A structure-activity relationship study of catechol-*O*-methyltransferase inhibitors combining molecular docking and 3D QSAR methods

Anu J. Tervo[a,b,*], Tommi H. Nyrönen[b], Toni Rönkkö[a] & Antti Poso[a]
[a]*Department of Pharmaceutical Chemistry, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland;*
[b]*CSC – Scientific Computing Ltd., P.O. Box 405, 02101 Espoo, Finland*

Summary

A panel of 92 catechol-*O*-methyltransferase (COMT) inhibitors was used to examine the molecular interactions affecting their biological activity. COMT inhibitors are used as therapeutic agents in the treatment of Parkinson's disease, but there are limitations in the currently marketed compounds due to adverse side effects. This study combined molecular docking methods with three-dimensional structure-activity relationships (3D QSAR) to analyse possible interactions between COMT and its inhibitors, and to incite the design of new inhibitors. Comparative molecular field analysis (CoMFA) and GRID/GOLPE models were made by using bioactive conformations from docking experiments, which yielded q^2 values of 0.594 and 0.636, respectively. The docking results, the COMT X-ray structure, and the 3D QSAR models are in agreement with each other. The models suggest that an interaction between the inhibitor's catechol oxygens and the Mg^{2+} ion in the COMT active site is important. Both hydrogen bonding with Lys144, Asn170 and Glu199, and hydrophobic contacts with Trp38, Pro174 and Leu198 influence inhibitor binding. Docking suggests that a large R_1 substituent of the catechol ring can form hydrophobic contacts with side chains of Val173, Leu198, Met201 and Val203 on the COMT surface. Our models propose that increasing steric volume of e.g. the diethylamine tail of entacapone is favourable for COMT inhibitory activity.

Introduction

Catechol-*O*-methyltransferase (COMT [1]) is an enzyme that catalyses the transfer of a methyl group from the cofactor *S*-adenosylmethionine (SAM) to a hydroxylate oxygen on the catechol or substituted catechol substrate [2]. COMT is important in both the peripheral and central nervous systems, where it metabolises various catechols, particularly dopamine, adrenaline and noradrenaline [3]. From a pharmaceutical point of view, an important substrate for COMT is levodopa, which remains the cornerstone in treating Parkinson's disease [3]. Inhibition of COMT by various inhibitors slows down the metabolism of levodopa and thereby enables a smaller-dose therapy with fewer side-effects [4, 5].

Recently developed COMT inhibitors include entacapone [6–8] (OR-611, (*E*)-2-cyano-3-(3,4-dihydroxy-5-nitro-phenyl)-*N*,*N*-diethyl-acrylamide, **43**) and tolcapone [8–10] (Ro-40-7592, (3,4-dihydroxy-5-nitrophenyl)-*p*-tolyl-methanone, **78**) (Figure 1). There are, however, some limitations in using these COMT inhibitors as therapeutic agents. Hepatic dysfunction, for example, may develop in association with tolcapone therapy [11]. Therefore, further development of COMT inhibitors is still needed.

With three-dimensional structure-activity relationship (3D QSAR) analysis it is possible to analyse the probable structural elements affecting the biological activity of compounds. The information can be used e.g. in designing new bioactive compounds. Taskinen et al. [12] and Lotta et al. [13] previously carried out structure-activity relationship studies of COMT inhibitors using quantum chemical molecular descriptors in their studies, but there are no previous 3D QSAR stud-

*To whom correspondence should be addressed. Tel: +358-9-4572228; Fax: +358-9-4572302; E-mail: anu.tervo@csc.fi.
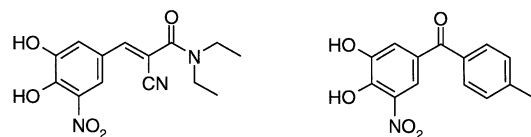
43 (Entacapone, OR-611)   **78** (Tolcapone, Ro-40-7592)

Figure 1. The structures of entacapone and tolcapone.

ies of COMT inhibitors. In this paper, we report 3D QSAR studies from a panel of 92 COMT inhibitors obtained from literature [13].

Several different approaches have been developed for 3D QSAR analysis. One of the most widely used procedures, and also the one applied in the present study, is comparative molecular field analysis (CoMFA) [14]. CoMFA produces information about the favoured and disfavoured properties of the examined compound set. The crucial steps in a CoMFA procedure are in determining the bioactive conformation of the examined compounds and the subsequent superpositioning of these conformations [14]. The X-ray crystallographic structure of COMT in the Protein Data Bank [15] (PDB ID: 1vid [16]) offers the possibility to make a molecular superposition by using the bioactive conformation of co-crystallized 1,5-dinitrocatechol as a template, or to dock all inhibitors into the active site of COMT. Co-crystallized 1,5-dinitrocatechol, however, does not give information about the bioactive conformation of catechols having large substituents. Thus, our approach was to obtain bioactive conformations of an inhibitor panel by automated molecular docking, and to use 1,5-dinitrocatechol to verify the docked position by the catechol ring of each inhibitor. Several examples in the literature indicate that molecular docking followed by 3D QSAR analysis can yield explanative models [17–19]. Automated molecular docking also gives information about the possible binding mechanisms of the inhibitors, and gives an estimation of their binding affinities [20].

The interaction between candidate inhibitors and the Mg^{2+} ion in the active site of COMT is important for the inhibitory activity [16]. Therefore, another 3D QSAR model was created for examining this interaction by using GRID/GOLPE [21, 22]. This procedure allows more possibilities to examine interaction types that may be important for biological activity, and allows a more accurate statistical treatment of the data compared to CoMFA.

Materials and methods

Data preparation

Inhibitors used for modeling were obtained from Lotta et al. [13], excluding compounds without known isomers. The structures of the 92 used inhibitors are presented in Table 1. The reported biological activity of the inhibitors (IC_{50}) was measured as the potency of a given compound to inhibit rat brain soluble COMT (S-COMT) *in vitro* [13] as has been described by Taskinen et al. [12].
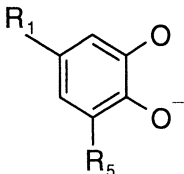
All modeling except GRID/GOLPE was performed using Sybyl 6.8 [23] molecular modeling software. Generation of the inhibitor set proceeded as follows. Inhibitors were built in their anionic forms, because most of them are ionised at pH 7.8 where their biological activities were measured [12]. 3D structures were generated using CORINA [24, 25]. CORINA was allowed to generate multiple conformations for inhibitors having flexible ring systems, using a maximum energy difference of 20 kJ/mol between the created conformations. By using this energy difference only a single conformation was created for each inhibitor, except inhibitor **21** (4 conformations) and **42**, **84** and **89** (2 conformations).

The X-ray structure of COMT co-crystallised with 1,5-dinitrocatechol (compound **20**) was retrieved from the Protein Data Bank [15] (PDB ID: 1vid [16]). 1,5-Dinitrocatechol and all water molecules except the water molecule which completes the octahedral coordination geometry of the Mg^{2+} ion [16] (Wat400 in 1vid) were removed from the protein structure. After adding hydrogens, the protein was minimised for 200 cycles using the Steepest Descent method and the Tripos force field [26] to remove possible bad geometries. The position of the Mg^{2+} ion was constrained with respect to atoms that belonged to the octahedral coordination in the crystal structure. During minimization, the distance of Mg^{2+} was constrained to 2.2 Å from the OD1 atom on the side chain of Asp 141, 2.2 Å from OD2 on the side chain of Asp169 and 2.3 Å from OD1 on the side chain of Asn170. The position of the water molecule was also frozen during minimization.

Molecular docking

All inhibitors were docked into the active site of COMT to generate the docked conformations using the molecular docking program FlexX 1.11.1 [27] in Sybyl. The active site of COMT was defined to contain all atoms within a 15 Å radius from Mg^{2+}, including

Table 1. Structures of the inhibitors.

Compd	R$_1$	R$_5$	Compd	R$_1$	R$_5$
1	(CH$_2$)COOH	NO$_2$	47	CHC(COCH$_3$)CO-phenyl	NO$_2$
2	CHCHCO-phenyl	NO$_2$	48	CO-(2-pyridyl)	NO$_2$
3	(CH$_2$)$_4$CONH-(1-adamantyl)	NO$_2$	49	5-methylene-2-thioxo-imidazolidin-4-one	NO$_2$
4	(CH$_2$)$_4$CONHCH(CH$_3$)$_2$	NO$_2$	50	CHC(CN)CONHC(CH$_3$)$_3$	NO$_2$
5	CHC(CH$_3$)COCH$_3$	NO$_2$	51	CHC(CN)CO-(1-piperidyl)	NO$_2$
6	CHC(CN)$_2$	NO$_2$	52	5-methylene-2-thioxo-thiazolidin-4-one	NO$_2$
7	CHCHCO-(3,4,5-trimethoxyphenyl)	NO$_2$	53	CHC(CN)CON(C$_3$H$_7$)$_2$	NO$_2$
8	CONH-(1-adamantyl)	NO$_2$	54	5-methylene-pyrimidine-2,4,6-trione	NO$_2$
9	CHCHNO$_2$	NO$_2$	55	CO-(4-pyridyl)	NO$_2$
10	CH$_2$CH(CN)$_2$	NO$_2$	56	CHC(COCH$_3$)CON(CH$_3$)$_2$	NO$_2$
11	CHCH-(4-pyridyl)	NO$_2$	57	CHC(COCH$_3$)CON(C$_2$H$_5$)$_2$	NO$_2$
12	CHCH-(4-quinolinyl)	NO$_2$	58	CHC(CN)-(2-pyridyl)	NO$_2$
13	CHO	NO$_2$	59	CO-(2-fluorophenyl)	NO$_2$
14	CHC(COCH$_3$)$_2$	NO$_2$	60	CHC(COCH$_3$)CONH-1-adamantyl	NO$_2$
15	COCHCH-(4-(dimethylamino)phenyl)	NO$_2$	61	CHC(CN)CO-(3,4,5-trimethoxyphenyl)	NO$_2$
16	CHCHCO-(2-hydroxyphenyl)	NO$_2$	62	CHC(COCH$_2$CH$_3$)$_2$	NO$_2$
17	CN	NO$_2$	63	CHC(CN)-phenylsulfonyl	NO$_2$
18	(CH$_2$)$_4$COO(CH$_2$)$_{13}$CH$_3$	NO$_2$	64	(CH$_2$)$_4$COOH	Cl
19	CHCHCO-(4-methylphenyl)	NO$_2$	65	CONH-(1-adamantyl)	Cl
20	NO$_2$	NO$_2$	66	COOH	CN
21	CO-(4-(cyclohexylcarbonyl)-1-piperidyl)	NO$_2$	67	COO(CH$_2$)$_3$CH$_3$	CN
22	CONHCH$_2$-phenyl	NO$_2$	68	CHCHCO-phenyl	CF$_3$
23	CHCHCO-(4-methoxyphenyl)	NO$_2$	69	CHO	CN
24	CHCHCO-(4-nitrophenyl)	NO$_2$	70	CHO	CHO
25	CHCHCO-(2-carboxyphenyl)	NO$_2$	71	CHO	CF$_3$
26	CHCHCO-(4-chlorophenyl)	NO$_2$	72	CHC(COCH$_3$)$_2$	CF$_3$
27	CHCHCO-(3,4-dichlorophenyl)	NO$_2$	73	CONHCH$_2$-phenyl	CN
28	CHCHCO-(4-hydroxy-3-methoxyphenyl)	NO$_2$	74	Cl	SO$_2$N-(CH$_2$CH$_3$)$_2$
29	CHCHCO-(4-hydroxyphenyl)	NO$_2$	75	CN	CN
30	CHC(CN)COOCH$_2$CH$_3$	NO$_2$	76	Cl	SO$_2$CH$_3$
31	CHC(COCH$_3$)CH$_2$COOCH$_3$	NO$_2$	77	NO$_2$	F
32	CONHCH$_2$-(2-pyridyl)	NO$_2$	78	CO-(4-methylphenyl)	NO$_2$
33	CONHCH$_2$-(2,4,6-trimethoxyphenyl)	NO$_2$	79	COCH(CH$_3$)$_2$	H
34	CH$_2$OH	NO$_2$	80	COOCH$_2$CH$_2$CH$_3$	OH
35	CH$_2$OCH$_2$CH$_2$OCH$_3$	NO$_2$	81	CHCHOOH	NO$_2$
36	CH$_2$SCH$_2$COOH	NO$_2$	82	(CH$_2$)$_4$CON(CH$_3$)CH$_2$CCH	NO$_2$
37	CH$_2$-(2-pyrrolyl)	NO$_2$	83	COCH$_3$	NO$_2$
38	H	NO$_2$	84	(CH$_2$)$_4$CO-(4-benzyl-1-piperazinyl)	NO$_2$
39	Cl	NO$_2$	85	CHCHCO-(3,4-dimethoxyphenyl)	NO$_2$
40	CHC(CN)CON(CH$_3$)$_2$	NO$_2$	86	CHC(CN)CONH$_2$	NO$_2$
41	CHC(CN)CONHCH(CH$_3$)$_2$	NO$_2$	87	CHC(CN)COOCH$_2$C(CH$_3$)$_3$	NO$_2$
42	CHC(CN)CO-(4-methyl-1-piperazinyl)	NO$_2$	88	CHC(COCH$_3$)CONHCH$_2$CH$_2$OH	NO$_2$
43	CHC(CN)CON(CH$_2$CH$_3$)$_2$	NO$_2$	89	CHC(COCH$_3$)CONH-cyclohexyl	NO$_2$
44	CHC(CN)CONH(CH$_2$)$_4$OH	NO$_2$	90	CONH-(1-adamantyl)	CN
45	CONH(CH$_2$)$_3$OH	NO$_2$	91	CHO	SO$_2$CH$_3$
46	COOH	NO$_2$	92	CHC(CN)CON(CH$_2$CH$_3$)$_2$	NO$_2$

SAM as a part of the enzyme active site structure. As the active site of COMT is an open cavity situated on the enzyme surface, a relatively large active site radius was chosen for the docking simulations in order to include all residues that could interact with the inhibitors. Because the inhibitors were examined in their anionic forms, a formal charge of -1 was assigned to all inhibitors. All conformations of the inhibitors **21, 42, 84** and **89** were docked separately. The docking procedure was repeated 30 times for each inhibitor. The chosen docked conformation for each inhibitor was selected from the docking results on the basis of their FlexX total energy score. The chosen conformation also had to fulfil the catechol ring position criterion obtained from the original co-crystallised 1,5-dinitrocatechol. The docked inhibitors were further minimized for 100 steps with the Steepest Descent method using the Tripos force field [26] and Gasteiger-Hückel [28, 29] charges in the active site of COMT, while keeping the enzyme rigid and allowing only the inhibitor to relax. After minimization, MOPAC [30] 6.0 MNDO/ESP charges [31] that were needed for 3D QSAR analysis were calculated for all inhibitors.

3D QSAR analyses

3D QSAR models were built using both CoMFA [14] and GRID/GOLPE [21, 22] procedures. CoMFA was used to explore the contributors of steric and electrostatic effects to inhibitory activity, and GRID/GOLPE was used to explore the interaction mode between inhibitors and Mg^{2+}. Both models were built using the docked conformations and inhibitor alignment obtained from molecular docking.

All 92 inhibitors were utilized in the model building. In both models, pIC_{50} activity values were used as dependent variables in a partial least squares (PLS) [32] statistical analysis of the resulting data. The predictive value of the PLS models was validated using five random groups cross-validation method. In this method, 20% of the inhibitors are randomly excluded from the inhibitor set, and their activities are predicted with a model derived from the rest of the inhibitors. In the absence of an external test set, a random groups cross-validation using two equal-sized groups gives a good estimate of model predictivity [33]. Therefore, cross-validated models using two random groups were also built.

CoMFA. CoMFA descriptor fields were calculated in a 3D cubic box extending 4 Å beyond the aligned inhibitors. Steric and electrostatic fields were calculated using an sp3 carbon probe atom with a $+1$ charge and a 2 Å grid spacing. The Tripos Standard field was used with a cutoff of 30.0 kcal/mol for both steric and electrostatic fields. The random groups cross-validations were carried out using both five and two as a number of randomly selected groups. A minimum σ value (standard deviation threshold) of 2.0 kcal/mol was utilized to improve the signal-noise ratio. Both cross-validation procedures were repeated 25 times, and average statistical values were calculated. The final non-cross-validated model was developed using the optimal number of components that had both the highest q^2 value and the smallest value of standard error of predictions (Spress) in both cross-validation procedures.

GRID/GOLPE. The interaction between inhibitors and the Mg^{2+} ion in the COMT active site was of particular interest. Thus, a series of GRID [21] calculations were performed using Mg^{2+} and hydrophobic DRY probes. GRID predicts interactions between a molecule and a small functional group (i.e. a probe). Several probes were tested as a combination with Mg^{2+}, and the DRY was selected as a complementary probe on the basis of q^2 and Spress values. Calculations were performed using GRID version 21 with GREATER interface. Interaction fields were calculated between all inhibitors, and both Mg^{2+} and hydrophobic DRY probes. A grid spacing of 0.5 Å in a 3D cubic box extending 5 Å beyond the aligned inhibitors was utilized with a default cutoff value of $+5$ kcal/mol.

Generated GRID interaction fields were analysed by GOLPE [22]. Initially, 176,105 active X-variables were obtained from the calculated fields. However, a major part of these variables were not important for describing an interaction between the inhibitors and the probes. Therefore, an advanced data pre-treatment and Smart Region Definition (SRD [34])/F. Factorial Selection procedures in GOLPE version 4.5.12 were used to reduce the number of X-variables in the final 3D QSAR model.

Advanced data pre-treatment was used to eliminate variables that might cause either noise or inadequate weighting in the 3D QSAR model building. Positive values smaller than 0.10 and negative values larger than -0.01 were zeroed to diminish the amount of noise in the data. The minimum X variable cutoff value was 0.10 (corresponding minimum σ value in CoMFA). Finally, all 2-level and 3-level variables, which might have forced the model to explain the vari-

ance of a few objects having a high leverage, were removed. After the advanced pre-treatment, a total of 42,411 active X-variables remained.

The reliability of the model was checked after the advanced data pre-treatment to assure that the model was good enough to allow the following SRD and F. Factorial Selection procedures. SRD was utilized to select groups of variables that were important in explaining interactions, instead of only using single variables. Groups were generated using a maximum dimensionality of 2, along with a critical distance cutoff value of 1.0 Å and a collapsing distance cutoff value of 2.0 Å. The calculated groups were then used in a F. Factorial Selection procedure using a maximum dimensionality of two and the random groups method, with five groups of randomly chosen inhibitors and 25 Spress calculations as a validation mode. The number of active X-variables was thereby reduced to 3,676.

PLS statistical analysis was used to correlate the biological activity values with variations in the interaction fields. The random groups cross-validations were carried out using both five and two as a number of randomly selected groups. Both cross-validation procedures were repeated 25 times. The final model was the one giving both the highest q^2 value and the smallest Spress for both procedures.

Results and discussion

Docking and alignment of the inhibitors

The automated molecular docking procedure created 30 docked conformations for each inhibitor. The docked conformation for each inhibitor was chosen on the basis of the FlexX total energy score and the position criterion obtained from COMT co-crystallised with 1,5-dinitrocatechol. The failed docked conformations could be easily identified by using the co-crystallised 1,5-dinitrocatechol, which showed the catechol ring position in the COMT binding site. FlexX total energy score estimated the binding free energy of a docked inhibitor [20]. The docking result giving the best FlexX total energy score was consistent with the co-crystallised 1,5-dinitrocatechol for all inhibitors except **82**. However, FlexX also found a binding mode for **82** that was consistent with the crystal structure, which was only 0.44 kcal/mol higher in energy than the best ranking conformation. The docking result yielding the best total score among the separate starting conformations of **21**, **42**, **84** and **89** was chosen as the representative docked conformation.
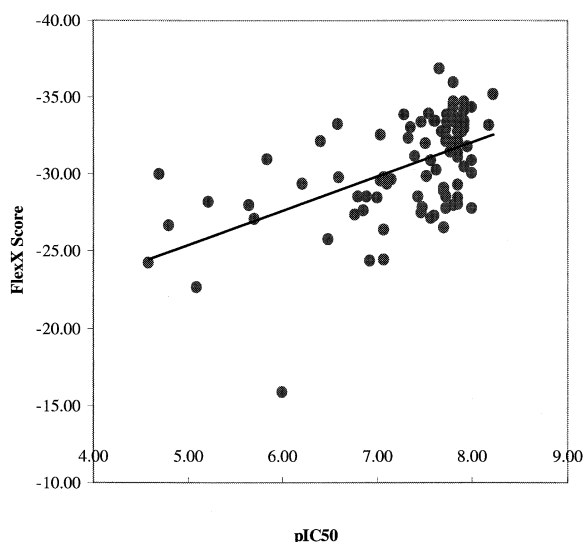

Figure 2. Correlation between the experimental activities (pIC$_{50}$) and FlexX total energy scores (kcal/mol).
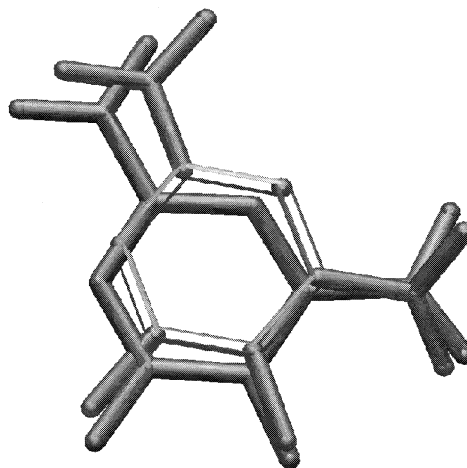
Figure 3. Comparison between 1,5-dinitrocatechol from the X-ray structure 1vid and the docked 1,5-dinitrocatechol (**20**, with transparent aromatic ring).

Interest was focused on how well the calculated FlexX total energy scores correlated with the experimental biological activity. FlexX total energy scores (Table 2) for the chosen conformations were compared with the corresponding pIC$_{50}$ value using the standard linear r^2 value. A value of $r^2 = 0.2994$ was obtained for the set of all 92 inhibitors (Figure 2). This indicates a weak correlation, although some correlation can be visually observed from the plot profile. The free energy of binding calculated with FlexX does not correlate with the pIC$_{50}$ values of inhibitors **12**, **18** and **91**, and the inhibitors emerge as outliers in Figure 2.

Table 2. Experimental activities, FlexX total energy scores of the inhibitors and RMSD values compared with 1,5-dinitrocat-echol from X-ray structure 1vid.

Compd	pIC$_{50}$	FlexX[a]	RMSD	Compd	pIC$_{50}$	FlexX[a]	RMSD
1	7.07	−24.50	0.256	**47**	7.80	−36.00	0.273
2	8.22	−35.20	0.274	**48**	7.54	−34.00	0.273
3	7.57	−27.20	0.272	**49**	7.72	−28.60	0.140
4	7.48	−27.90	0.271	**50**	7.92	−33.00	0.275
5	7.92	−33.30	0.260	**51**	7.85	−31.40	0.445
6	7.70	−29.10	0.273	**52**	7.85	−28.50	0.140
7	8.18	−33.20	0.274	**53**	7.85	−28.10	0.280
8	7.72	−32.90	0.273	**54**	6.60	−29.80	0.270
9	7.40	−31.20	0.146	**55**	7.80	−33.50	0.273
10	7.00	−28.50	0.274	**56**	7.46	−33.40	0.278
11	7.80	−34.70	0.272	**57**	7.57	−30.90	0.277
12	7.66	−36.90	0.143	**58**	7.92	−33.50	0.279
13	7.62	−30.30	0.272	**59**	7.80	−33.60	0.142
14	7.68	−32.80	0.275	**60**	7.51	−32.00	0.275
15	7.92	−33.50	0.266	**61**	7.77	−31.50	0.272
16	7.85	−33.70	0.274	**62**	7.85	−31.10	0.275
17	7.52	−29.90	0.271	**63**	7.95	−31.80	0.275
18	6.00	−15.90	0.137	**64**	5.10	−22.70	0.086
19	7.82	−33.40	0.275	**65**	6.40	−32.20	0.087
20	7.92	−30.50	0.256	**66**	5.70	−27.10	0.075
21	7.70	−28.90	0.436	**67**	7.72	−27.80	0.073
22	7.74	−33.90	0.280	**68**	6.59	−33.30	0.050
23	7.89	−32.80	0.276	**69**	6.80	−28.60	0.074
24	7.92	−34.70	0.439	**70**	6.89	−28.60	0.068
25	7.80	−34.60	0.274	**71**	5.65	−28.00	0.052
26	7.92	−34.20	0.440	**72**	5.84	−31.00	0.073
27	7.60	−33.50	0.440	**73**	7.35	−33.10	0.075
28	7.80	−34.40	0.275	**74**	4.59	−24.30	0.142
29	7.85	−32.20	0.440	**75**	7.46	−27.50	0.074
30	7.85	−29.30	0.273	**76**	4.80	−26.70	0.127
31	7.80	−28.00	0.277	**77**	6.48	−25.80	0.235
32	7.28	−33.90	0.281	**78**	7.61	−33.50	0.140
33	7.82	−33.40	0.273	**79**	5.22	−28.20	0.075
34	6.76	−27.40	0.453	**80**	5.70	−27.10	0.088
35	7.07	−26.40	0.274	**81**	7.43	−28.60	0.274
36	6.92	−24.40	0.271	**82**	7.70	−26.57	0.138
37	7.15	−29.70	0.274	**83**	7.80	−32.10	0.273
38	6.85	−27.70	0.264	**84**	7.07	−29.80	0.273
39	7.60	−27.30	0.158	**85**	8.00	−34.40	0.275
40	7.74	−33.40	0.277	**86**	7.74	−32.30	0.274
41	7.85	−32.80	0.275	**87**	8.00	−30.10	0.274
42	7.72	−32.20	0.278	**88**	7.04	−29.60	0.276
43	8.00	−30.90	0.278	**89**	7.33	−32.40	0.275
44	8.00	−27.80	0.273	**90**	7.04	−32.60	0.073
45	7.10	−29.40	0.274	**91**	4.70	−30.00	0.087
46	6.21	−29.40	0.273	**92**	7.70	−29.10	0.274

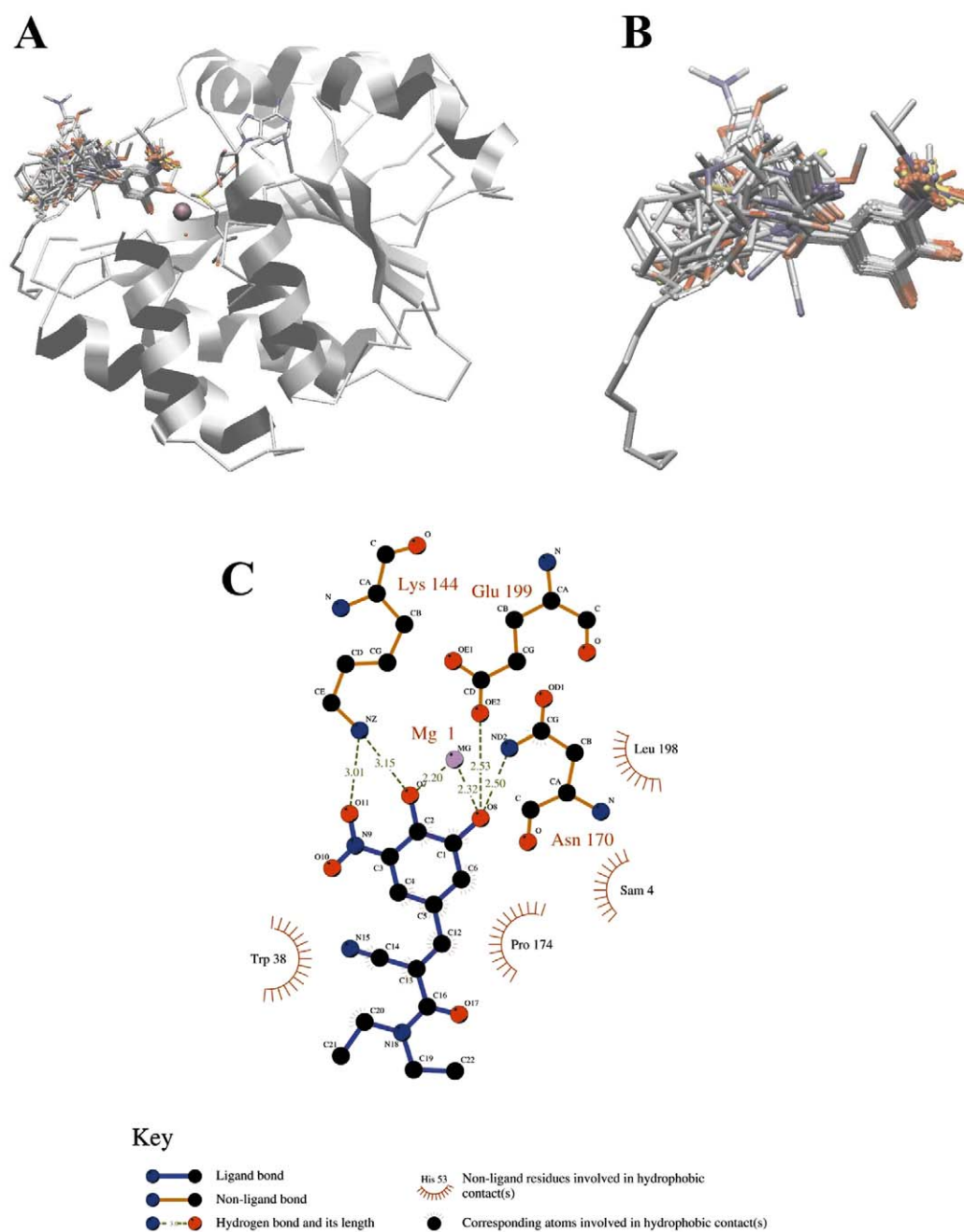[a] FlexX total energy score in kcal/mol.

Figure 4. (A) Location of the docked inhibitors, Mg^{2+} ion, Wat400 and SAM in the active site of COMT. (B) Predicted binding conformations of COMT inhibitors and their alignment. (A) and (B) are rendered by the BODIL [37] program, and hydrogens are omitted for clarity. (C) Proposed interaction model of inhibitor **43** in the active site of COMT, drawn using the LIGPLOT [36] program.

The outliers may have negative consequences for the 3D QSAR model performance statistics. In general, inhibitors having similar pIC_{50} values and still slightly different molecular structures had large differences in FlexX binding free energy when compared to differences in their experimentally determined biological activity.

The X-ray structure of COMT co-crystallised with 1,5-dinitrocatechol was compared with the docking of 1,5-dinitrocatechol (**20**) by FlexX (Figure 3). The heavy atom root mean square deviation (RMSD) between the docked and co-crystallised 1,5-dinitrocatechol conformations was 0.256 Å, which is considered to be a good reproduction of the 1,5-dinitrocatechol X-ray structure, particularly when the X-ray structure is resolved at a resolution of 2.00 Å [35]. Moreover, FlexX successfully predicted the important interaction between the catechol oxygens and Mg^{2+}.

All inhibitors were docked in the active site of COMT similarly to the co-crystallised 1,5-dinitrocatechol (Figure 4). To evaluate the docking results, the heavy atom RMSD was calculated between the largest common fragment of the inhibitors and the co-crystallised 1,5-dinitrocatechol (Table 2). The average RMSD value of the superimposed largest common fragments of the inhibitor set was 0.237 Å, which indicates a good overall alignment of the catechol rings. COMT complexed with 1,5-dinitrocatechol helped verify the predicted interaction between Mg^{2+} and the catechol ring oxygens in the inhibitor set.

Interaction model of the docked inhibitors

Figure 4C represents the interaction model of the docked inhibitor **43** (entacapone) with COMT, as generated with the program LIGPLOT [36]. The interaction between the Mg^{2+} ion and both catechol oxygens of **43** was of major importance. The anionic catechol oxygen of **43** forms a hydrogen bond with one of the hydrogens of the protonated NZ nitrogen on Lys144. The catechol hydroxyl group acts both as a hydrogen bond donor to oxygen OE2 of Glu199 and as a hydrogen bond acceptor from the amino group of Asn170 (hydrogen of ND2). Finally, the oxygen in the nitro group of **43** at position R_5 forms a hydrogen bond with hydrogen of the protonated NZ nitrogen of Lys144. Although the importance of the interaction with Mg^{2+} is already known [16], the multiple hydrogen bonding between the active site and the inhibitor must also have an effect on activity. Taskinen et al. [12] previ-

ously found that the electron-withdrawing R_1 and R_5 substituents increase ionisation of the catechol 4-OH and favour the inhibitory effect. Inhibitors with good biological activity have a R_5 substituent that is capable of forming hydrogen bonds, and therefore we suggest that this structural feature enhances activity.

The active site of COMT is sterically not very restricted. However, hydrophobic parts of both Pro174 and Trp38 sterically limit the volume available for the skeleton of **43** (Figure 4C). Amino acids Val173, Leu198, Met201 and Val203 on the enzyme surface can form hydrophobic contacts with bulky R_1 substituents. The methyl group of SAM also restricts the active site near the R_5 substituent of the inhibitor. This could be the reason for the experimentally observed trend that bulky R_5 groups reduce biological activity [13]. Leu198 and residues involved in hydrogen bonding with the inhibitor (Lys144, Asn170, Glu199) restrict the volume available for the catechol ring. The placement of the catechol ring was well defined in the COMT active site by these interactions.

3D QSAR models

3D QSAR models were created to obtain information about the structural elements and properties affecting the activity of the inhibitors. The chosen conformations of the inhibitors play a crucial role in the 3D QSAR model building. Therefore entire structures, not only the most crucial structural elements, have to be properly aligned to produce a predictive model. The automated compound alignment must produce a good alignment for all structural elements in the inhibitor set in addition to the superposition of the catechol substructure important for the inhibitory activity. The inhibitor alignment was made automatically with docking and evaluated with experimental structural data.

CoMFA. PLS analysis of the created CoMFA steric and electrostatic fields using the cross-validation of five random groups gave an average q^2 of 0.594 with five components, and an average Spress of 0.538. Five components were thus chosen for the final model, which produced a non-cross-validated PLS model with an r^2 of 0.903 and a standard error of estimate of 0.263. These values (also shown in Table 3) indicate that the CoMFA model is statistically significant. The predicted inhibitory activities of the inhibitors are listed in Table 4 and a graph of experimental versus predicted activities is presented in Figure 5. There were no significant outliers in the predictions, in-

Table 3. Statistical parameters of the CoMFA and GRID/GOLPE models.

		Cross-validated				Conventional	
	Groups[a]	q^2	Comp[b]	Spress[c]		r^2	s^d
CoMFA	5	0.594	5	0.538		0.903	0.263
CoMFA	2	0.483	5	0.603		0.903	0.263
GOLPE	5	0.636	3	0.494		0.830	
GOLPE	2	0.591	3	0.523		0.830	

[a]Number of groups used in the cross-validation.
[b]Optimal number of components.
[c]Standard error of predictions.
[d]Standard error of estimate.
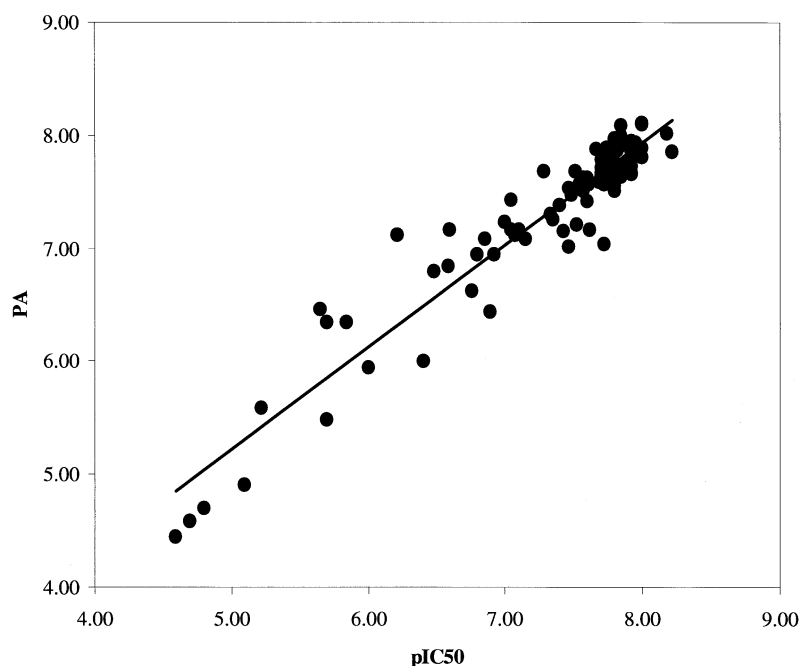
Figure 5. Correlation between the experimental activities (pIC_{50}) and predicted activities (PA) using the CoMFA model.

dicating the robustness and predictive power of the model.

Random groups cross-validation with two groups produced an average q^2 of 0.483 with five components, and an average Spress of 0.603. The average q^2 value below 0.5 indicates that the CoMFA model may have difficulties in predicting activities for inhibitors not used in the model building. The large gap between the r^2 (0.903) and the q^2 (0.483) values indicates that the model may also contain irrelevant X-variables.

GRID/GOLPE. The GRID/GOLPE procedure was used to examine the correlation between biological activity and the interaction between inhibitors, Mg^{2+} and DRY probes. The PLS analysis of 42,411 active X-variables, obtained from the advanced data pretreat-

ment, gave a conventional r^2 of 0.865, followed by a q^2 of 0.481 with three components and Spress of 0.589. These values were good enough to allow for SRD treatment. With a similar pretreatment procedure, Mg^{2+} and DRY probes alone yielded q^2 values of 0.465 and 0.313, respectively. These values give an estimation of the amount of each probe in explaining the data.

Further analysis of the interaction field data with SRD and F. Factorial Selection gave 3,676 active X-variables, which were used in building the final 3D QSAR model. The X-variable matrix and activity Y-vector were analysed using GOLPE PLS, which produced a statistically significant conventional r^2 of 0.830 with three components. From the cross-

Table 4. Activity predictions of the CoMFA and GRID/GOLPE models.

Compd	pIC$_{50}$	Pred.CoMFA[a]	Pred.GOLPE[b]	Compd	pIC$_{50}$	Pred.CoMFA[a]	Pred.GOLPE[b]
1	7.07	7.16	6.91	47	7.80	7.69	7.99
2	8.22	7.86	7.98	48	7.54	7.57	7.85
3	7.57	7.51	7.17	49	7.72	7.57	7.99
4	7.48	7.48	7.26	50	7.92	7.91	7.68
5	7.92	7.95	7.81	51	7.85	7.68	7.56
6	7.70	7.79	7.88	52	7.85	7.64	8.02
7	8.18	8.02	8.01	53	7.85	8.09	8.10
8	7.72	7.85	7.53	54	6.60	7.17	7.23
9	7.40	7.39	7.15	55	7.80	7.77	7.63
10	7.00	7.24	7.12	56	7.46	7.54	7.50
11	7.80	7.98	8.18	57	7.57	7.62	7.73
12	7.66	7.89	7.36	58	7.92	7.86	8.11
13	7.62	7.17	7.00	59	7.80	7.62	7.50
14	7.68	7.59	7.79	60	7.51	7.69	7.63
15	7.92	7.80	7.63	61	7.77	7.66	8.09
16	7.85	8.00	8.00	62	7.85	7.65	7.80
17	7.52	7.22	7.49	63	7.95	7.94	7.96
18	6.00	5.94	6.54	64	5.10	4.91	5.62
19	7.82	7.71	7.65	65	6.40	6.01	6.55
20	7.92	7.67	7.06	66	5.70	6.35	6.36
21	7.70	7.72	7.33	67	7.72	7.05	7.15
22	7.74	7.77	7.36	68	6.59	6.85	6.34
23	7.89	7.78	7.95	69	6.80	6.95	6.57
24	7.92	7.74	7.85	70	6.89	6.44	6.38
25	7.80	7.70	7.88	71	5.65	6.47	5.60
26	7.92	7.67	7.84	72	5.84	6.35	5.84
27	7.60	7.63	7.86	73	7.35	7.26	7.34
28	7.80	7.76	8.01	74	4.59	4.45	4.36
29	7.85	7.70	7.75	75	7.46	7.01	7.75
30	7.85	7.92	8.03	76	4.80	4.71	5.32
31	7.80	7.57	7.73	77	6.48	6.80	6.12
32	7.28	7.69	7.13	78	7.61	7.57	7.74
33	7.82	7.87	7.69	79	5.22	5.59	6.17
34	6.76	6.63	6.48	80	5.70	5.49	5.83
35	7.07	7.13	6.96	81	7.43	7.16	6.87
36	6.92	6.95	7.09	82	7.70	7.67	7.26
37	7.15	7.08	7.38	83	7.80	7.52	7.13
38	6.85	7.09	6.72	84	7.07	7.14	7.06
39	7.60	7.42	6.83	85	8.00	7.90	8.00
40	7.74	7.82	7.89	86	7.74	7.89	7.82
41	7.85	7.98	7.99	87	8.00	7.82	7.83
42	7.72	7.70	8.03	88	7.04	7.43	7.41
43	8.00	8.11	8.15	89	7.33	7.30	7.57
44	8.00	8.10	7.69	90	7.04	7.17	7.14
45	7.10	7.17	7.05	91	4.70	4.59	5.52
46	6.21	7.13	7.00	92	7.70	7.65	8.06

[a] Activities predicted by CoMFA model.
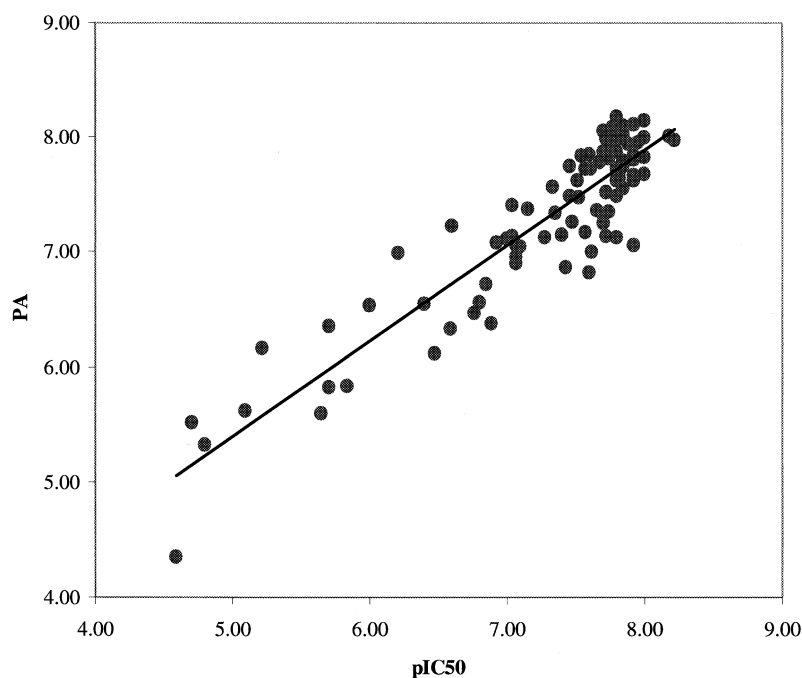[b] Activities predicted by GRID/GOLPE model.

Figure 6. Correlation between the experimental activities (pIC$_{50}$) and predicted activities (PA) using the GRID/GOLPE model.

validation with five random groups, three components also produced the lowest Spress of 0.494 and a good q^2 of 0.636, which was subsequently chosen for the final model. The statistics of the model are presented in Table 3. Predicted activities of the inhibitors are presented in Table 4, and a graph of the experimental versus predicted activities is presented in Figure 6. There were no significant outliers in the activity predictions, again indicating the robustness and predictive power of the model.

The cross-validation procedure with two random groups produced a q^2 of 0.591 with three components, and the lowest Spress of 0.523. The statistically significant q^2 value indicates that the GRID/GOLPE model is capable of predicting activities for inhibitors not used in the model building.

The statistical quality of both GRID/GOLPE and CoMFA 3D QSAR models is similar. The main difference between the models is in the difficulty of the CoMFA model to predict activities for inhibitors when the two random groups cross-validation method is used. However, this cross-validation method is very demanding, and therefore a q^2 value of 0.483 is not considered as a failure of the whole model. Overall, success was achieved in creating 3D QSAR models of COMT inhibitors with two different procedures, which indicates the general usability of the models

in both activity prediction and in the design of new inhibitors.

The interpretation of CoMFA and GOLPE maps

Statistically relevant information of both CoMFA and GRID/GOLPE procedures was also represented as 3D contour maps. This is an efficient way to examine the structural features and locations that are significant to the biological activity of an inhibitor. The reliability of the maps was evaluated by comparing the information with interactions that can be formed with the inhibitors and the amino acids in the enzyme's active site.

The contour maps obtained for CoMFA and GOLPE are displayed in Figure 7. Inhibitor **43** is displayed in the active site of COMT in these maps to assist in visualisation. The CoMFA contour maps are displayed as stdev*coeff plots in Figures 7A and 7B. Green regions show those areas where an increase of steric bulk is connected with enhanced activity and yellow regions show areas where steric bulk decreases activity. Red regions indicate areas where an increase in electronegativity enhances activity and blue regions those where electronegativity decreases activity.

The properties of substituents on the catechol ring are important for activity. The CoMFA maps show several contour features near the R$_5$ substituent. A blue region near the catechol ring and R$_5$ indicates that
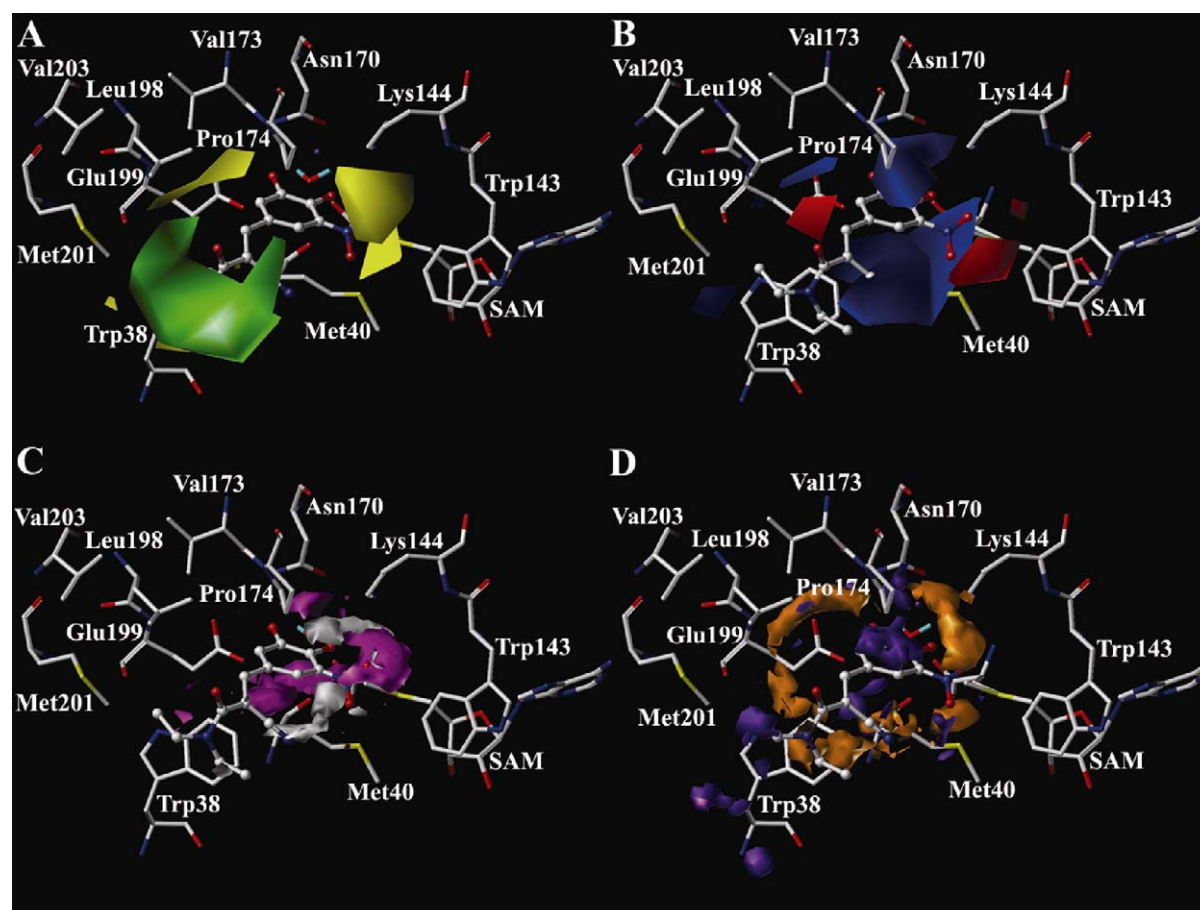
Figure 7. Contour maps of the 3D QSAR models with inhibitor **43** and amino acids in the COMT active site. All hydrogens except those of the water molecule are omitted for clarity. (A) The steric fields of CoMFA. Sterically favoured regions are in green; sterically disfavoured regions are in yellow. (B) The electrostatic fields of CoMFA. Red regions indicate that negative potential is favoured; blue regions indicate that positive potential is favoured. (C) The Mg^{2+} interaction fields of GOLPE. Regions that indicate favourable interaction with Mg^{2+} are in magenta (negative coefficients under -0.0005) and unfavourable interaction regions are in white (positive coefficients over 0.0007). (D) The DRY interaction fields of GOLPE. Regions that indicate favourable interaction with DRY are in orange (negative coefficients under -0.0005) and unfavourable interaction regions are in purple (positive coefficients over 0.0001).

a positive partial charge in that area enhances activity. Yellow regions and one green region around this area suggest that a large R_5 substituent is disfavoured due to steric limitations, and that the correct orientation for R_5 is important. The red region indicates that negatively charged R_5 enhances activity, and also shows the importance of hydrogen bonding between an inhibitor and Lys144 identified in the docking simulations. Thus, the CoMFA maps support the findings of Taskinen et al. [12] concerning the importance of an electron-withdrawing nature for R_5. We found that the size of the R_5 substituent was also important in docking simulations, and CoMFA results also suggest that the size of R_5 correlates with observed biological activity. Because there were no variations in the pos-

itions of the catechol oxygens in the inhibitor panel, there were no CoMFA regions in that area. The green regions and one yellow region near the R_1 substituent indicate areas where steric elements enhance and decrease activity, respectively. A red region near the oxygen atom in R_1 of **43** suggests that electronegativity in that area could enhance activity. A large blue region under the whole carbon skeleton of **43** suggests that this area favours positive interactions. This correlates with the suggestion by Taskinen et al. [12] that electron-withdrawing R_1 substituents are favoured.

GOLPE maps are displayed in Figures 7C and 7D. The magenta regions represent favourable interactions and the white regions represent unfavourable interactions with the Mg^{2+} probe. The orange regions

show favourable interactions and the purple regions show unfavourable interactions with the hydrophobic DRY probe. Two magenta regions close to the catechol oxygens of **43** suggest that the area is favourable for interaction with Mg^{2+}. This result supports the findings of Vidgren et al. [16], who show that the interaction between an inhibitor and the active site Mg^{2+} ion is important for activity. The magenta region near the R_5 substituent favours placement of the positive Mg^{2+} probe. That area partly overlaps with a red region of CoMFA (indicating that electronegativity enhances activity; Figure 7B), which may indicate that the area favours an interaction with an electronegative group. The interaction between an inhibitor and Mg^{2+} is disfavoured in areas perpendicular to the plane of the catechol ring, and some areas near the R_1 substituent of **43**. Hydrophobic interactions are favoured below and above the skeleton of **43**. Hydrophobic contacts between an inhibitor and the active site residues (Trp38, Pro174, Leu198, SAM) that were found to be important in the docking simulations support this result. Interestingly, a large orange region of GOLPE illustrates areas where the inhibitor favours hydrophobic contacts partly overlap with a blue region of CoMFA maps, which in turn suggests that positive charge in that area would increase activity; thereby, these regions support each other. Hydrophobic contacts are disfavoured at the end of the R_1 substituent of **43**, according to GOLPE maps. However, according to CoMFA maps, there are no considerable steric correlations in this regard. The inhibitor alignment in this area was poor, which may explain the quality of the 3D QSAR models. However, it is apparent that R_1 should not be too large or hydrophobic, as that part of the inhibitor may be pointing towards the solvent.

Both CoMFA and GOLPE maps show compatibility with the COMT active site. The large blue region under the carbon skeleton of **43** (Figure 7B), consistent with the orange DRY region in GOLPE (Figure 7D), is located in the volume close to the hydrophobic parts of Trp38, Met40 and SAM, which indicate favourable hydrophobic contacts with these residues. The yellow regions are sterically disfavoured, and located in a sterically restricted volume where Met40, Lys144 and SAM limit the volume available for inhibitors. The yellow region close to Pro174 indicates that a steric bulk in that area interferes with the enzyme. The green areas indicating sterically favoured regions do not overlap with the amino acids in the COMT active site. The two magenta regions indicating a favourable interaction with the Mg^{2+} probe are both close to the

Mg^{2+} ion in the active site. White areas where Mg^{2+} interactions are disfavoured are located near the hydrophobic portions of Trp143, Lys144 and Pro174. In addition to regions common with the blue CoMFA regions, hydrophobic contacts in the orange regions are located in front of Pro174 and Leu198. The sterically disfavoured purple DRY area attached to Trp38 and Pro174 indicates the importance of limitations within the COMT active site volume.

The 3D QSAR models and the docking results are in agreement, and the interpretation of the contour maps correlates well with the COMT active site structure. Because the 3D QSAR contour maps of the regions that explain the variance in the experimental activity are in consensus with the amino acids in the COMT active site, we can presume that the docked conformations of the inhibitors are reliable. Information obtained from the 3D QSAR models created in this study supports and improves understanding of factors affecting the biological activity of COMT inhibitors and thereby assists in the design of new inhibitors.

Conclusions

In this study we have explored the binding conformations and structure-activity relationships of a panel of 92 COMT inhibitors using automated docking with the FlexX program, and 3D QSAR with CoMFA and GRID/GOLPE. Results show that although FlexX had problems in estimating the binding free energies of inhibitors, it correctly predicted the interaction between the catechol oxygens of inhibitors and the Mg^{2+} ion in the COMT active site. The interaction model shows that the interaction between inhibitors and the Mg^{2+} ion is important for activity, and hydrogen bonds between the inhibitor and Lys144, Asn170 and Glu199 side chains influence the binding mode of the inhibitors.

Molecular docking has been previously combined with different 3D QSAR analyses [17–19]. Our study further demonstrates that molecular docking followed by 3D QSAR analysis can yield explanative models. On the basis of the conformations produced with docking, predictive CoMFA and GRID/GOLPE models were created with satisfactory q^2 values as a result, and statistically important interactions in both models can be traced back to the amino acid structure of the COMT active site.

Our 3D QSAR models support previous findings concerning the importance of the Mg^{2+} interaction [16], and the electron-withdrawing character of R_1 and R_5 substituents on the inhibitors [12]. Thereby, replacing the nitro R_5 substituent with a more electronegative group, which is also able to form hydrogen bonds, could increase the pIC_{50} of entacapone. Results also suggest that the position of the carbonyl oxygen in R_1 of entacapone is more favourable than that of tolcapone. In the docked conformation of tolcapone the carbonyl oxygen is towards hydrophobic parts of Val173 and Leu198, whereas in entacapone the carbonyl oxygen is situated closer to the enzyme surface and could therefore interact with the solvent.

3D QSAR indicates that steric volume is favoured in the tail of R_1. For example, increasing the size of the entacapone diethylamine group increases pIC_{50} according to our model. Docking indicates that this type of modification is favourable, if the new substituent can form hydrophobic contacts with side chains of Val173, Leu198, Met201 and Val203. However, the R_1 substituent is also oriented towards the solvent water, and thus a too hydrophobic R_1 is not appropriate.

Recognising the molecular properties that affect the enzyme-inhibitor binding and biological activity is essential in drug design. As shown in this paper, these properties can be analysed by combining receptor-based and inhibitor-based methods. The expanding amount of experimental information increases the accuracy of molecular modeling and improves the prospects of future drug design.

Acknowledgements

References

1. Männistö, P.T. and Kaakkola, S., Pharmacol. Rev., 51 (1999) 593.
2. Axelrod, J. and Tomchick, R., J. Biol. Chem., 233 (1958) 702.
3. Guldberg, H.C. and Marsden, C.A., Pharmacol. Rev., 27 (1975) 135.
4. Bonifati, V. and Meco, G., Pharmacol. Ther., 81 (1999) 1.
5. Gordin, A., Kaakkola, S. and Teräväinen, H., Adv. Neurol., 91 (2003) 237.
6. Guttman, M., Leger, G., Reches, A., Evans, A., Kuwabara, H., Cedarbaum, J.M. and Gjedde, A., Mov. Disord., 8 (1993) 298.
7. Najib, J., Clin. Ther., 23 (2001) 802.
8. De Santi, C., Giulianotti, P.C., Pietrabissa, A., Mosca, F. and Pacifici, G.M., Eur. J. Clin. Pharmacol., 54 (1998) 215.
9. Zurcher, G., Colzi, A. and Da Prada, M.J., Neural. Transm. Suppl., 32 (1990) 375.
10. Micek, S.T. and Ernst, M.E., Am. J. Health Syst. Pharm., 56 (1999) 2195.
11. Olanow, C.W., Arch. Neurol., 57 (2000) 263.
12. Taskinen, J., Vidgren, J., Ovaska, M., Bäckström, R., Pippuri, A. and Nissinen, E., Quant. Struct.-Act. Relat., 8 (1989) 210.
13. Lotta, T., Taskinen, J., Bäckström, R. and Nissinen, E., J. Comput.-Aided Mol. Des., 6 (1992) 253.
14. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
15. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, I.N., Shindyalov, P.E. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
16. Vidgren, J., Svensson, L.A. and Liljas, A., Nature, 368 (1994) 354.
17. Buolamwini, J.K. and Assefa, H., J. Med. Chem., 45 (2002) 841.
18. Sippl, W., J. Comput.-Aided Mol. Des., 14 (2000) 559.
19. Liu, H., Huang, X., Shen, J., Luo, X., Li, M., Xiong, B., Chen, G., Yang, Y., Jiang, H. and Chen, K., J. Med. Chem., 45 (2002) 4816.
20. Böhm, H.J., J. Comput.-Aided Mol. Des., 8 (1994) 243.
21. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
22. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R. and Clementi, S., Quant. Struct.-Act. Relat., 12 (1993) 9.
23. Sybyl 6.8, Tripos Inc., St. Louis, MO.
24. Sadowski, J. and Gasteiger, J., Chem. Rev., 93 (1993) 2567.
25. Sadowski, J., Gasteiger, J. and Klebe, G., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
26. Clark, M., Cramer III, R.D. and Van Opdenbosch, N., J. Comput. Chem., 10 (1989) 982.
27. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
28. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
29. Purcell, W.P. and Singer, J.A., J. Chem. Eng. Data 12 (1967) 235.
30. Stewart, J.J., J. Comput.-Aided Mol. Des., 4 (1990) 1.
31. Besler, B.H., Merz, K.M. and Kollman, P.A., J. Comput. Chem., 11 (1990) 431.
32. Cramer III, R.D., Bunce, J.D. and Patterson, D.E., Quant. Struct.-Act. Relat., 7 (1988) 18.
33. Oprea, T.I. and Garcia, A.E., J. Comput.-Aided Mol. Des., 10 (1996) 186.
34. Pastor, M., Cruciani, G. and Clementi, S., J. Med. Chem., 40 (1997) 1455.
35. Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F. and Matthew, J.B., J. Mol. Graph. Model., 20 (2002) 281.
36. Wallace, A.C., Laskowski, R.A. and Thornton, J.M., Protein Eng., 8 (1995) 127.
37. Lehtonen, J. V., Rantanen, V.-V., Still, D.-J., Ekholm, J., Björklund, D., Iftikhar, Z., Huhtala, M., Jussila, A., Jaakkola, J., Pentikäinen, O.T., Nyrönen, T.A., S.T., Gyllenberg, M. and Johnson, M.S., BODIL: a molecular modeling environment for structure-function analysis and drug discovery, unpublished, http://www.abo.fi/fak/mnf/bkf/research/johnson/bodil.html.