBG, and also for determining the disposition of the ligands inside the receptor. However, the use of the SMC-ChemScore method provided similar results.

For further investigation of the obtained results, 69 inhibitors[21–28] that had been tested previously for their activities on MMP2, MMP3, and MMP9, all bearing hydroxamate functionalities as ZBGs (see Table 1 of the Supporting Information), were docked into the MMP2, MMP3, and MMP9 binding sites. These compounds were characterized by seven different central scaffolds, and had inhibitory activities that spanned five orders of magnitude, with IC_{50} that ranged from 1 nM to 16 μM (Fig. 3).

The docking results were then scored using the ChemScore and GoldScore fitness functions, but the correlations between the calculated and the experimental free energies of binding were not in accord, since the quadratic correlation (R^2) showed values below 0.4.

2.3. Receptor-based 3D-QSAR analysis

With the aim of obtaining quantitative results, the docking poses were used as receptor-based alignments for 3D-QSAR models. The reliabilities of the 3D-QSAR models were characterized by their correlation coefficient (r^2), predictive correlation coefficient (q^2) and cross-validated standard deviation of errors of prediction ($SDEP_{CV}$). Compounds from the original set were divided into a training set and a test set, according to the usual guidelines. In detail, the training and test sets should contain compounds in such a way to maximize their structural diversity (i.e., compounds belonging to the training set and to the test set should be representative of the molecular diversity of all the compounds under study) and uniformly span the full range of activities. As a result, the training set was composed of 54 compounds, while 15 molecules were selected for the test set to validate the predictive power of the models (see Table 1 of the Supporting Information).

The combination of the C3 (corresponding to a methyl group) and N2 (corresponding to neutral NH_2) probes using the alignment originated by the SMC-GoldScore method showed the best results, and was used for the further studies. Table 3 and Figure 4 report the 3D-QSAR analysis data obtained for MMP2, MMP3, and MMP9 using the poses obtained from the SMC-GoldScore docking method and applying C3 and N2 probes. The third PLS component of the MMP2 model rationalized 92% of variance, and was predictive (q^2 = 0.81, SDEP of test set = 0.63); the fourth PLS component showed worse predictive results (r^2 = 0.95, q^2 = 0.80 and the SDEP of the test set = 0.68); and the fifth PLS component had no further predictivity improvement. As shown in Table 2, similar results were also obtained for MMP3 and MMP9. As a result, the optimal

RMSD < 2.0	2.0 < RMSD < 3.5	3.5 < RMSD < 5.0	RMSD > 5.0

SMC-GoldScore MMP3

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	0.7	1.5	3.4	5.8	1.0	2.3	2.0	1.5	2.3
1BIW	0.9	1.7	3.3	5.9	0.9	6.3	2.5	1.4	2.9
1BQO	0.6	1.5	4.4	4.0	1.0	2.4	2.6	1.7	2.3
1D5J	0.9	0.8	3.4	5.9	0.8	2.7	1.9	1.5	2.2
1D7X	0.7	0.6	4.7	4.1	0.9	2.5	1.7	1.6	2.1
1D8F	0.5	0.7	3.2	1.6	1.0	2.6	1.6	1.6	1.6
1D8M	0.6	0.8	4.6	1.3	0.9	3.8	6.8	1.5	2.5
1G49	0.6	0.7	4.6	1.2	1.0	2.9	1.5	1.6	1.8
								ARMSD	2.2

SMC-ChemScore MMP3

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	1.0	1.2	3.2	1.1	1.1	4.2	1.6	1.3	1.8
1BIW	0.6	1.8	3.8	6.1	1.0	3.8	1.4	1.5	2.5
1BQO	1.3	1.1	3.4	0.8	1.0	3.0	1.4	1.5	1.7
1D5J	1.1	1.3	1.8	1.4	2.2	2.7	1.6	1.3	1.7
1D7X	1.2	1.2	3.4	0.9	0.9	4.2	2.0	1.3	1.9
1D8F	1.0	1.2	3.3	3.0	1.0	1.8	1.3	1.4	1.8
1D8M	1.4	1.0	3.5	0.7	1.1	4.3	1.3	1.4	1.8
1G49	1.1	1.1	3.5	1.1	0.9	1.8	1.6	1.4	1.6
								ARMSD	1.8

ZBG

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	0.4	0.7	0.5	0.9	0.7	0.6	0.6	0.6	0.6
1BIW	0.6	0.5	0.2	0.4	0.5	0.4	5.0	0.5	1.0
1BQO	0.7	0.9	0.8	0.5	0.7	0.7	0.7	0.6	0.7
1D5J	0.9	0.9	0.7	0.5	0.6	5.0	5.0	0.6	1.8
1D7X	0.5	0.8	0.9	0.9	0.7	0.6	0.8	0.7	0.7
1D8F	0.5	0.7	0.4	1.1	0.5	0.4	0.6	0.6	0.6
1D8M	0.7	1.0	0.9	0.3	0.8	0.6	0.7	0.7	0.7
1G49	0.7	0.9	1.2	5.0	0.9	0.8	0.9	0.8	1.4
								ARMSD	0.9

ZBG

Proteins \ Ligands	1B3D	1BIW	1BQO	1D5J	1D7X	1D8F	1D8M	1G49	RMSD
1B3D	0.5	0.6	0.2	0.6	0.3	0.9	0.4	0.4	0.5
1BIW	0.6	0.4	0.3	0.6	0.3	0.4	0.2	0.4	0.4
1BQO	0.4	0.7	0.3	0.6	0.5	0.4	0.5	0.5	0.5
1D5J	0.5	0.6	0.5	0.4	5.0	0.4	5.0	0.7	1.6
1D7X	0.3	0.7	0.2	0.3	0.3	0.3	0.6	0.5	0.4
1D8F	0.4	0.6	0.2	0.6	0.3	0.4	0.3	0.4	0.4
1D8M	0.5	0.7	0.3	0.2	0.4	0.4	0.5	0.5	0.4
1G49	0.6	0.8	0.4	5.0	0.5	0.4	0.6	0.7	1.1
								ARMSD	0.7

SMC-GoldScore MMP8

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	2.0	2.1	2.2	1.5	1.7	1.0	1.8
1A86	1.8	1.7	3.4	1.4	1.9	2.0	2.0
1JAQ	1.7	1.4	1.5	1.6	1.2	6.0	2.2
1KBC	1.5	1.4	3.9	1.5	4.5	4.0	2.8
1MMB	1.8	6.3	5.0	1.9	1.7	1.0	3.0
1MNC	1.8	1.9	3.7	1.7	1.6	0.5	1.9
						ARMSD	2.3

SMC-ChemScore MMP8

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	2.3	2.0	2.3	3.6	1.7	5.3	2.9
1A86	2.3	1.5	4.1	2.9	1.7	0.6	2.2
1JAQ	4.9	1.6	3.5	8.7	1.9	4.8	4.2
1KBC	5.9	6.4	6.1	3.5	2.8	4.3	4.8
1MMB	7.0	6.3	5.0	2.0	1.5	1.4	3.9
1MNC	2.3	5.4	5.0	2.5	1.7	0.7	2.9
						ARMSD	3.5

ZBG

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	0.7	0.6	0.6	0.8	0.9	1.0	0.8
1A86	0.8	0.8	1.1	1.1	0.9	1.1	1.0
1JAQ	0.5	0.5	0.5	0.9	0.9	0.4	0.6
1KBC	0.5	0.6	0.5	0.5	0.5	0.5	0.5
1MMB	0.4	0.7	0.7	0.6	0.7	0.7	0.6
1MNC	0.6	0.6	0.5	0.6	0.5	0.6	0.5
						ARMSD	0.7

ZBG

Proteins \ Ligands	1A85	1A86	1JAQ	1KBC	1MMB	1MNC	RMSD
1A85	0.3	0.5	0.5	0.9	0.8	0.5	0.6
1A86	0.3	0.4	0.3	0.7	0.6	0.6	0.5
1JAQ	0.5	0.3	0.6	0.5	0.8	0.3	0.5
1KBC	0.7	0.8	0.9	0.8	0.4	0.6	0.7
1MMB	0.7	0.6	0.9	0.6	0.7	0.6	0.7
1MNC	0.4	0.7	1.2	0.7	0.4	0.4	0.6
						ARMSD	0.6

Figure 2. Matrix of RMSDs obtained from the cross-docking study using the GoldScore and ChemScore with the SMC function for MMP3 and MMP8.

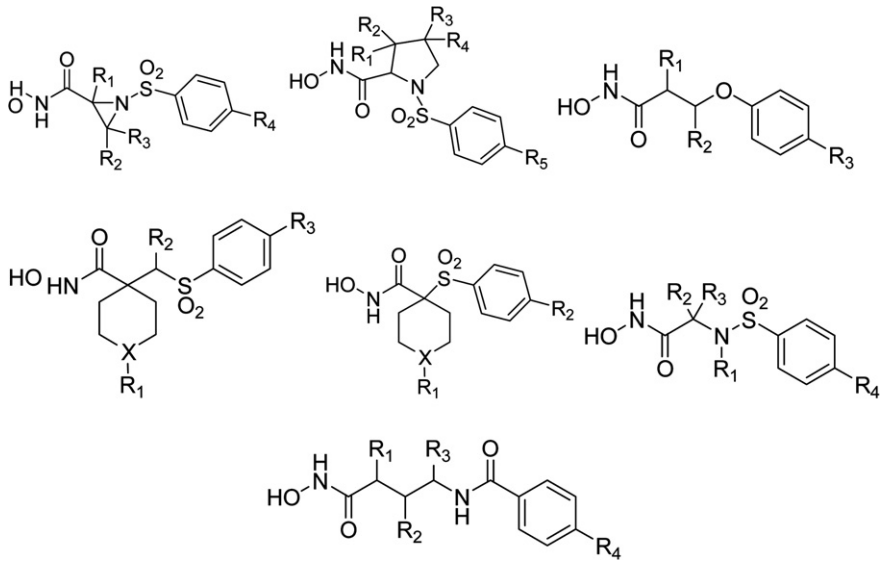


Figure 3. General scaffolds of the compounds used for the docking study.

dimensionality for the three 3D-QSAR models was provided by three components. For the two probes, a score was determined according to the information obtained through the Consensus Principal Component Analysis (CPCA).[29] As shown in Table 2, the N2 probe that represented the electrostatic and/or hydrogen-bonding interaction was the most relevant for all the three models.

One important feature of 3D-QSAR analysis is graphical model representation, which is used to make interpretation easier. In the GOLPE program,[30] there are several options for displaying the final model. Among these, the PLS pseudocoefficient plots are very useful since they make it possible to visualize favorable and unfavorable interactions between the probes and the molecules under study.

Figure 5 illustrates the PLS coefficient plots for the model for the C3 and N2 probes obtained for MMP2, MMP3, and MMP9. Compound **48** (see Table 1 of the Supporting Information), a diphenyl ether α-sulfone derivative that showed a high activity for all the three MMPs, is also reported.

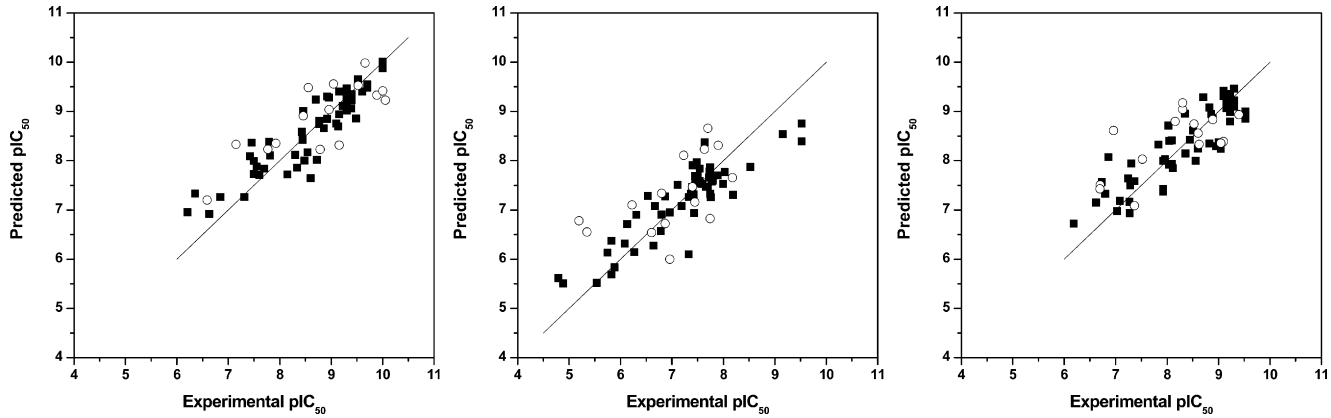


Figure 4. Plots of the 3D-QSAR model for MMP2 (left), MMP3 (center), and MMP9 (right), experimental/predicted pIC_{50} is reported. The training set is represented by (■) and the test set by (○).

Table 2
Statistical results of the receptor-based 3D-QSAR models

MMP2					MMP3					MMP9				
Vars	PC	r^2	q^2	$SDEP_{TS}$	Vars	PC	r^2	q^2	$SDEP_{TS}$	Vars	PC	r^2	q^2	$SDEP_{TS}$
879	**3**	**0.92**	**0.81**	**0.63**	**1177**	**3**	**0.91**	**0.77**	**0.79**	**1034**	**3**	**0.93**	**0.75**	**0.70**
879	4	0.95	0.80	0.68	1177	4	0.94	0.75	0.79	1034	4	0.95	0.74	0.73
879	5	0.97	0.78	0.66	1177	5	0.97	0.73	0.82	1034	5	0.97	0.72	0.73
C3 (%)		N2 (%)			C3 (%)		N2 (%)			C3 (%)		N2 (%)		
30.7		69.3			25.8		74.2			33.6		66.4		

The contribution for C3 and N2 probes was reported in the last two rows. In bold is reported the optimal dimensionality.

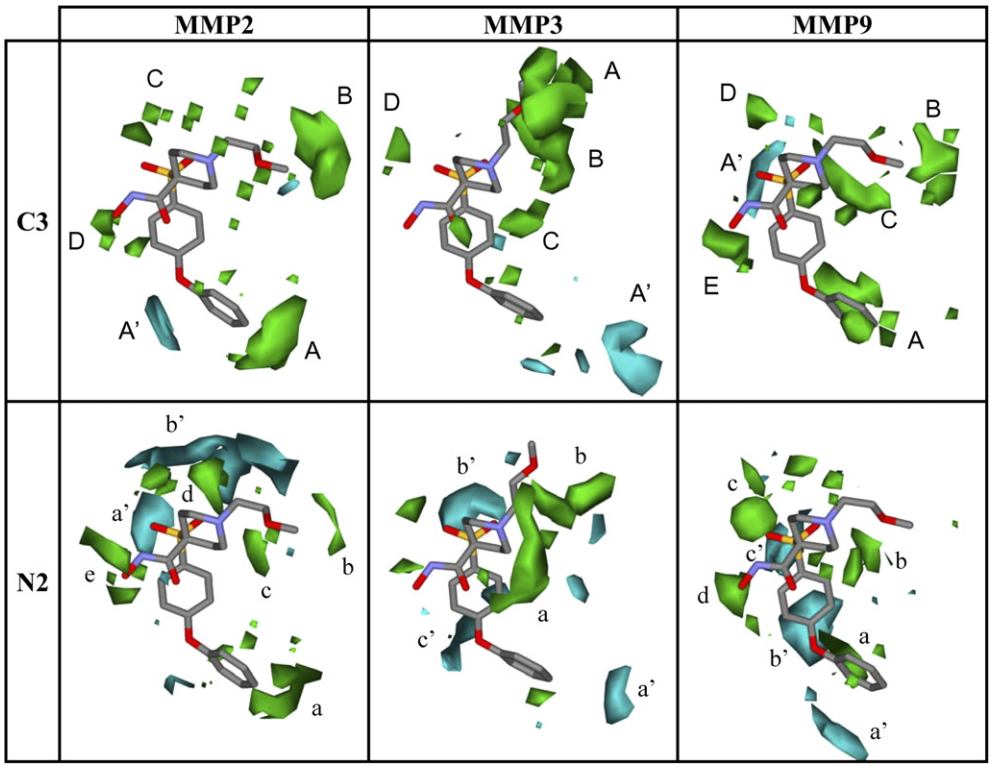


Figure 5. Negative (cyan) and positive (green) regions of the PLS coefficient plot obtained with the C3 and N2 probes. Compound **48** is also displayed as a reference structure.

In the first row of Figure 5, the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the C3 probe for the three MMPs are illustrated. For MMP2, there are four principal regions (A–D) with positive values, in which a favorable interaction between a substituent and the probe results in an increase in activity, whereas an unfavorable interaction between a substituent and

the probe results in a decrease in activity. There is only one negative region (A′) in which a favorable interaction between a substituent and the probe results in a decrease in activity, whereas an unfavorable interaction between a substituent and the probe results in an increase in activity. For MMP3, there are four positive regions (A–D) and one negative region (A′), and for MMP9, there are five positive regions (A–E) and one negative region (A′).

In the second row of Figure 5, the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the N2 probe for the MMPs are reported. For MMP2, there are two principal regions with negative values (a′, b′) in which a favorable interaction between a substituent and the probe results in an increase in activity, whereas an unfavorable interaction between a substituent and the probe results in a decrease in activity, and there are five positive regions (a–e) where a favorable interaction between a substituent and the probe results in a decrease in activity. For MMP3, there are three negative regions (a′–c′) and two positive regions (a, b), and for MMP9, there are three negative (a′–c′) and four positive regions (a–d).

Since the alignment of the ligands was performed using the structures docked into the MMP2, MMP3, and MMP9 proteins, identification of matching alignments among the MMP2, MMP3, and MMP9 receptors and the 3D-QSAR maps was conducted. In Figure 6, the binding sites of the MMP2, MMP3, and MMP9 overlap with the 3D PLS coefficient maps of the C3 and the N2 probes. For MMP2, region A of the C3 probe is situated in the proximity of Ala422, while surface B corresponds to Ile424, and region C is in the proximity of Leu190. With regard to the N2 probe, regions a′ and b′ are situated near to the backbone of Gly189, Leu190, Leu191, and Ala192.

For MMP3, region C of the C3 probe is in the proximity of Pro238, while for the N2 probe, region a′ is in the proximity of His241 and surface b′ corresponds to Leu226 and Ala227. For MMP9, region A of the C3 probe is in the proximity of Tyr423 and residues Leu188 and Leu187 are in the proximity of regions C and D, respectively. Finally region c′ is near to the backbone of Leu188 and Ala189.

With the aim of further analyzing the results obtained for the three MMPs, the PLS coefficients maps of each MMP were compared to those of the other two MMPs in order to evaluate whether the different dispositions of the PLS maps could be correlated to the interaction with non-conserved residues. From this approach, it was determined that regions A and B (probe C3) of MMP2 were not present in MMP3 and MMP9. Ala422 in MMP2, which corresponded to region A, was not conserved, and it was replaced by a tyrosine residue in both MMP3 and MMP9 (Tyr237 and Tyr420, respectively). Region B in MMP2 corresponded to Ile424, which was also not conserved, and it was substituted by Leu239 and Met422 in MMP3 and MMP9, respectively.

For the N2 probe, the only noted difference was that region a′ of MMP3 was not present in MMP2 and MMP9. This region in the MMP3 corresponded to His241 and this residue was replaced by Thr429 in the MMP2 and Arg424 in the MMP9.

The analysis of the other polar and hydrophobic residues suggested that there were no other residues important for the selectivity and directly correlated with PLS maps; the side chains of these residues were completely aligned in MMP2, MMP3, and MMP9 crystal structures, and could not justify the different dispositions of the PLS maps.

However, the surface analysis of the crystallographic structure of the three MMP subtypes revealed that the S1′ cavity, where the not conserved residues suggested as important for the MMP selectivity are located, possessed a different shape due to the presence of different residues. As shown in Figure 7 in the Supporting Information, the S1′ channel of MMP2 was mainly delimited by Pro417, Gly418, Ala422, Ile424, and Thr426, and showed a large aperture at the end of the pocket. With respect to MMP2, the S1′ pocket of MMP3 was delimited by different residues (Thr232, Glu233, Tyr237, Leu239, and His241) and showed a smaller aperture at the end of the cavity. Finally, the analysis of S1′ of MMP9

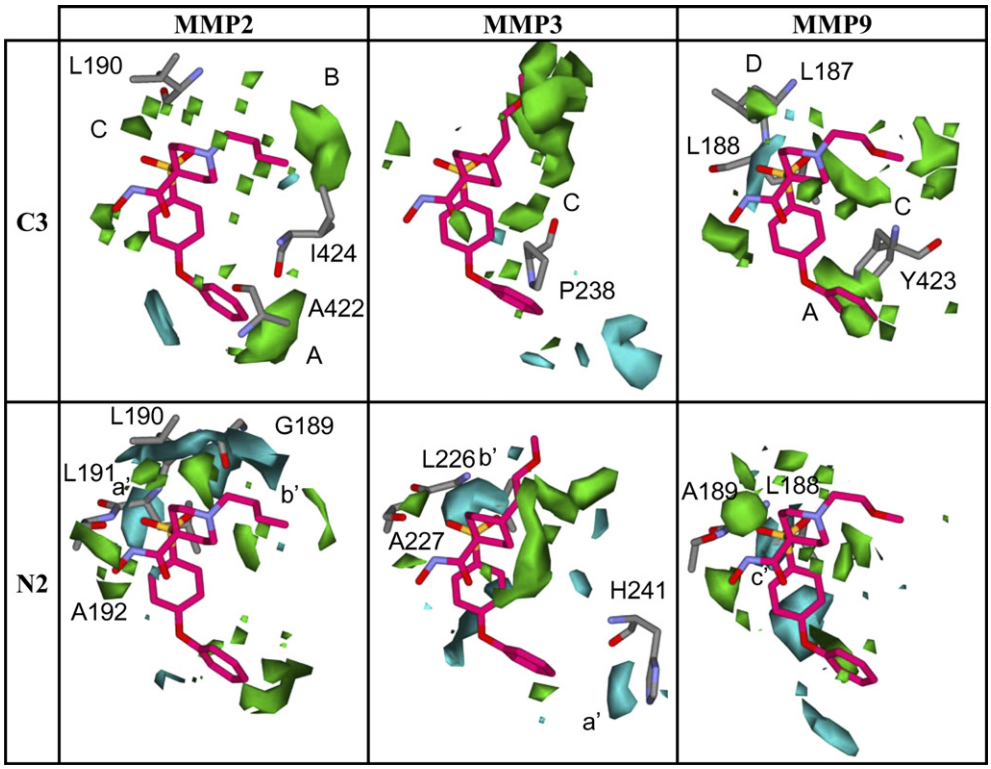
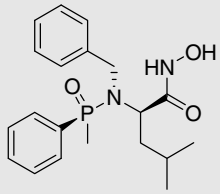
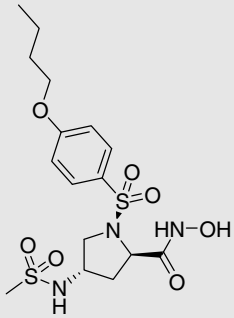
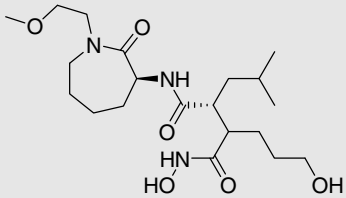
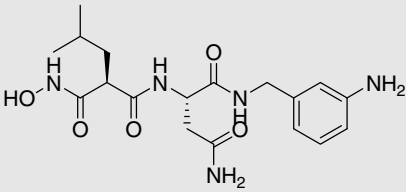
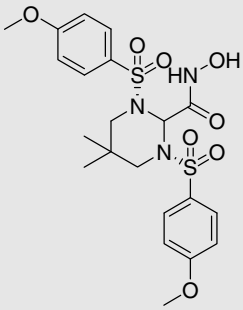
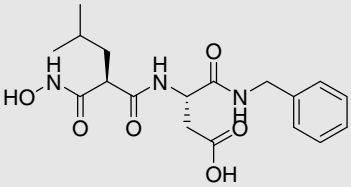
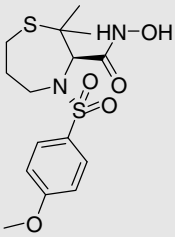
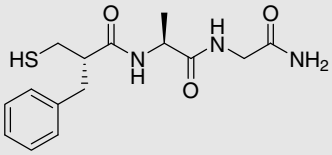
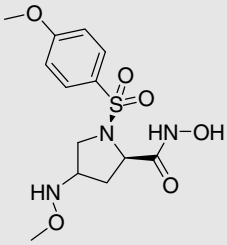
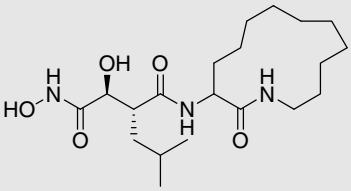


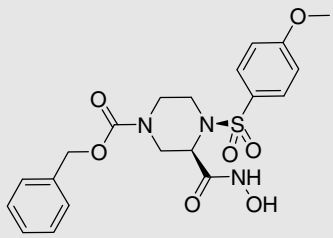
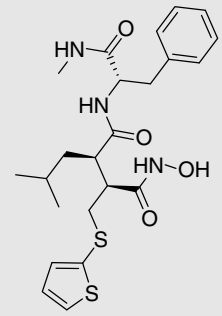
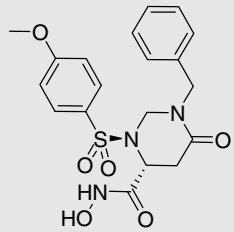
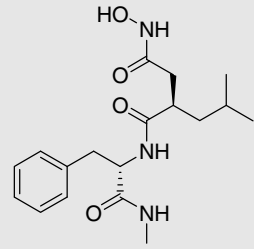
Figure 6. PLS coefficient plots obtained with the C3 and N2 probes superimposed on the MMP2, MMP3, and MMP9 binding sites.

Table 3
Ligand structures of the MMP X-ray complexes used in this study

PDB code	MMP	Ligand	PDB code	MMP	Ligand
1B3D	MMP3		1G49	MMP3	
1BIW	MMP3		1A85	MMP8	
1BQO	MMP3		1A86	MMP8	
1D5J	MMP3		1JAQ	MMP8	
1D7X	MMP3		1KBC	MMP8	

(*continued on next page*)

Table 3 (*continued*)

PDB code	MMP	Ligand	PDB code	MMP	Ligand
1D8F	MMP3		1MMB	MMP8	
1D8M	MMP3		1MNC	MMP8	

highlighted that, differently from the other two MMP subtypes, it was capped by the presence of the not conserved residue Arg424.

Not all these residues can be directly correlated to PLS coefficients for C3 and N2 probes, however, their effect on the ligand binding could be correlated to a different interaction of the ligands with the adjacent residues, highlighted as important by the 3D-QSAR study.

3. Conclusions

MMP2 is among the most studied matrix metalloproteinases due to its high level of expression in many human tumor samples, its association with highly invasive cells, and its ability to degrade native type IV collagen. Herein, we tested the potential of developing receptor-based 3D-QSAR models for MMPs. First, we analyzed the ability of the Gold program to predict the binding disposition of known MMPIs. The best procedure was then used to dock compounds previously tested for their inhibitory activities against MMP2, MMP3, and MMP9, and the best poses were used as an alignment tool for the development of three 3D-QSAR models. The results obtained allowed us to extract both qualitative and quantitative information about the ligand–receptor interactions.

The 3D-QSAR analysis indicated that the MMP2-inhibitor activity is determined by both electrostatic and lipophilic interactions, while the selectivity appears to be correlated primarily with lipophilic interactions. Interestingly, these interactions involve two non-conserved residues, Ala422 (in region A of the MMP2 3D-QSAR model) and Ile424 (in region B of the MMP2 3D-QSAR model).

Our goal was the development and validation of a computational approach for the study of MMP2 inhibitors. As such, inhibitors tested at the same time on MMP2, MMP3, and MMP9 were taken into account. The obtained 3D-QSAR models were able to rationalize both the affinity and the selectivity of the inhibitors, and therefore are potentially applicable toward the design of new selective MMP2 inhibitors. Our results encourage us to extend this methodology toward the study of others MMPs.

4. Experimental

4.1. MMP-inhibitor complex structures and docking

For evaluating docking reliability, the available X-ray structures of MMP3- and MMP8-inhibitor complexes were taken into account. We selected these two MMP subtypes since they have the highest number of deposited crystallographic structures. The MMP complexes were taken from the Protein Data Bank,[31] and used in this work for a total of 14 proteins (see Table 3). The hydrogen atoms were added by means of Maestro 7.5[32] using the all-atom model, and water molecules and Ca^{2+} ions were eliminated.

The ligands were extracted from the X-ray complexes, and were subjected to a CS of 1000 steps in a water environment (using the Generalized-Born/Surface-Area model) using the Macromodel program.[33] The Monte Carlo algorithm was used with the MMFFs forcefield, and a distance-dependent dielectric constant of 1.0. The ligands were then minimized using the Conjugated Gradient method to a convergence value of 0.05 kcal/Åmol, using the same forcefield and parameters as for the CS.

The minimized ligands were docked into their corresponding proteins using Gold 3.0.1.[18] The region of interest in Gold was defined in such a manner that it contains all the residues which stayed within 10 Å from the ligand in the X-ray structures. The zinc ion was set as a bipyramidal trigonal geometry atom. The 'allow early termination' command was deactivated, while the possibility for the ligand to flip ring corners was activated. The default Gold parameters were used for all remaining variables, and ligands were submitted to 40 Genetic Algorithm runs.

Four docking methods were carried out. In the first two cases, the two fitness functions implemented in Gold, GoldScore, and ChemScore were used. Subsequently, the scaffold match constraint (SMC) function was combined with the GoldScore and ChemScore functions. The scaffold match constraint weight, a parameter determining how closely ligand atoms fit onto the scaffold, was

set to the value 10.0. The best docked pose for each ligand was then used for further studies.

The docking results were evaluated through the comparison of the ligand docked positions and the experimental poses derived from the X-ray structures. As a measure of the docking reliability, the RMSD between the position of heavy atoms of the ligand in the calculated and experimental structures was taken into account.

4.2. Cross-docking

Each MMP3- and MMP8-inhibitor was docked, in turn, into each available MMP3 and MMP8 X-ray structure. A total of 100 docking calculations were developed for each of the four above-mentioned procedures (GoldScore, ChemScore, SMC-GoldScore, SMC-ChemScore). In order to evaluate these four procedures, the MMP3-inhibitor complex 1B3D and the MMP8-inhibitor complex 1A85 were selected as reference structures. All the other X-ray complexes were aligned to them through the superimposition of the MMP backbone atoms. In this manner, it was possible to obtain an experimental reference position for each of the 14 ligands into the 14 X-ray crystal structures.

The docking reliabilities were then evaluated by calculating the RMSD (heavy atoms) between the above-mentioned experimental reference position and that calculated by Gold for each ligand and each of the 14 X-ray complexes.

4.3. 3D-QSAR model

4.3.1. Alignment of the molecules

Sixty-nine compounds were docked inside the MMP2 (1QIB[34] pdb code), MMP3 (1D8F[35]), and MMP9 (1GKC[36]) binding sites using the SMC-GoldScore and SMC-ChemScore fitness functions. For each ligand, the best docked structure was chosen, and this receptor-based alignment was used for further studies.

4.3.2. Data set

The GOLPE program[30] was used to define a 3D-QSAR model, using GRID interaction fields[37] as descriptors (see below). The training set was composed of the same 54 compounds for all the three MMPs. These compounds were characterized by activity values spanning five orders of magnitude. Similarly, the 15 compounds belonging to the external test set showed pIC_{50} values ranging from 5 to 10, and were chosen in a manner that allowed uniform distribution along the activity range for all the three MMPs.

4.3.3. Probe selection

The GRID program[37] was used to describe the previously superimposed molecular structures. Interaction energies between the selected probes and each molecule were calculated using a grid spacing of 1 Å. Initially, many probes were tested, and the preliminary PLS analyses suggested that the use of a combination of C3 (corresponding to a methyl group) and N2 (corresponding to neutral NH_2) probes, using the alignment originating from the SMC-GoldScore method, best described the system.

4.3.4. Variable selection

The molecular interaction fields (MIFs) of the training set were imported into the GOLPE program. It is well known that many of the variables derived from the GRID analyses can be considered noise, which decreases the quality of the model. For this reason, variable selection was conducted by zeroing values with absolute values smaller than 0.06 kcal/mol, and removing variables with a standard deviation below 0.1. Further, variables which exhibited only two values and had a skewed distribution were also removed.

The smart region definition (SRD) algorithm[38] was applied with 10% of the active variables as the seed number (selected in the PLS weights space), a critical distance cutoff of 2.5 Å, and a collapsing distance cutoff of 4.0 Å. The groups were then used in the fractional factorial design (FFD) procedure. FFD selection was applied until the r^2 and q^2 values did not increase significantly, using the cross-validation routine with five random sets of compounds.

Acknowledgments

Financial support provided by Fondazione Monte dei Paschi di Siena ('Strumenti terapeutici innovativi per la modulazione dell'attività di metalloproteasi della matrice implicate nelle patologie tumorali cerebrali') is gratefully acknowledged. Many thanks are due to Prof. Gabriele Cruciani and Prof. Sergio Clementi (Molecular Discovery and MIA srl) for the use of the GOLPE program in their chemometric laboratory (University of Perugia, Italy) and for having provided the GRID program.

Supplementary data

Actual versus predicted data of the receptor-based 3D-QSAR models (Tables 1), the analysis of the S1' pocket of MMP2, MMP3 and MMP9 (Figure 7). This material is available free of charge via the Internet. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.07.004.

References and notes

1. Nagase, H.; Woessner, J. F., Jr. *J. Biol. Chem.* **1999**, *274*, 21491.
2. Whittaker, M.; Floyd, C. D.; Brown, P.; Gearing, A. J. *Chem. Rev.* **1999**, *99*, 2735.
3. Rush, T. S.; Powers, R. *Curr. Top. Med. Chem.* **2004**, *4*, 1311.
4. Coussens, L. M.; Fingleton, B.; Matrisian, L. M. *Science* **2002**, *295*, 2387.
5. Marcq, V.; Mirand, C.; Decarme, M.; Emonard, H.; Hornebeck, W. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2843.
6. Skiles, J. W.; Gonnella, N. C.; Jeng, A. Y. *Curr. Med. Chem.* **2004**, *11*, 2911.
7. Peterson, J. T. *Cardiovasc. Res.* **2006**, *69*, 677.
8. Augé, F.; Hornebeck, W.; Decarme, M.; Laronze, J. Y. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1783.
9. Pavlaki, M.; Zucker, S. *Cancer Metastasis Rev.* **2003**, *22*, 177.
10. Björklund, M.; Koivunen, E. *Biochim. Biophys. Acta* **2005**, *1755*, 37.
11. Overall, C. M.; Kleifeld, O. *Nat. Rev. Cancer* **2006**, *6*, 227.
12. Itoh, T.; Tanioka, M.; Yoshida, H.; Yoshioka, T.; Nishimoto, H.; Itohara, S. *Cancer Res.* **1998**, *58*, 1048.
13. Tester, A. M.; Waltham, M.; Oh, S. J.; Bae, S. N.; Bills, M. M.; Walzer, E. C.; Kern, F. G.; Stetler-Stevenson, W. G.; Lippman, M. E.; Thompson, E. W. *Cancer Res.* **2004**, *64*, 652.
14. Fingleton, B. *Semin. Cell Dev. Biol.* **2008**, *19*, 61.
15. López-Otín, C.; Matrisian, L. M. *Nat. Rev. Cancer* **2007**, *7*, 800.
16. Coussens, L. M.; Tinkle, C. L.; Hanahan, D.; Werb, Z. *Cell* **2000**, *103*, 481.
17. Hamano, Y.; Zeisberg, M.; Sugimoto, H.; Lively, J. C.; Maeshima, Y.; Yang, C.; Hynes, R. O.; Werb, Z.; Sudhakar, A.; Kalluri, R. *Cancer Cell* **2003**, *3*, 589.
18. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
19. Tuccinardi, T.; Nuti, E.; Ortore, G.; Supuran, C. T.; Rossello, A.; Martinelli, A. *J. Chem. Inf. Model.* **2007**, *47*, 515.
20. Tuccinardi, T.; Ortore, G.; Rossello, A.; Supuran, C. T.; Martinelli, A. *J. Chem. Inf. Model.* **2007**, *47*, 2253.
21. Yamamoto, S.; Nakatani, S.; Ikura, M.; Sugiura, T.; Nishita, Y.; Itadani, S.; Ogawa, K.; Ohno, H.; Takahashi, K.; Nakai, H.; Toda, M. *Bioorg. Med. Chem.* **2006**, *14*, 6383.
22. Becker, D. P.; Villamil, C. I.; Barta, T. E.; Bedell, L. J.; Boehm, T. L.; Decrescenzo, G. A.; Freskos, J. N.; Getman, D. P.; Hockerman, S.; Heintz, R.; Howard, S. C.; Li, M. H.; McDonald, J. J.; Carron, C. P.; Funckes-Shippy, C. L.; Mehta, P. P.; Munie, G. E.; Swearingen, C. A. *J. Med. Chem.* **2005**, *48*, 6713.
23. Hanessian, S.; MacKay, D. B.; Moitessier, N. *J. Med. Chem.* **2001**, *44*, 3074.
24. Nakatani, S.; Ikura, M.; Yamamoto, S.; Nishita, Y.; Itadani, S.; Habashita, H.; Sugiura, T.; Ogawa, K.; Ohno, H.; Takahashi, K.; Nakai, H.; Toda, M. *Bioorg. Med. Chem.* **2006**, *14*, 5402.
25. Hanessian, S.; Moitessier, N.; Cantin, L. D. *Tetrahedron* **2001**, *57*, 6885.
26. Rossello, A.; Nuti, E.; Carelli, P.; Orlandini, E.; Macchia, M.; Nencetti, S.; Zandomeneghi, M.; Balzano, F.; Uccello-Barretta, G.; Albini, A.; Benelli, R.; Cercignani, G.; Murphy, G.; Balsamo, A. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1321.
27. Ikura, M.; Nakatani, S.; Yamamoto, S.; Habashita, H.; Sugiura, T.; Takahashi, K.; Ogawa, K.; Ohno, H.; Nakai, H.; Toda, M. *Bioorg. Med. Chem.* **2006**, *14*, 4241.
28. Chollet, A. M.; Le Diguarher, T.; Murray, L.; Bertrand, M.; Tucker, G. C.; Sabatini, M.; Pierre, A.; Atassi, G.; Bonnet, J.; Casara, P. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 295.
29. Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T. *J. Med. Chem.* **2000**, *43*, 3033.

30. GOLPE 4.5. Multivariate Infometric Analysis Srl., Viale dei Castagni 16, Perugia, Italy, 1999.
31. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
32. Maestro, ver. 7.5; Schrödinger Inc.: Portland, OR, 1999.
33. Macromodel, ver. 8.5; Schrödinger Inc.: Portland, OR, 1999.
34. Dhanaraj, V.; Williams, M. G.; Ye, Q.-Z.; Molina, F.; Johnson, L. L.; Ortwine, D. F.; Pavlovsky, A.; Rubin, J. R.; Skeean, R. W.; White, A. D.; Humblet, C.; Hupe, D. J.; Blundell, T. L. *Croat. Chem. Acta* **1999**, *72*, 575.
35. Cheng, M.; De, B.; Pikul, S.; Almstead, N. G.; Natchus, M. G.; Anastasio, M. V.; McPhail, S. J.; Snider, C. E.; Taiwo, Y. O.; Chen, L.; Dunaway, C. M.; Gu, F.; Dowty, M. E.; Mieling, G. E.; Janusz, M. J.; Wang-Weigand, S. *J. Med. Chem.* **2000**, *43*, 369.
36. Rowsell, S.; Hawtin, P.; Minshull, C. A.; Jepson, H.; Brockbank, S. M.; Barratt, D. G.; Slater, A. M.; McPheat, W. L.; Waterson, D.; Henney, A. M.; Pauptit, R. A. *J. Mol. Biol.* **2002**, *319*, 173.
37. Cruciani, G.; Goodford, P. Copyright Molecular Discovery Ltd 2001–2003 GREATER graphical interface for GRID version, 1.1.7 GRIB, UPF/IMIM, Barcelona (Spain) http://www.moldiscovery.com.
38. Pastor, M.; Cruciani, G.; Clementi, S. *J. Med. Chem.* **1997**, *40*, 1455.