European Journal of Medicinal Chemistry 44 (2009) 3524–3532

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Docking and 3D-QSAR studies of BMS-806 analogs as HIV-1 gp120 entry inhibitors

C. Teixeira, N. Serradji, F. Maurel, F. Barbault*

Laboratoire ITODYS (Interface, Traitement, Organisation et DYnamiques des Systèmes), Université Paris Diderot, 15 rue Jean de Baïf, bâtiment Lavoisier, 75013 Paris, France

ARTICLE INFO

Article history:
Received 10 September 2008
Received in revised form
27 January 2009
Accepted 19 March 2009
Available online 28 March 2009

Keywords:
gp120 inhibitors
3D-QSAR
CoMFA
CoMSIA
Docking

ABSTRACT

BMS-378806 (BMS-806) is a small molecule that blocks the binding of host-cell CD4 with viral gp120 protein and therefore inhibits the first steps of HIV-1 infection. Recently, 36 analogs compounds of BMS-806 were synthesized and their biological activity evaluated. Based on these compounds, a molecular docking was firstly performed with BMS-806 to the gp120 cavity in order to get a representative ligand conformation for the 3D-QSAR process. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were then conducted for these 36 compounds. CoMFA and CoMSIA models give reliable correlative and predictive abilities but the CoMFA model performance was slightly better than CoMSIA. CoMFA contours were analysed and have been correlated to the gp120 viral protein. The discussion indicates several key fragment positions on the ligands and their implications on the gp120 protein binding. The computational approach used in this paper provides reliable clues for further design of small molecules gp120/CD4 inhibitors based on the BMS-806.

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is due to the infection by the HIV-1 virus and remains a major concern for worldwide health. The temporary control over viral progression is made thanks to a combination of several reverse transcriptase and protease inhibitors which is called highly active anti-retroviral therapy (HAART). Currently, HAART toxicities restrict their uses in anti-HIV therapies. Furthermore, the continuing emergence of new viral resistance strains limits the biological activity of such drugs. Therefore, original and efficient anti-retroviral drugs elaboration is still desired.

The gp120 and gp41 HIV-1 proteins play essential roles in orchestrating the viral entry process and therefore, represent promising antiviral targets. These viral proteins' behaviours are concerted in order to lead to the cellular infection [1]. The first event is the binding process between CD4 and gp120 is the interaction of the N-terminal moiety of the receptor with the viral cavity of gp120 protein. This binding creates the necessary exposure of an interacting surface for the CCR5 or CXCR4 host-cell chemokine receptor. This chain of events triggers the release of gp41 which, ultimately, undergoes a large conformational change responsible for the viral and cellular membranes' apposition, thereby allowing entry of the viral genetic material into the cytosol [2,3]. All these steps provide potential HIV-1 targets for new drugs and are attractive because they are at the first steps of HIV-1 life cycle, preceding the cellular infection [4,5].

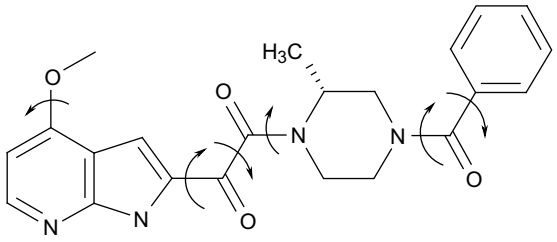
Three crystal structures of gp120, bounded to a protein which mimics CD4, were previously obtained by X-ray diffraction and are available in the protein databank [1,6,7]. All possess a highly conserved structural motif between gp120 and CD4. The recognition between the two proteins is essential and the amino-acids belonging to this area of interaction are highly conserved among HIV-1 isolates. Consequently, the CD4-gp120 cavity has been suggested to be a potential target for drug-design [1,7,8].

To date, most of the gp120 inhibitors are proteins or peptides [9]. Therefore, a peptidomimetic strategy was already tempted, with several aminoglycoside antibiotic compounds modified with additions of arginine residues [2,10–14]. Concerning small organic molecules, BMS-806 (Scheme 1) and BMS-043 (formally BMS-378806 and BMS-488043 – see Table 1) are the most promising agents and demonstrated the ability to reduce viral load in man [15–17]. Their mode of action was controversial [18–21] but a study finally stated that the binding pocket on gp120 for these molecules was the same as the CD4-gp120 interacting cavity [22].

Recently, a series of BMS-806 analogs were synthesized and biologically evaluated [23]. The main originality of these molecules dwells on the replacement of a ketoamide group by a sulfonamide function which roughly adopts the same arrangement and then constitutes an isosteric group. This fact was demonstrated previously by a theoretical approach that led to a molecular

* Corresponding author. Tel.: +33 1 57 27 88 50; fax: +33 1 57 27 72 63.
E-mail address: florent.barbault@univ-paris-diderot.fr (F. Barbault).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.03.028



Scheme 1. Molecular scheme of the BMS-806 molecule and its dihedral angles used for the docking calculation.

superposition which has clearly shown the isosteric behaviour of sulfonamide function to the ketoamide group [23]. Among the 36 compounds synthesized, several compounds present good and promising affinities to gp120. The purpose of this work was to rationalize the molecular variations to the inhibitory activities. To do that, comparative molecular similarity analysis (CoMSIA) along with comparative molecular fields analysis (CoMFA) were used to build 3D-QSAR models. The predictive abilities of such models were checked and validated statistically with a representative test set of compounds. Furthermore, the 3D-QSAR CoMFA contour maps were superimposed to the BMS-806/gp120 docked structure. The good correspondences between the contour maps and the properties of the gp120 amino acids interacting with the ligand led us to identify several key points of the binding mode for the BMS-806 analogs. To our knowledge, this work provides the first 3D-QSAR study for this family of compounds and provides a platform for the prediction of HIV-1 entry inhibitors derived from BMS-806.

2. Computational details

2.1. Molecular docking calculations

In a 3D-QSAR analysis, the ligand geometry must correspond to the one which is responsible for the biological effect. As no experimental gp120 complex structure suits with BMS-806, docking calculations were conducted to obtain a structure of the complex. Then the complex structure was optimized.

Autodock [24] software version 3.0 was used for this molecular docking process. After removing the CD4 protein of the complex structure [7] (pdb code 1G9N), a grid box of 70 Å in the three axis directions was constructed around the gp120 hydrophobic binding cavity. The Lamarckian Genetic Algorithm method [24] was used for the global optimum binding position search. One hundred cycles of calculations were performed in order to get a final binding position as accurate as possible. All ligand flexible dihedral angles (see Scheme 1) were set free to move during the simulation in order to authorize all binding conformations possibilities.

Because Autodock software works with only polar hydrogen, the missing hydrogen atoms were added, with SYBYL software, to the BMS-806 conformation which displayed the lowest docking energy.

An optimization of the whole system (protein and ligand) was then performed, in the Hakwins et al. generalized Born (GB) implicit solvent model [25,26], with AMBER 9.0 software [27]. The calculation started with 1000 steps of steepest descent followed by 1000 steps of conjugate gradient minimization with a gradient tolerance of 0.05 kcal mol^{-1}. Visualization and analysis of protein ligand interactions were made with the help of the visual molecular dynamic software [28] (VMD).

2.2. Dataset compounds

In this work, all activity data, reported as IC_{50} antiviral activity, were determined in the same conditions [23] following a previously published protocol [29,30]. This point is essential since homogeneity of experimental conditions is required to get good 3D-QSAR models. The activities of 36 compounds were reported and they constitute the global dataset. These compounds, listed in Table 1, were constructed with the SYBYL software using BMS-806 docked conformation as scaffold. Minimizations were performed with SYBYL program along with Tripos force field and Gasteiger–Marsili atomic partial charges. To obtain the molecular spatial alignment required for the 3D-QSAR models, the constructed molecules were superimposed to the docked conformation of BMS-806 according to the heavy atoms of the piperazine ring and the carbonyl of benzoyl moiety. These atoms were chosen according to the maximum common sub-structure method (MCSS). The yielded superposition, represented in Fig. 1, is homogeneous.

2.3. 3D-QSAR models

To calculate the required 3D descriptors, a grid box, with a 2 Å spacing, was created to contain all molecules. The dimensions were automatically determined in such a way that the boundaries were extended beyond 4 Å in the *x*, *y* and *z* directions for all molecules.

For the CoMFA field generation, the van der Waals (Lennard Jones) and Coulombic potentials, which respectively represent the steric and electrostatic interactions, were calculated at each lattice intersection of the defined box. These calculations were performed, for all molecules, with a carbon probe atom positively charged which was moved to every lattice point. To speed up the analysis and reduce noise, a minimum column filtering value of 2.0 kcal/mol was used for the cross-validation.

CoMSIA fields were derived with the same lattice box carried out for CoMFA. CoMSIA incorporates five descriptors: steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor. These descriptors, called similarity indices, were calculated according to the method defined by Klebe et al. [31].

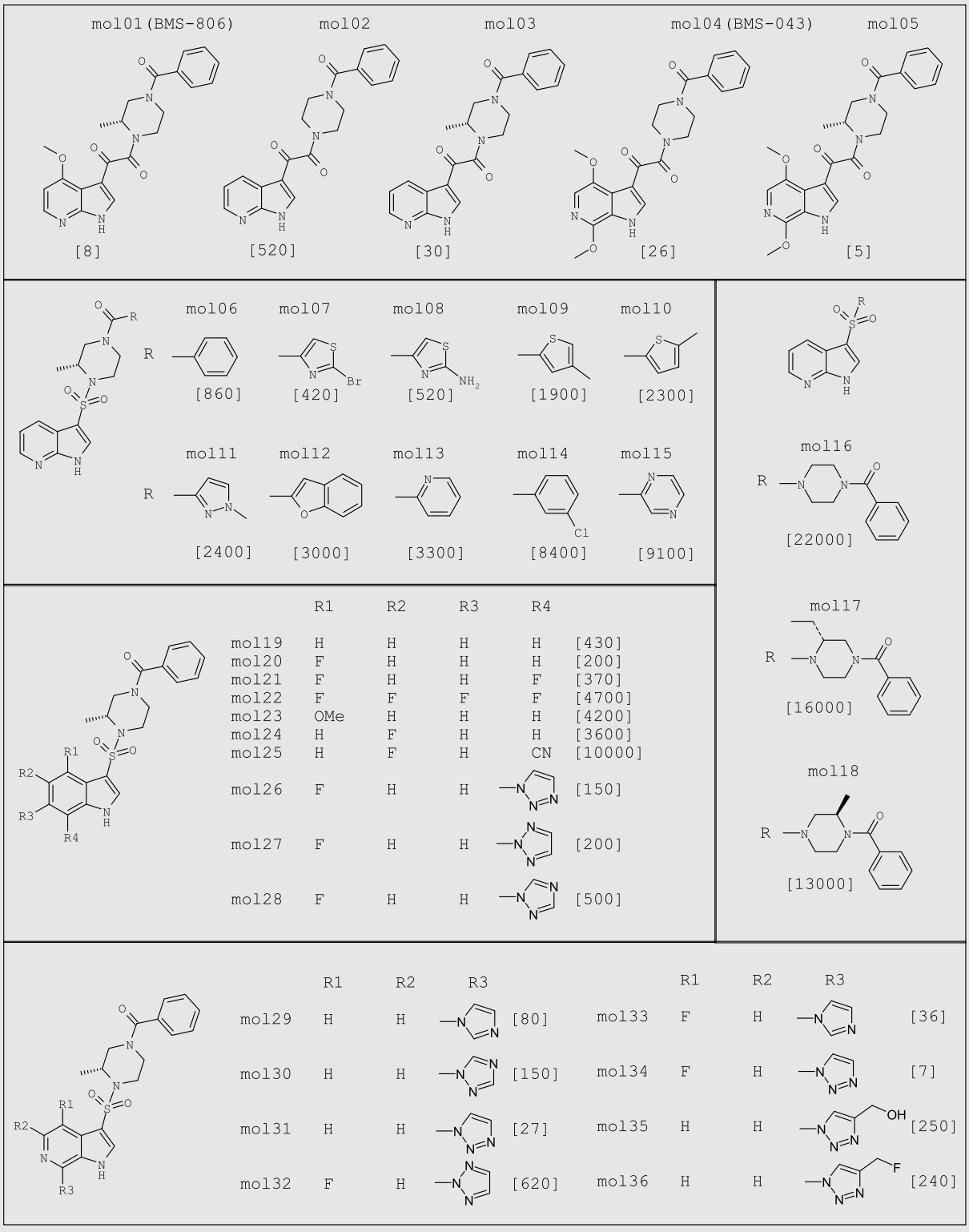
Partial Least Square (PLS) methodology [32,33] was performed to quantify the relationships between the CoMFA and CoMSIA descriptors and the biological activities. To evaluate the reliability of the models, cross-validation analysis were accomplished with the "leave one out" methodology (LOO) wherein one compound was moved away from the dataset and its activity was predicted by using the model derived from the rest of the dataset. A cross-validated coefficient, q^2, was then obtained and provided a glimpse of model predictive power. To minimize the data overrating tendency, the number of components was systematically optimized to produce the highest q^2 value. Finally, a non-cross-validation analysis was computed and the Pearson coefficient, r^2, was calculated. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis [34] were performed for 100 runs and the corresponding r^2 and standard error of estimates (SEE) were reported. The 3D-QSAR contour maps generated from the models were displayed and analysed in the protein environment with the SYBYL software.

The predictive power of the models was evaluated by predicting the activities of the six compounds belonging to the test set. The selection of these six compounds (marked with an asterisk in Table 3) was made by considering the fact that the test set must represent structural diversity and a range of antiviral activities similar to that of the training set. A predictive *r* square (r^2_{pred}) value was then obtained with the following formula:

$$r^2_{pred} = (SD - PRESS)/SD$$

In this equation, SD represents the sum of squared deviation between the biological activities of the test set molecule and the

Table 1
Structures of the compounds used in this article. Values in bracket are the experimental IC_{50}, in nM.



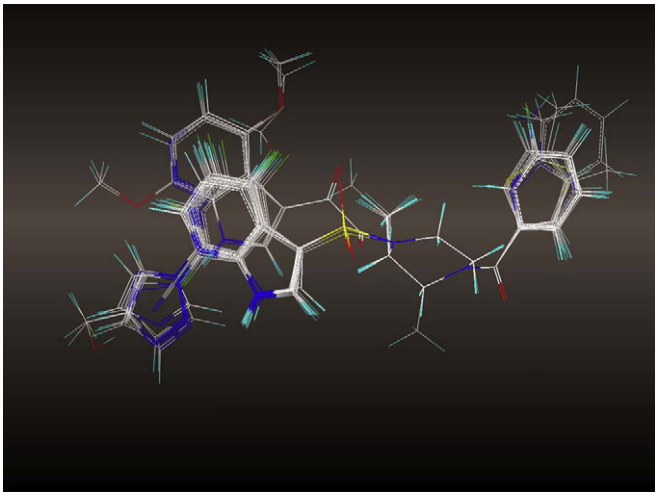


Fig. 1. Alignment of the 36 ligands' 3D structures employed for the molecular field generation and analysis.

mean activity of the training set molecules, PRESS is the sum of squared deviations between the experimental and predicted activities of the test molecules.

3. Results and discussion

3.1. Molecular docking studies

Docking calculations were used to find the optimal position of BMS-806 in the binding pocket of gp120 protein. The hundred resulting docking structures were clustered into conformations families according to a RMSD lower than 1 Å. It should be noted that the selection of a representative docking structure is often complicated as the results are particularly sensitive to the scoring function. In this work, the conformation selected was the one which presented the lowest docking free energy of binding (−10.5 kcal/mol) in the most populated cluster (6). Then, to provide an accurate view of the complex structure and authorize small adaptations of the protein geometry to the ligand, the gp120/BMS-806 complex was energetically minimized with AMBER software. The coordinates of the bound orientation of the gp120/BMS-806 complex are available in Supplementary material.

In the resulting geometry of BMS-806, the piperazine group is orthogonally placed with the azaindole part, at the entrance of the binding pocket formed by the protein residues Ile193, Gly287, Trp241 and Asn239. This last residue makes hydrophobic interactions with the piperazine methyl group and also, with the help of Val244 with the phenyl group of BMS-806, at the outside of the gp120 cavity. Moreover, a hydrogen bond (distance 2 Å and angle 149°) is identified between the NH atoms of the azaindole moiety and the side chain oxygen omega of Thr108. This hydrogen bond might be responsible for the azaindole group location in the hydrophobic pocket.

Although the validation of these calculations cannot be checked by comparison with experimental structure (not yet available) it could be noted that our results are in agreement with previous theoretical work [22]. Comparison with the complex structure of CD4 with gp120 points out that the azaindole group is more deeply inserted in the gp120 hydrophobic cavity than phenyl 43 of CD4. This observation was correlated with experimental mutagenesis data [35]. Indeed, large side chain modifications, like S375W or T257R, were made in the gp120 cavity. These mutations abate the BMS-806 affinity but stabilize the CD4-gp120 complex by residue filling. This result can only be explained if the ligand is deeply inserted in gp120 binding pocket, therefore when the azaindole is in its cavity [22].

Docking calculation was also performed for the most active compound: mol05 ($IC_{50} = 5$ nM). Its binding geometry inside gp120 is displayed in Fig. 2 (left) and is similar to the orientation obtained with BMS-806. Mol05 contains a ketoamide group. Therefore, in order to see the isosteric behaviour of sulfonamid on gp120 cavity, mol06 was also docked. This compound has been chosen because

Table 2
Summary of CoMFA and CoMSIA statistical results.

QSAR parameter	CoMFA model	CoMSIA model
q^2	0.534	0.583
Number of components	4	4
Standard error of prediction (SEP)	0.763	0.864
r^2	0.921	0.884
Standard error of estimate (SEE)	0.151	0.363
r^2 prediction	0.651	0.532
F-test value	73.022	47.513
r^2 bootstrapping	0.957 ± 0.017	0.931 ± 0.024
SEE bootstrapping	0.212 ± 0.126	0.274 ± 0.151
Percentage of field contribution		
Steric	50.7%	22.2%
Electrostatic	49.3%	–
Hydrophobic	–	36.7%
Acceptor H-bond	–	41.1%

Table 3
Experimental activities (IC_{50}) and predicted values of CoMFA and CoMSIA. Residual values represent the predicted minus the observed values. Compounds with a star represent the test set.

Name	IC_{50} (nM)	pIC_{50}	CoMFA		CoMSIA	
			Predicted	Residual	Predicted	Residual
MOL01	8	8.10	8.34	0.24	7.61	−0.49
MOL02	520	6.28	6.07	−0.22	6.50	0.21
MOL03*	30	7.52	6.88	−0.65	7.11	−0.41
MOL04	26	7.59	7.54	−0.04	7.82	0.23
MOL05	5	8.30	8.26	−0.04	8.28	−0.02
MOL06	860	6.07	5.45	−0.62	5.44	−0.63
MOL07*	420	6.38	5.78	−0.60	5.34	−1.04
MOL08	520	6.28	6.03	−0.26	6.21	−0.08
MOL09	1900	5.72	5.69	−0.03	5.59	−0.14
MOL10	2300	5.64	5.72	0.08	5.54	−0.10
MOL11*	2400	5.62	6.00	0.38	5.88	0.26
MOL12	3000	5.52	5.47	−0.05	5.46	−0.07
MOL13	3300	5.48	5.57	0.08	5.56	0.08
MOL14	8400	5.08	5.41	0.34	5.04	−0.04
MOL15	9100	5.04	5.37	0.33	5.30	0.26
MOL16	22,000	4.66	4.80	0.14	5.01	0.36
MOL17	16,000	4.80	5.15	0.35	5.17	0.38
MOL18	13,000	4.89	4.49	−0.39	4.48	−0.41
MOL19	430	6.37	6.15	−0.22	6.21	−0.15
MOL20	200	6.70	6.57	−0.13	6.23	−0.47
MOL21	370	6.43	6.19	−0.24	6.32	−0.11
MOL22*	4700	5.33	5.71	0.38	6.31	0.98
MOL23	4200	5.38	5.60	0.23	6.10	0.72
MOL24	3600	5.44	5.96	0.52	5.98	0.54
MOL25	10,000	5.00	4.99	−0.01	5.27	0.27
MOL26	150	6.82	7.29	0.47	7.11	0.29
MOL27	200	6.70	6.56	−0.14	6.17	−0.53
MOL28	500	6.30	6.54	0.24	6.38	0.08
MOL29	80	7.10	6.89	−0.21	7.30	0.20
MOL30*	150	6.82	6.21	−0.61	6.47	−0.36
MOL31	27	7.57	7.05	−0.51	7.32	−0.25
MOL32	620	6.21	6.44	0.23	6.45	0.25
MOL33	36	7.44	7.24	−0.21	7.35	−0.09
MOL34*	7	8.15	7.20	−0.95	7.29	−0.86
MOL35	250	6.60	6.59	−0.01	6.07	−0.53
MOL36	240	6.62	6.70	0.08	6.87	0.25

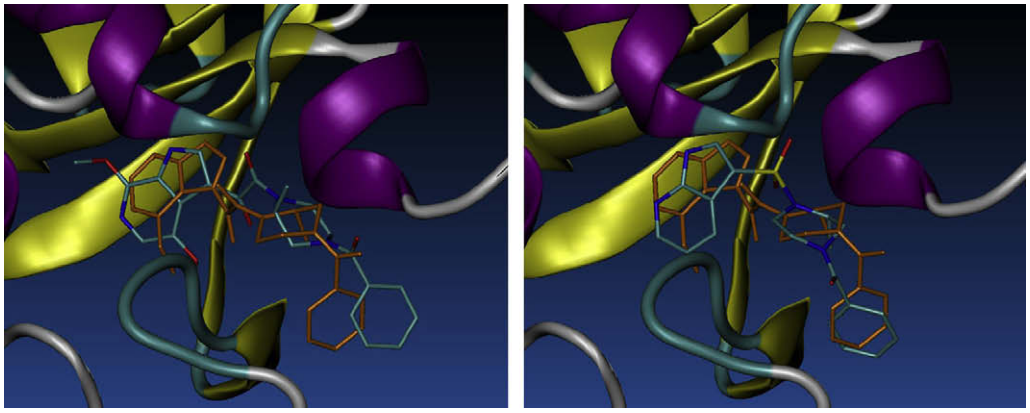


Fig. 2. Docking orientations of mol05 (left) and mol06 (right) on gp120 viral protein. For comparison, BMS-806 is displayed with orange color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Mol06 contains a sulfonamid group instead of a ketoamide and it is the most similar to BMS-806. The result docking orientation of mol06, shown in Fig. 2 (right), indicates that the binding position is identical to BMS-806 (in orange in Fig. 2) and demonstrates that the sulfonamide group is really isosteric to the ketoamide group, as suggested by Lu and coworkers [23].

3.2. 3D-QSAR studies

3.2.1. CoMFA model

To determine the optimum number of PLS components and minimize the tendency to over fit the data, the number of components, corresponding to the lowest SEP (Standard Error of Prediction) and highest q^2, was chosen to derive the final PLS regression model. In the CoMFA model, an optimum number of components of 4 with a q^2 of 0.534 and an SEP of 0.763 were obtained. According to the fact that q^2 coefficient is usually used as a measure of 3D-QSAR quality, the value of 0.534 indicates a reasonable correlation.

A non-cross-validated regression was then made and yielded a value of conventional r^2 of 0.921 which indicates that the correlation is statistically correct. To support the statistical validity of these models, 100 runs of bootstrapping method were produced to determine the error on the r^2 (r^2_{boot}) and the standard error of estimate (SEE_{boot}). An average r^2_{boot} value of 0.957 ± 0.017 and an SEE_{boot} value of 0.212 ± 0.126 demonstrate that the CoMFA model is very stable, possesses a statistical significance and a good predictive ability.

3.2.2. CoMSIA model

CoMSIA method adds three other descriptors to the CoMFA method. These are: hydrogen bond donor, hydrogen bond acceptor and hydrophobicity. Furthermore, to calculate 3D descriptors, CoMSIA includes a gaussian type distance dependent factor which smoothes the different fields contributions and can provide a better interpretable 3D-QSAR model [31]. As a consequence, several studies reported that CoMSIA models are less dependent on the spacing and positions (translation and rotation) in the grid [36,37].

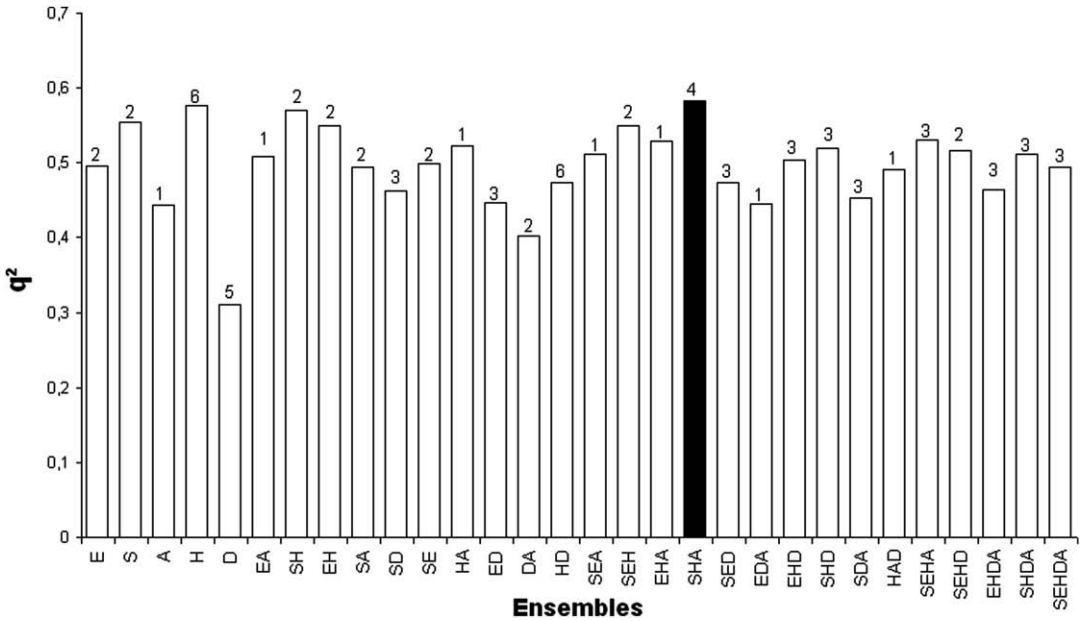


Fig. 3. Graph of the 31 possible CoMSIA descriptors' combinations (S = steric, E = electrostatic, H = hydrophobic, D/A = H-bond donor/acceptor) with their respective q^2 values. Above bars are reported their optimal number of components.

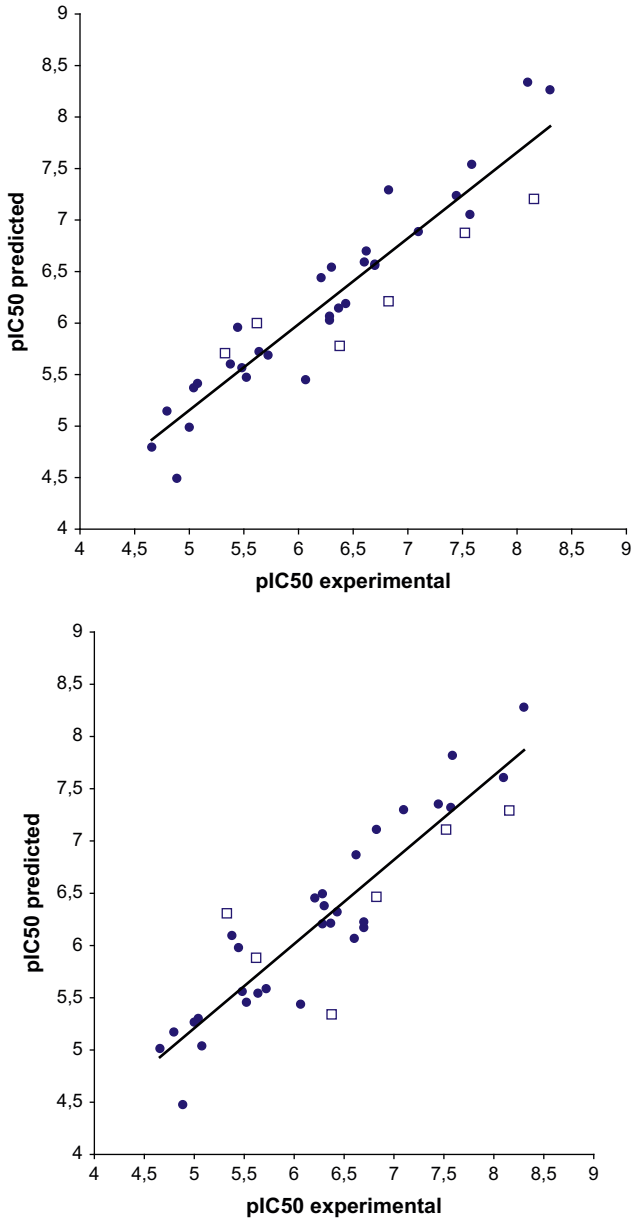


Fig. 4. Plots of predicted versus observed pIC_{50} values of CoMFA (top) and CoMSIA (bottom) models. Square points represent the compounds of the test set.

Former studies demonstrated that the five CoMSIA descriptors dependency may reduce the model significance and predictivity [36,38]. For this reason, all 31 possible descriptors' combinations were calculated with their respective q^2 value and optimum number of components (Fig. 3). The combination of steric, hydrophobicity and hydrogen bond acceptor presents the highest q^2 value and was then selected for further analysis. This best CoMSIA correlation was reported with a q^2 of 0.583 which is slightly better than the value of the previous CoMFA model (0.534). However, the standard error of prediction was higher with 0.864 (0.763 in CoMFA). Hence, with only these results, it was impossible to decide which 3D-QSAR model was the most meaningful.

The non-cross-validated regression gave a value of conventional r^2 of 0.884 which indicates that the CoMSIA model was statistically good, even if it was less reliable than CoMFA (0.921). The 100 bootstrapping runs produced an average r^2_{boot} value of 0.931 ± 0.024 and SEE_{boot} value of 0.274 ± 0.151. Once again, CoMSIA model appeared less attractive.

3.2.3. Model validation and selection

All 3D-QSAR statistical results are summarized in Table 2. The predictive r^2 was calculated to assess the quality of the 3D-QSAR models' predictive abilities. Values of 0.651 (see Table 2) and 0.532 were obtained for CoMFA and CoMSIA methods.

The CoMFA and CoMSIA correlations are plotted in Fig. 4 whereas the predicted and residual values for both models are compiled in Table 3. The test set molecules are also reported in the table and in the graphs. These points are placed above and below the correlation line; this indicates that both 3D-QSAR models' prediction ability is correct. However, like this has been demonstrated statistically, CoMSIA predictive power appears to be lower than CoMFA. Therefore, taking all statistical results into account, the CoMFA model has been selected for the structural analysis.

3.3. Analysis of CoMFA contours and structural implications

The contour maps generated from such CoMFA model can provide precious insights about the molecular binding to the targeted protein. With this objective, the CoMFA contours have been superimposed on the binding pocket of gp120 protein. Statistically, steric effect accounts for 50.7% of the antiviral activities whereas electrostatic effect represents 49.3% (Table 2). This ratio emphasizes the good balance between these two descriptors indicating that both are essential to explain the activity.

3.3.1. Steric contours

In the CoMFA steric fields, green areas indicate region where bulky groups increase the inhibitory activity (favourable) whereas yellow contours indicate where bulky groups decrease the activity (unfavourable). These contours are represented in Fig. 5. Four sterically unfavourable contours are localized on the azaindole ring with two above and two below. Azaindole ring is a planar fragment and deeply inserted in the gp120/CD4 cavity. Therefore, the unfavourable contours represent the limitation of the binding pocket (Fig. 6). The fact that unfavourable contours are placed above and below the azaindole ring suggests that a planar group, like an aromatic moiety, is required for increasing the inhibitory activity. As a consequence, we strongly propose to keep an aromatic fragment at this molecular position in order to fill the gp120 the binding pocket.

A supplementary unfavourable contour (yellow) is situated near the phenyl group of BMS-806. This one suggests that, even if this part is outside the cavity, ligand extension should be limited to small fragments. In fact, the phenyl part is close to a backbone coil of gp120 which restricts the size of this group. This area might explain some activity variations like for mol06 ($IC_{50} = 860$ nM) and mol14 ($IC_{50} = 8400$ nM) where the phenyl part is increased in size with a chlorine atom and then the IC_{50} is multiplied by a 10 factor.

Favourable contours, displayed in green in Fig. 5, are presented with the protein environment in Fig. 6. Only two favourable areas are reported, the first one is placed near the methyl group of the piperazine ring and the second one is situated above the NH atoms of the azaindole ring. The first favourable contour, located near the methyl group of the piperazine ring, explains why compounds mol03 and mol06 are better than, respectively, molecules mol02 and mol16. Indeed, the methyl group in the piperazine group seems to enhance the inhibitory activity.

The behaviour of this methyl fragment can be explained with the protein structure. Indeed, the azaindole and ketoamide group are deeply inserted in the gp120 cavity whereas the piperazine group is at the entrance of the interaction zone. The methyl fragment is making a hydrophobic interaction with the CH_2 group of Asn239 amino acid side chain (Fig. 6) which is also placed at the entrance of the gp120 cavity. This interaction suggests that the

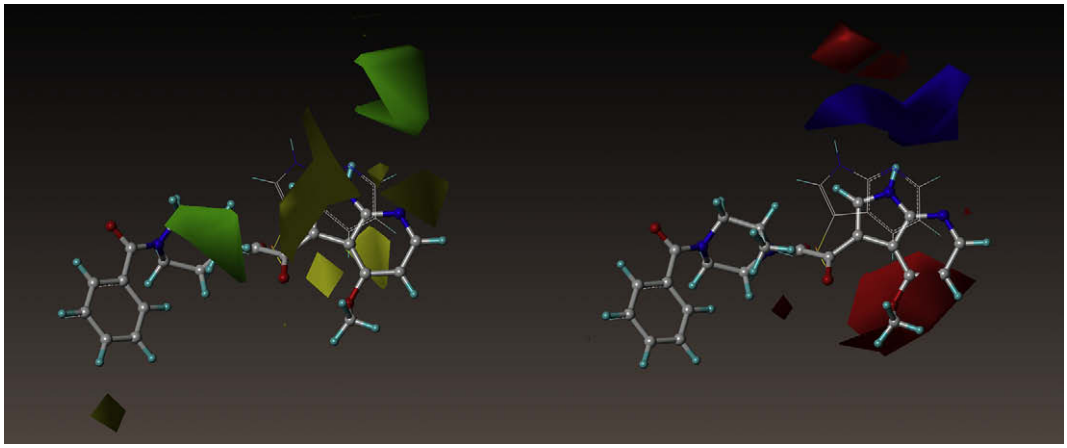


Fig. 5. CoMFA steric (left) and electrostatic (right) contours. Green areas indicate regions where bulky groups increase activity whereas yellow contours indicate where bulky groups decrease activity. Blue areas represent regions where positive groups enhance activity and red contours where negative groups increase the activity. BMS-806 compound is shown with ball and stick representation. For the comparison, mol06 which possess a sulfonamide instead of a ketoamide group is displayed in wireframe. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

methyl group is essential for a good binding with gp120 protein. This conclusion has been previously proposed and experimentally demonstrated [23].

More surprising is the favourable contour positioned above the NH group of the azaindole ring. In fact, the gp120 binding cavity geometry hints that there is not enough room to extend the azaindole ring. However, compounds mol34, mol31 or even mol33 display a good affinity while they possess a supplementary five member ring. A first analysis of the gp120 cavity prohibits every extension in the azaindole ring indicating the binding pocket is already filled. Nevertheless, as it has been demonstrated by molecular dynamics [22], gp120 is a flexible molecule and movements in the binding cavity cannot be excluded. The CoMFA sterically favourable contour is directly pointed to the peptide coil IVTHS from 193 to 197. This amino acids sequence does not present significant intramolecular interaction like salt bridges, hydrogen bonding or even hydrophobic contacts. The evidence suggests that this peptide sequence could be easily stretched to let enough room at a small fragment, like a five member ring, which could then reach other amino acids, more deeply inserted into gp120 core, to generate new interactions.

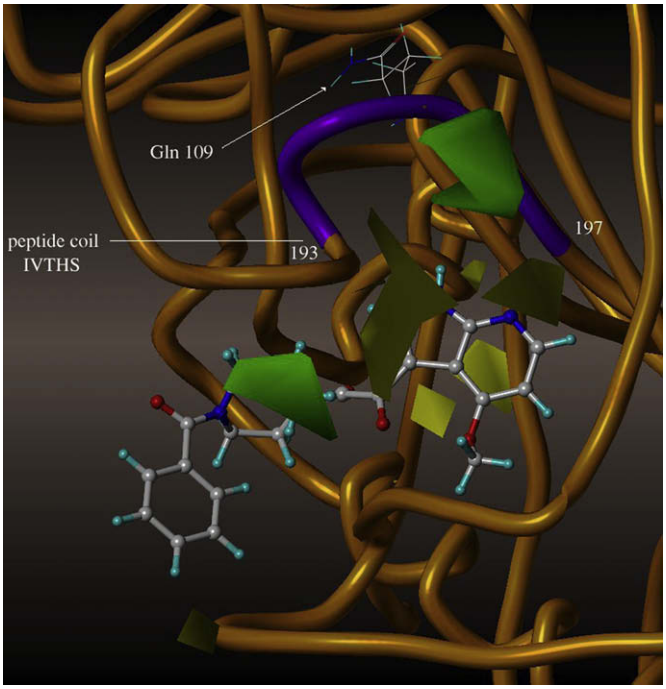


Fig. 6. CoMFA steric contour maps superimposed on the gp120 structure with BMS. Green contours indicate regions where bulky groups increase activity whereas yellow areas indicate regions where bulky groups decrease activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

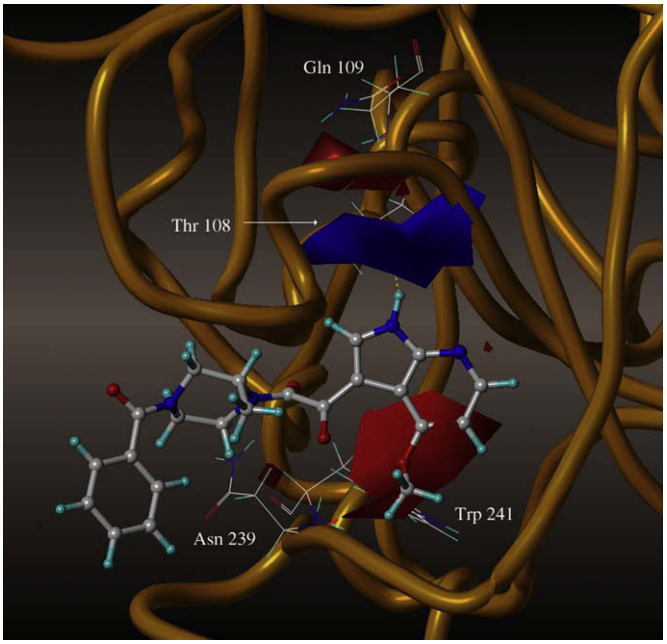


Fig. 7. CoMFA electrostatic contours. Blue areas represents region where positive group increases activity whereas red contours indicates region where negative charge increases the biological affinity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3.2. Electrostatic contours

In the CoMFA electrostatic fields, the blue and red contours represent the electropositive favourable and unfavourable regions, respectively. Therefore, an electropositive molecular fragment near a blue area would enhance the inhibitory activity whereas in the proximity of a red contour an electronegative fragment would be favourable. The CoMFA electrostatic contours are shown in Fig. 5.

Only one electropositive map is observed. This one is placed above the NH group of the azaindole ring. So, from this analysis, it appears that a positive group is welcome at this position for a good binding to gp120. The electropositive contour is broadly placed around the NH vector. A glimpse to Fig. 5 shows that the NH atoms are differently located on the superposition; this is due to the global compounds' superposition which was made on the central piperazine ring and carbonyl of benzoyl group. Thus, since the NH atoms are present in all compounds, the CoMFA surface is as broad as the different NH positions observed on the compounds' superposition. Moreover, the nitrogen of the azaindole ring participates to this electropositive map to discriminate indolic and azaindolic compounds. This fact explains the difference of inhibitory activity between mol06 and mol19, respectively of 860 and 430 nM, where 7-azaindole is replaced by an indole. Indeed, this nitrogen induces a modification of the electronic environment for mol06 and generating an electronegative area which is not favourable for a good antiviral activity like this has been defined with the electropositive surface of CoMFA.

The localization of this electropositive area onto the gp120 protein (Fig. 7) indicates why an electropositive group is needed at this position. Indeed, a hydrogen bond between the hydrogen of the NH indolic group of the ligand and the oxygen omega of threonine 108 is reported. This hydrogen bond is essential to the binding onto gp120.

In Fig. 5, a large electronegative contour is observed near the $-OCH_3$ group of BMS-806. This area explains why this last group is beneficial for the biological activity. For example, a comparison between compounds mol01 and mol03 (or mol29 and mol33) shows that an electronegative fragment increases the biological activity. The superposition of this contour onto the gp120 binding site gives a clue of the behaviour of this supposed interaction (Fig. 7). A supplemental hydrogen bond could justify the presence of this electronegative area around this ligand position, but such direct interaction is not observed with gp120. However, a slightly movement of gp120 could approximate two NH groups near the ligand, at the -OCH3 substituent, and therefore promote a hydrogen bond between them. These groups are the NH atoms of the indole side chain of Trp241 and the NH atoms of the peptide chain of Asn239 (Fig. 7). Indeed, these two candidates are close to the $-OCH_3$ but their NH vectors need a small motion for being aligned and producing a hydrogen bond. We suggest that a rotation of the dihedral angle between Cα and Cβ of Trp241 might be involved to completely expose the NH indole atoms for this hydrogen bond.

The last electrostatic contour of the CoMFA model is an electronegative area located above the electropositive area discussed previously (Fig. 5). The presence of two consecutive surfaces with opposite requirements (electronegative and electropositive) might be surprising. Yet, these two contours are clearly separated and then denote that these areas involve distinct interactions. Since this electronegative contour is a little bit far away from the azaindole ring, this area implicates compounds which possess a bulky fragment at the vicinity of the indolic NH like the ortho position. Eleven compounds present a nitrogenated five-membered ring at this position and these ligands generally own a good affinity. This strongly suggests that a supplementary interaction to gp120 is observed for these molecules.

Fig. 7 localizes the proximity of this electrostatic contour to the corresponding amino acids of gp120. The closest amino acid of this electronegative map is glutamine 109 in which the amido moiety is hydrogen bond donor. Thus, the presence of an electronegative group in the ligand may permit the creation of a hydrogen bond. As a consequence, compounds with a five-membered nitrogenated ring should make this hydrogen bond with Gln109. However, this amino acid is buried in gp120 and is only accessible if the peptide coil IVTHS is stretched (Fig. 6). The flexibility of this peptide sequence was previously proposed, according to the position of a sterically favourable area at this position (Figs. 5 and 6). The supplemental CoMFA area is then an additional argument to our hydrogen bond hypothesis with Gln109.

Among the eleven compounds in which there is a five-membered nitrogenated ring, the affinities are various. Among them, three compounds, mol34, mol33 and mol32 with respective IC_{50} of 7, 36 and 620 nM, differ only by the heterocycle ring. Due to the fact that an electronegative fragment is required for affinity, electrostatic potentials were calculated at semi-empirical quantum level with PM3 method using the Ampac software [39]. The electrostatic potentials are then displayed on their molecular surfaces in Fig. 8. The comparison between these three compounds clearly suggests that a nitrogen atom is desired at the position marked with an arrow in this last figure. One might suppose that this

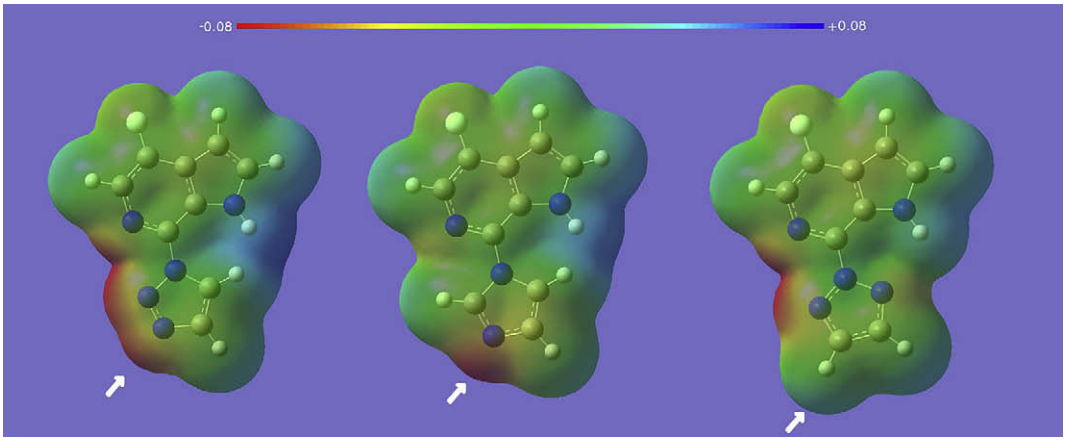


Fig. 8. Electrostatic potentials (ESP) plot on molecular surfaces of fragments making the distinction between, from left to right, mol34, mol33 and mol32. The scale is provided on the top of the figure and the unit is volt. Arrows indicate supposed position of the hydrogen donor atom of Gln109.

nitrogen is the best candidate as hydrogen bonding acceptor with Gln109 amido group of gp120.

4. Conclusions

The global purpose of this work was to produce a 3D-QSAR model for the prediction of BMS-806 analogs. To summarize, molecular docking study was firstly performed to obtain a satisfactory position of the ligand in the binding pocket of gp120. This one was then optimized with the protein to get an accurate view of the interactions with gp120. Secondly, the geometry of BMS-806 obtained from this calculation was then used to construct the 36 analog compounds. 3D-QSAR models were then computed with the CoMFA and CoMSIA method. The resulting models have been generated and give reliable correlative and predictive abilities. However, the CoMSIA model, even if all permutations between the five available 3D descriptors were considered, presented a lowest predictive performance than the CoMFA model. This last model was then chosen to elaborate structure–activity relationships.

The CoMFA 3D contour maps are in good agreement with the chemical variations of the dataset. Even if the protein structure is not required for a QSAR-3D study, the superimposition of gp120 with the 3D contour maps gives us key features on the ligand binding mode of interaction. For example, we observed that the azaindole ring position, located in the gp120 cavity, is crucial and must be kept as an aromatic. Like it was previously determined, the hydrogen bond of the indolic amino group with Thr108 appears to be essential to the binding. Two others modifications on this moiety might also give hydrogen bonds. It is the case for an electronegative group at the R1 position (see Table 1) of the ligand NH indolic group to target the amino group of Trp241 sidechain. The second hydrogen bond may be obtained by an extension with an electronegative fragment at the ortho position to target the amido group of Gln109. Furthermore, the methyl group on the piperazine ring seems to take the role of an anchor for the gp120 cavity with a hydrophobic interaction with the lateral chain of Asn239. All these results yield reliable and precious informations for further structure-based and ligand-based drug-design optimization.

Acknowledgements

We are grateful for the financial support from the French research association against AIDS (ANRS). Ph.D. fellowship of Catia Teixeira is supported by the "Fundação para a Ciência e a Tecnologia" of Portugal. We also warmly thank Professor Françoise Heymans for her suggestions and her careful reading.

Appendix. Supplementary material

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2009.03.028.

References

[1] P.D. Kwong, R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, W.A. Hendrickson, Nature 393 (1998) 648–659.
[2] G. Borkow, A. Lapidot, Curr. Drug Targets Infect. Disord. 5 (2005) 3–15.
[3] S.C. Harrison, Adv. Virus Res. 64 (2005) 231–261.
[4] W.S. Blair, P.F. Lin, N.A. Meanwell, O.B. Wallace, Drug Discov. Today 5 (2000) 183–194.
[5] J.P. Moore, R.W. Doms, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 10598–10602.
[6] C.C. Huang, M. Tang, M.Y. Zhang, S. Majeed, E. Montabana, R.L. Stanfield, D.S. Dimitrov, B. Korber, J. Sodroski, I.A. Wilson, R. Wyatt, P.D. Kwong, Science 310 (2005) 1025–1028.
[7] P.D. Kwong, R. Wyatt, S. Majeed, J. Robinson, R.W. Sweet, J. Sodroski, W.A. Hendrickson, Structure 8 (2000) 1329–1339.
[8] R. Wyatt, P.D. Kwong, E. Desjardins, R.W. Sweet, J. Robinson, W.A. Hendrickson, J.G. Sodroski, Nature 393 (1998) 705–711.
[9] K. Vermeire, D. Schols, Expert Opin. Investig. Drugs 14 (2005) 1199–1212.
[10] A. Berchanski, A. Lapidot, Biochim. Biophys. Acta 1768 (2007) 2107–2119.
[11] G. Borkow, V. Vijayabaskar, H.H. Lara, A. Kalinkovich, A. Lapidot, Antiviral Res. 60 (2003) 181–192.
[12] C. Cabrera, A. Gutierrez, J. Barretina, J. Blanco, A. Litovchick, A. Lapidot, B. Clotet, J.A. Este, Antiviral Res. 53 (2002) 1–8.
[13] M.V. Catani, M.T. Corasaniti, M. Ranalli, D. Amantea, A. Litovchick, A. Lapidot, G. Melino, J. Neurochem. 84 (2003) 1237–1245.
[14] R. Hegde, G. Borkow, A. Berchanski, A. Lapidot, FEBS J. 274 (2007) 6523–6536.
[15] Q. Guo, H.T. Ho, I. Dicker, L. Fan, N. Zhou, J. Friborg, T. Wang, B.V. McAuliffe, H.G. Wang, R.E. Rose, H. Fang, H.T. Scarnati, D.R. Langley, N.A. Meanwell, R. Abraham, R.J. Colonno, P.F. Lin, J. Virol. 77 (2003) 10528–10536.
[16] H.G. Wang, R.E. Williams, P.F. Lin, Curr. Pharm. Des. 10 (2004) 1785–1793.
[17] T. Wang, Z. Zhang, O.B. Wallace, M. Deshpande, H. Fang, Z. Yang, L.M. Zadjura, D.L. Tweedie, S. Huang, F. Zhao, S. Ranadive, B.S. Robinson, Y.F. Gong, K. Ricarrdi, T.P. Spicer, C. Deminie, R. Rose, H.G. Wang, W.S. Blair, P.Y. Shi, P.F. Lin, R.J. Colonno, N.A. Meanwell, J. Med. Chem. 46 (2003) 4236–4239.
[18] N. Madani, A.L. Perdigoto, K. Srinivasan, J.M. Cox, J.J. Chruma, J. LaLonde, M. Head, A.B. Smith 3rd, J.G. Sodroski, J. Virol. 78 (2004) 3742–3752.
[19] A. Schon, N. Madani, J.C. Klein, A. Hubicki, D. Ng, X. Yang, A.B. Smith 3rd, J. Sodroski, E. Freire, Biochemistry 45 (2006) 10973–10980.
[20] Z. Si, N. Madani, J.M. Cox, J.J. Chruma, J.C. Klein, A. Schon, N. Phan, L. Wang, A.C. Biorn, S. Cocklin, I. Chaiken, E. Freire, A.B. Smith 3rd, J.G. Sodroski, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 5036–5041.
[21] F. Stricher, L. Martin, P. Barthe, V. Pogenberg, A. Mechulam, A. Menez, C. Roumestand, F. Veas, C. Royer, C. Vita, Biochem. J. 390 (2005) 29–39.
[22] R. Kong, J.J. Tan, X.H. Ma, W.Z. Chen, C.X. Wang, Biochim. Biophys. Acta 1764 (2006) 766–772.
[23] R.J. Lu, J.A. Tucker, T. Zinevitch, O. Kirichenko, V. Konoplev, S. Kuznetsova, S. Sviridov, J. Pickens, S. Tandel, E. Brahmachary, Y. Yang, J. Wang, S. Freel, S. Fisher, A. Sullivan, J. Zhou, S. Stanfield-Oakley, M. Greenberg, D. Bolognesi, B. Bray, B. Koszalka, P. Jeffs, A. Khasanov, Y.A. Ma, C. Jeffries, C. Liu, T. Proskurina, T. Zhu, A. Chucholowski, R. Li, C. Sexton, J. Med. Chem. 50 (2007) 6535–6544.
[24] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.
[25] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. 100 (1996).
[26] V. Tsui, D.A. Case, Biopolymers 56 (2001) 275–291.
[27] D. Case, T. Darden, T. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Merz, D. Pearlman, M. Crowley, R. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. Mathews, C. Schafmeister, W. Ross, P. Kollman, AMBER, University of California, San Francisco, 2006.
[28] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. 14 (1996) 33–38.
[29] Q.X. Hu, A.P. Barry, Z.X. Wang, S.M. Connolly, S.C. Peiper, M.L. Greenberg, J. Virol. 74 (2000) 11858–11872.
[30] M. Mink, S.M. Mosier, S. Janumpalli, D. Davison, L. Jin, T. Melby, P. Sista, J. Erickson, D. Lambert, S.A. Stanfield-Oakley, M. Salgo, N. Cammack, T. Matthews, M.L. Greenberg, J. Virol. 79 (2005) 12447–12454.
[31] G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994) 4130–4146.
[32] L. Staahle, S. Wold, J. Chemom. 1 (1987) 185–196.
[33] S. Wold, C. Albano, W.J. Dunn, U. Edlund, K. Esbenson, P. Geladi, S. Hellberg, W. Lindburg, M. Sjostrom. Chemometrics: Mathematics and Statistics in Chemistry, The Netherlands, Dordrecht, 1984.
[34] R. Vong, P. Geladi, S. Wold, K. Esbensen, J. Chemom. 2 (1988) 281–296.
[35] S.H. Xiang, P.D. Kwong, R. Gupta, C.D. Rizzuto, D.J. Casper, R. Wyatt, L. Wang, W.A. Hendrickson, M.L. Doyle, J. Sodroski, J. Virol. 76 (2002) 9888–9899.
[36] M. Bohm, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994).
[37] S.J. Cho, A. Tropsha, J. Med. Chem. 38 (1995) 1060–1066.
[38] G. Bringmann, C. Rummey, J. Chem. Inf. Comput. Sci. 43 (2003) 304–316.
[39] AMPAC, Semichem, Inc., PO Box 1649, Shawnee, KS 66222, 1992–2004.