RESEARCH ARTICLE

MOLECULAR MODELLING AND VIRTUAL SCREENING APPLICATION TO THE COMPUTER-AIDED DESIGN OF ANTICANCER INHIBITORS WITH A FAVORABLE PHARMACOKINETIC PROFILE AGAINST E6 PAPILLOMAVIRUS TYPE 16

Ingrid Clémence Bléhoué[1] , Mawa Koné[2,3] , Elvice Akori Esmel[1] , Issouf Fofana[1] ,
Mélalie Kéïta[1] , Eugene Megnassan[*1,2,4]

[1]Laboratory of Fundamental and Applied Physics, University of Abobo-Adjamé (Now Nangui Abrogoua), Côte d'Ivoire.
[2]Laboratory of Constitution and Reaction of Matter, University of Cocody (Now Felix Houphouët Boigny), Côte d'Ivoire.
[3]National Laboratory for Quality Testing, Metrology and Analysis, BP V 174 Abidjan, Côte d'Ivoire.
[4]International Centre for Theoretical Physics (ICTP), Strada Costiera, Trieste, Italy.

Article Info:

Abstract

Background: Worldwide, cervical cancer (CC) is the fourth most common women cancer. It is crucial to develop more effective treatments for this disease. We aim at designing new anticancer compounds with a favorable pharmacokinetic profile, targeting the E6 oncoprotein of human papillomavirus type 16 (HPV16 E6).

Methods: Computer-Aided Molecular Design (CAMD) has been carried out by the elaboration of a Quantitative Structure-Activity Relationship (QSAR) model of molecular complexation, through the correlation between the relative Gibbs free energy (rGFE) and the observed biological activities of a series of flavonoids inhibitors separated in training (TS) and validation sets (VS). Starting from the 3D crystallographic structure of the oncoprotein HPV16 E6 (Protein Data Banck (PDB) input code: 4GIZ), enzyme – inhibitor complexes were built by *in situ* modification of the native ligand (FLAV1, IC_{50}^{exp}= 850 nM), replacing substitution R-groups at nine different positions $R_1 - R_9$.

Results: The complexation QSAR model equation (pIC_{50}^{exp}=-0.5494 $\times$ $\Delta\Delta G_{com}$+5.9983 (1); n=16; R^2=0.98) explains 98% of the variation in biological activity by that of rGFE of the analogues used. The subsequent 3D pharmacophore model (PH4) generated from the active conformations of FLAVS (pIC_{50}^{exp}= 1.0177$\times$ pIC_{50}^{pre} – 0.0927(2); n=16; R^2=0.90) was used as a virtual selection tool to identify new analogues from a virtual library (VL) reaching two digits nanomolar range predicted activity.

Conclusions: The combination of molecular modelling and in silico screening of VL using the PH4 pharmacophore has led to the discovery of new promising anticancer candidates, with favorable pharmacokinetic profiles against HPV16 E6. Among them, the top two predicted respective inhibitory powers IC_{50}^{pre} (50 nM and 61 nM).

Keywords: Cervical cancer, flavonoids, HPV16 E6 oncoprotein, prediction of ADME properties, Quantitative Structure-Activity Relationship (QSAR).

INTRODUCTION

In 2022, cervical cancer was the fourth most common cancer among women worldwide, with approximately 660.000 new cases and around 350.000 deaths[1]. The main cause of cervical cancer is persistent infection with human papillomaviruses (HPV), particularly the high-risk (HR) genotypes 16 and 18, responsible for about 50% and 20% of cases respectively[2,3]. High-risk HPV infections are a major global medical issue, accounting for 99% of cervical cancer and contributing to approximately 5% of all cancers worldwide[4,5]. HPV-induced keratinocyte transformation is a complex process, marked primarily by the intracellular accumulation of two key viral oncoproteins, E6 and E7, playing a key role in disrupting cellular control mechanisms, thus promoting progression to malignancy[3]. The oncoprotein E6 of papillomavirus of type 16 (HPV16 E6) plays a crucial role in cervical oncogenesis[4,5]. It promotes the degradation of the p53 protein, a key regulator of the cell cycle and apoptosis, through the E6 ubiquitin ligase (E6AP)[6]. The inhibition

of this destructive interaction is therefore a promising strategy for the treatment of cervical cancer. HPV vaccines have no therapeutic potential with limited access worldwide making design of new treatments a crucial need[7]. Recently, several innovative approaches have been explored E6 oncoprotein inhibition including the use of RNA interference[8] to reduce viral protein expression. On the other hand, inhibitoric peptides[9] have been developed against E6 and p53. Finally, small molecules such as luteolin[10] were identified as potential candidates, due to their ability to restore the tumor suppressive activity of p53.The availability of the E6/E6AP crystal structure in the Protein Data Bank (PDB)[11] opened the gate to precisely identify amino acid residues involved in critical interactions with E6. Among them, residues in the LxxLL binding pocket are particularly important because their key role in the recognition of E6 protein binding partners. Natural products (NP) are drawing interest as anti-cancer agents, particularly due to their accessibility, affordability, and generally lower toxicity compared to other synthetic molecules. Among them, flavonoids stand out particularly, having demonstrated various biological effects, such as antibacterial, antiviral and antioxidant ones. Furthermore, these compounds are known to be able to induce apoptosis in cancer cells, making them promising candidates for the development of anti-cancer treatments based on NP[12].

In this context, Srikanth Kolluru and *al.*[13], reported flavonols, with the most active, FLAV1, exhibits anIC_{50}^{exp} of 850 nM, and the x-rays crystallography structure of "HPV16 E6 – FLAV1"complex (PDB entry code 4GIZ)[13]. Analysis of the interactions from this seminal study and literature exploration led to a series of enough compounds with the same scaffold in order to build QSAR model of HPV16 E6 inhibition.

Therefore, CAMD is suitable to release novel FLAVs analogs with better inhibitory potency. To achieve our objective, we first developed a QSAR model that establishes a correlation between the free energy of Gibbs released during the formation of "HPV16 E6-FLAVx" complexes and their corresponding experimental inhibition activities. Next, we designed a 3D-QSAR pharmacophore model (PH4) for HPV16 E6 inhibition, based on the active conformations obtained through the complexation method. Subsequently, we generated and screened a virtual library using the PH4 model. Finally, we evaluated the predicted activity of the most promising PH4-identified analogs and calculated their ADME profile. In a last step the stability of the best analogs has been checked by molecular dynamics (MD) simulations. The expected results could provide new therapeutic leads in the design of novel small molecule inhibitors for treatment of E6-targeted cancers.

METHODS

In this work, the chemical structures and biological activities studied belong to the flavonoid family. These structures were divided into two sets: one for training set (TS) and the other for validation set (VS).

Training and validation sets
Flavonoids belong to the large family of polyphenols. They share a common basic structure consisting of two benzene rings connected by a linear three-carbon chain, forming an oxygenated heterocycle[14].
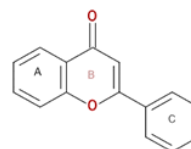

Figure 1: Basic skeleton common to flavonoids.

There are a total of 14 flavones and 5 flavonols, with experimental inhibition concentrations taken from the literature respectively[10,13]. There are a total of nineteen (19) molecules with an experimental concentration of HPV16 E6 oncoprotein inhibition ranging from 850 to 671000 nM respectively, wide sufficiently to allow the construction of a reliable QSAR model. They were distributed as follows:
- 16 compounds for the trainingset (TS) and;
- 3 compounds for model validation (VS)set.
This distribution was done using a protocol called "Generate Training and Test Data" in the Discovery Studio 2.5 software[15].

Model building
The construction of the QSAR model begins with the download of the crystallographic structure of the HPV16 E6–FLAV1 complex (PDB code:4GIZ). Indeed, the Protein – Inhibitor (IP) complexes were made from the crystallographic structure of HPV16 E6–FLAV1 using the Molecular Modelling program Insight II[16]. No water molecule is included in the model. The "HPV16 E6 – FLAVx" complexes were built by *in situ* modifications on the reference structure of "HPV16 E6 – FLAV1" complex with FLAV1 inside the receptor active site by replacing atom by atom, the appropriate fragments of the molecular structure of FLAV1. During these *in situ* modifications, the addition of an atom requires a local minimization within a radius of 5 Å around it. When adding a group of atoms, a systematic conformational search around the torsion angles (torsion drive), is necessary, followed by a global minimization of the protein-ligand complex. This procedure has a double advantage to avoid explosive steric contacts and take into account the reality (flexibility of the lateral chains of the receptor active site residues).

The complete description of the computation of relative ligand binding affinity ($\Delta\Delta G_{com}$) has been reported earlier[17].

$\Delta\Delta G_{com} = \Delta G_{com}(I) - \Delta G_{com}(I_{ref}) = \Delta\Delta H_{MM} - \Delta\Delta TS_{vib} + \Delta\Delta G_{sol}$ (1).

$\Delta\Delta H_{MM}$ describes the relative enthalpic contribution to the GFE change related to the intermolecular interactions in E:I complex derived by molecular mechanics (MM), $\Delta\Delta G_{sol}$ and $\Delta\Delta TS_{vib}$ represent, respectively, the relative solvation rGFE and simplified relative vibrational entropy.

Molecular mechanics

The FLAVs, HPV16 E6 –FLAVx complexes and HPV16 E6 have been refined by Molecular Mechanics (MM) using methodology previously described earlier[18].

Solvation Gibbs free energies

The electrostatic component of solvation-free energy (GFE), which incorporates the effects of ionic force through the resolution of the nonlinear Poisson-Boltzmann (PB) equation[19], was calculated by the Delphi module in the Discovery Studio[15] suite as previously detailed[18].

Molecular complexation QSAR

First, the thermodynamic quantities $\Delta\Delta H_{MM}$, $\Delta\Delta G_{sol}$, $T\Delta\Delta S_{vib}$, and $\Delta\Delta G_{com}$ were calculated for each ligand in the test set. The ligand FLAV1 being the most active of the set ($IC_{50}^{exp}= 850nM$) was chosen as the reference ligand. The values of the variation of the relative free enthalpy of complexation are obtained by the equation:

$\Delta\Delta G_{com}= \Delta G_{com} (I) - \Delta G_{com} (I_{ref})$

$\qquad =\Delta\Delta H_{MM} - T\Delta\Delta S_{vib} + \Delta\Delta G_{sol}$ (3) with:

- $\Delta\Delta H_{MM}$: relative variation of the complexation enthalpy for each inhibitor;

- $T\Delta\Delta S_{vib}$:relative variation of the complexation entropy;

- $\Delta\Delta G_{solv}$: relative variation of complexation solvation enthalpy.

Then, the experimental biological activities IC_{50}^{exp} were correlated to the variation of the Gcom complexation enthalpy (model in solvent) relative to the reference ligand. Finally, to validate this 3D-QSAR model of molecular complexation, the statistical indicators of reliability of linear regression were calculated.

Interaction energy

The molecular mechanics (E_{int}) interaction energy calculation protocol available in Discovery Studio 2.5[15] was used to determine unrelated interactions, including van der Waals and electrostatic interatomic potential terms, between two groups of atoms belonging one to P and the other to I in the P:I complexes to reach E_{int} breakdown into contributions of residues from individual as fully described earlier[18].

Pharmacophore generation

The linked conformations of inhibitors, extracted from P:I complex models were used to build the 3D QSAR pharmacophore using the HypoGen algorithm of Catalyst[20], integrated in Discovery Studio[15]. The generation process is carried out in three main stages using the set of most active inhibitors: (i) the construction phase, (ii) the subtraction phase and (iii) the optimization phase, as described in the work of Kouassi *et al.*[17]. Inactive molecules were used to define the volume excluded. Five HypoGen features were selected: hydrogen binding acceptor (HBA), hydrophobic, negative ionizable (Neg-Ionizable), heavy hydrogen binding acceptor (HBA heavy) and projection hydrogen bonding acceptor (HBA projection). The majority of protocol parameters are maintained at their default value except of the biological activity uncertainty set to 1.5. The uncertainty is a value between 1 and 3. This adjustment has reduced the uncertainty interval of experimental

biological activity from a wide range $\left[\frac{IC_{50}}{3} ; 3IC_{50}\right]$ to a relatively narrow range $\left[\frac{4IC_{50}}{5} ; \frac{5IC_{50}}{4}\right]$ due to the accuracy and homogeneity of the values of the available experimental biological activities[10,13]. The top ten pharmacophores were generated with a missing number of functionalities to the PH4 hypothesis set at 0. No new conformers are created when developing the PH4 model. The active site conformation is preserved during PH4 creation and superposition. Finally, the best pharmacophore model is selected.

Virtual library generation

Due to the high number of substituents (R_1 to R_9) on the scaffold, the generation of the virtual library (VL) was done using the molecular modeling and simulation software MOE 2015[21] (Molecular Operating Environment). This would require first to install and configure MOE 2015[22] so that all the functionalities are accessible and the necessary licenses are activated. Then, prepare the molecular structures, which consists of importing structures (Load your molecules via files of type SDF, MOL, PDB or other supported formats). Molecules can be imported from an external database or by drawing them directly in MOE[22]. In this work, we imported the common basic structure of inhibitors in SD format that we converted to a SDF file type, taken into account by MOE and then we redefined the order of appearance of the different R_{Groups} on the scaffold, we also specified their attachment point using the MOE[22] "builder" menu and then draw the fragments. Finally, we used the option "Databases>Build Database" to generate a database from the imported scaffold and fragments drawn.

ADME properties

The ADME properties of inhibitors were calculated using the QikProp[i]program based on Jorgensen's methods[2,3,4] as detailed previously[18].

Pharmacophore-based library searching

The pharmacophore model (PH4) mentioned above was derived from the conformations of the FLAVs linked to the active site of HPV16 E6. The virtual library was then scanned using the "pharmacophore" mapping protocol available in Discovery Studio 2.5[15]. as described earlier[18].

Inhibitory potency prediction

The conformer with the best match for PH4 pharmacophore in each group of the target library subset was selected for *in silico* screening by complexation QSAR model. The relative GFE of the formation of the P:I complex in water ($\Delta\Delta G_{com}$) was calculated for each new analogue selected and then used to predict their inhibitory power of HPV16 E6.

$$pIC_{50}^{pre} = -\log_{10}IC_{50}^{pre} = a \times \Delta\Delta G_{com} + b$$

Molecular Dynamics Simulations

In order to check 3D P:I complexes stability, molecular dynamics (MD) simulations to evaluate the stability of complexes formed between HPV16 E6 and flavonoid inhibitors (FLAVs), as well as the conformation flexibility of the active ligand FLAV1 and its three new potentials analogues *in silico* at the active site of HPV16 E6. The complexes obtained by in situ modification of the reference inhibitor FLAV1, followed by refinement through molecular mechanics

(MM) methods, were used as starting geometries for the molecular dynamics simulations performed with the software Desmond[28]. Table 8 shows the averages of total energy (E_{tot}) and potential energy (E_{pot}) of the systems studied during the 2000time step simulation. Figure 9 shows the temporal evolution of the properties of the linked inhibitor, such as mean quadratic deviation (RMSD) from initial conformation, the radius of gyration (rGyr), the number of intramolecular hydrogen bonds (intraHB), the molecular surface (molSA), the solvent-accessible surface (SASA)and the polar surface (PSA). The interactions between the protein and ligand were examined throughout the simulation trajectory to identify specific interactions that were maintained during the calculation (Figure 10). Finally, interactions that occur more than 20% of the time during the simulation (from 0 to 200 ns) are illustrated in a detailed 2D diagram (Figure 11). Furthermore, we superimposed the conformations of the ligands obtained after minimization of the complexes from the molecular dynamics simulations with those derived from the in situ modification and refinement of FLAV1 by molecular mechanics.

RESULTS

Training and validation sets

All sixteen (16) training set compounds and all three (3) validation set compounds, including their experimental inhibitory activities, were retrieved from the literature and they are listed in Table 1[10,13].

Table 1: Training and validation sets of HPV16E6 inhibitors used for the design of the Quantitative Structure Activity Relation (QSAR) model.
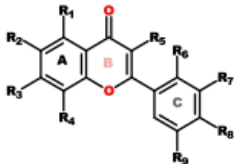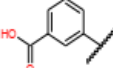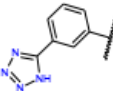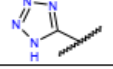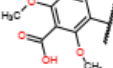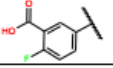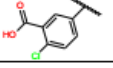

Training set									
Ligands	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	IC_{50}^{exp} (nM)
FLAV1	OH	OH	OH	H	OH	H	OH	OH	OH	850
FLAV2	H	(tetrazole)	H	H	H	H	CF_3	H	H	1 100
FLAV3	OH	H	OH	H	OH	OH	H	OH	H	4 000
FLAV4	H	(tetrazole)	H	H	H	H	OCH_3	H	H	5 200
FLAV5	H	(carboxyphenyl)	H	H	H	H	OH	H	H	5 200
FLAV6	H	(tetrazolylphenyl)	H	H	H	H	H	H	H	6 900
FLAV7	H	(tetrazole)	H	H	H	H	H	H	H	8 100
FLAV8	H	(dimethoxy carboxyphenyl)	H	H	H	H	H	H	H	12 500
FLAV9	OH	H	OH	H	H	H	OH	OH	H	23 000
FLAV10	H	OH	H	H	OH	H	H	H	H	40 000
FLAV11	OH	OH	OH	H	OH	H	OH	OH	H	40 000
FLAV12	H	H	S	H	H	H	H	H	H	47 300
FLAV13	H	H	(fluoro carboxyphenyl)	H	H	H	H	H	H	48 000
FLAV14	OH	H	OH	H	H	H	H	OH	H	382 000
FLAV15	OH	H	H	H	H	H	H	H	H	540 000
FLAV16	OH	H	H	H	H	H	OH	OH	H	671 000
Validation set									
Ligands	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	IC_{50}^{exp} (nM)
FLAV17	OH	H	OH	H	OH	H	H	H	H	40 000
FLAV18	H	H	(chloro carboxyphenyl)	H	H	H	H	H	H	62 200
FLAV19	OH	H	(acetoxy)	H	H	H	(diester)	(ester)	H	77 000

Table 2: Gibbs free energy (binding affinity) and its components for the HPV16E6 FLAV1-16 inhibitor drive assembly and the FLAV17-19 inhibitor validation assembly.

Training set [a]	M_W[b]	$\Delta\Delta H_{MM}$[c]	$\Delta\Delta G_{sol}$[d]	$\Delta\Delta TS_{vib}$[e]	$\Delta\Delta G_{com}$[f]	IC_{50}^{exp}[g] [nM]
FLAV1	318	0.00	0.00	0.00	0.00	850
FLAV2	358	0.85	-1.15	-0.62	0.31	1100
FLAV3	302	2.48	0.09	1.48	1 .10	4000
FLAV4	320	1.94	-0.54	0.40	1.00	5200
FLAV5	358	0.21	1.07	0.27	1.01	5200
FLAV6	366	1.93	-0.22	-0.18	1.90	6900
FLAV7	290	2.45	0.55	1.38	1.62	8100
FLAV8	402	3.66	0.74	2.12	2.28	12500
FLAV9	286	5.95	-1.24	2.51	2.20	23000
FLAV10	254	5.66	-2.27	0.65	2.74	40000
FLAV11	302	1.51	-1.09	-2.47	2.88	40000
FLAV12	342	1.57	2.36	1.09	2.84	47300
FLAV13	360	0.58	2.63	-0.14	3.35	48000
FLAV14	270	7.17	0.99	3.17	4.98	382000
FLAV15	238	9.37	-2.20	2.52	4.65	540000
FLAV16	270	6.29	0 .29	1.44	5.14	671000
Validation set	M_W[b]	$\Delta\Delta H_{MM}$[c]	$\Delta\Delta G_{sol}$[d]	$T\Delta\Delta S_{vib}$[e]	$\Delta\Delta G_{com}$[f]	$\frac{pIC_{50}^{pre}}{pIC_{50}^{exp}}$[h]
FLAV17	270	6.39	-2.12	1.62	2.65	0.81
FLAV18	376	1.23	2.11	0.50	2.83	0.87
FLAV19	412	5.96	-2.58	-0.63	4.01	0.91

[a]for the chemical structures of the training set of inhibitors see Table 1; [b] Mw (g/mol) is the molar mass of inhibitors;[c]$\Delta\Delta HMM$ (kcal/mol) is the relative enthalpic contribution; [d]$\Delta\Delta Gsol$ (kcal/mol) is the relative solvent effect contribution to the GFE change of E-I complex formation [e]$-\Delta\Delta TSvib$ (kcal/mol) is the relative entropic contribution of inhibitor to the GFE of P-Ix complex formation; [f]$\Delta\Delta Gcom$ (kcal/mol) is the overall relative GFE change of P-Ix complex formation: $\Delta\Delta G_{com} \approx \Delta\Delta H_{MM} + \Delta\Delta G_{sol} - \Delta\Delta TS_{vib}$; [g]$IC_{50}^{exp}$ (nM) is the experimental inhibitory concentration of FLAV obtained from ref[10,13]; [h]ratio of predicted and experimental half-maximal inhibition concentrations.

Their IC_{50}^{exp} cover a relatively wide range (about $10^{2.9}$) 850 nM $\leq IC_{50}^{exp} \leq 671000$ nM sufficient to allow the construction of a valid QSAR model.
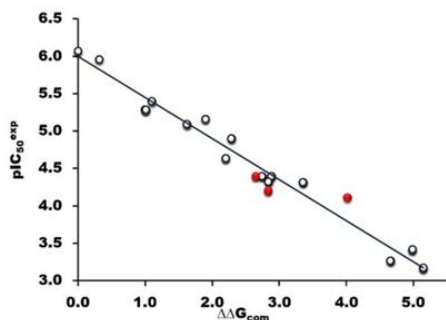

Figure 2: Graph of the correlation between pIC_{50}^{exp} and rGFE$\Delta\Delta G_{com}$, HPV16E6-FLAVs in aqueous medium.
White dots indicate the training set point cloud and three red dots the validation set point cloud.

One-descriptor QSAR model
The relative Gibbs free energy (rGFE) of the P:I complex formation ($\Delta\Delta G_{com}$) was calculated for the complexes HPV16 E6: FLAVs. The 16 formation complexes (TS) and V complexes (VS) were prepared respectively by *in situ* modification of the crystal structure of the refined model (input code PDB 4GIZ[13] of HPV16/E6APin complex with FLAV1) as described in the Methods section. Table 2 lists the values of the rGFE ($\Delta\Delta G_{com}$). The QSAR model obtained highlights the correlation between the experimental inhibition potency of flavonoids and the variations in relation to a reference of the rGFE ($\Delta\Delta G_{com}$) calculated by linear regression. The statistical data of linear regression are given in Table3. $\Delta\Delta G_{com}$ reflects the mutual binding affinity between the protein and the inhibitor. The resulting QSAR model of molecular complexation explains about 98% of the variation in biological activity by $\Delta\Delta G_{com}$ (Figure 3).

Table 3: Linear regression equation of the QSAR model of complexation, statistical model reliability parameters and experimental inhibitory concentrations interval IC_{50}^{exp}.

| Statistical data of linear regression	
$pIC_{50}^{exp} = -0.5494 \times \Delta\Delta G_{com} + 5.9983$	
Number of compound n	16
Squared correlation coefficient of regression R^2	0.98
LOO cross-validated Squared Correlation coef. R^2_{XV}	0.97
Standard error of regression σ	0.14
Statistical significance of regression. Fisher F-test	456.94
Level of statistical significance α	>95%
Range of activities IC_{50}^{exp} [nM]	850-671000
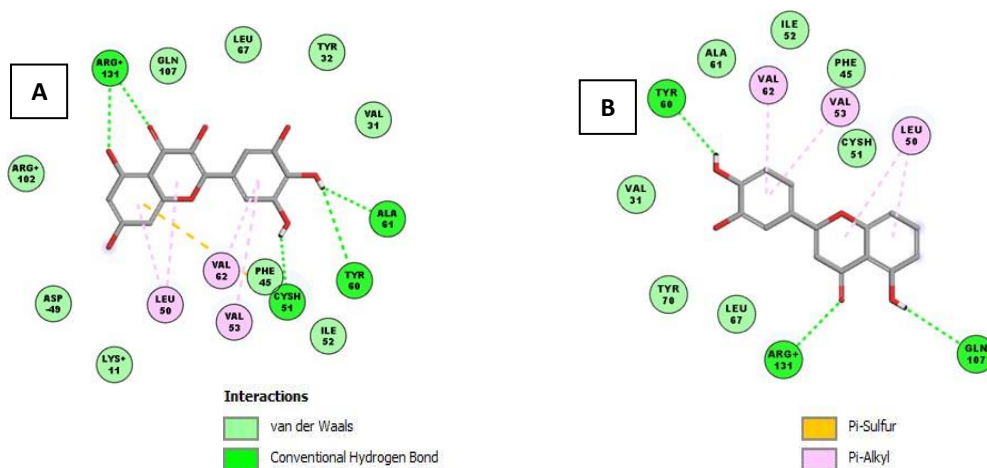
Figure 3: (A) 2D Diagram showing the interactions of the most active inhibitor, FLAV1, within the active site of HPV16 E6. (B): 2D diagram showing the interactions of the least active inhibitor, FLAV16, within the active site of HPV16 E6.
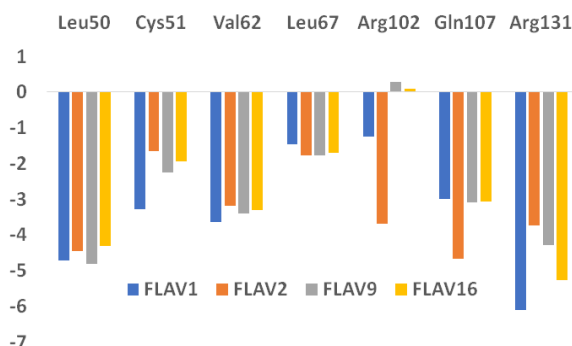

Figure 4: Molecular mechanics intermolecular interaction energy Decomposition (in kcal/mol) of the contributions of residues for FLAV1, FLAV2, FLAV9 and FLAV16.

Binding mode of FLAVs

The analysis of interactions of HPV16 E6 − FLAV1 and HPV16 E6 − FLAV16 complexes corroborates the experimental inhibitory activities obtained by Srikanth Kolluru *et al*[13]. Indeed, Figure 3 below displays the binding mode of the most active ligand (FLAV1) and least active (FLAV16) in the active site of HPV16 E6. We can see Arg131 in two hydrogen bonds (HB) contact with FLAV1 instead of one with FLAV16. The oxygen atom binding domain at R_8 of the FLAV1 is in HB with Ala61and Tyr60 against only one with Tyr60

for FLAV16. FLAV1 is in one HB with the catalytic residue Cys51 whereas its loss with FLAV16.

These specific HB contacts reported by Kolluru *et al.*[13], were observed for the new class of HPV16 E6 inhibitors, except the interaction between the inhibitor and Tyr 60 which was not reported prior to the work of Kolluru *et al.*[13], and is observed in this work. The superiority of interactions, notably of FLAV1's hydrogen with residues from the active site of HPV16 E6 as opposed to FLAV16, could explain the important stability of FLAV1 within the active site of the E6 protein compared to other ligands in the series.

Table 4: Parameters of 10 PH4 hypotheses generated for the inhibition of HPV16 E6 after the Cat-Scramble validation procedure (49 tests scrambled for each hypothesis with a confidence level of 98%).

Hypothesis	RMSD [a]	R^{2b}	Total costs[c]	Costs difference[d]	Closest Random[e]	Feature[f]
Hypo1	1.473	0.95	73.3	150.6	77.0	HBA, HBA, HBA, HBA, HYD
Hypo2	1.984	0.91	90.5	133.3	102.4	HBA, HBA, HBA, HBA
Hypo3	2.427	0.86	102.5	121.3	106.0	HBA, HBA, HBA, HBA, HBA, HBA
Hypo4	2.429	0.86	102.8	121.1	107.3	HBA, HBA, HBA, HBA, HBA, HBA, HBA
Hypo5	2.523	0.85	105.1	118.7	107.9	HBA, HBA, HBA, HBA
Hypo6	2.628	0.84	109.2	114.7	110.3	HBA, HBA, HBA, HBA, HBA
Hypo7	2.805	0.81	117.0	106.9	110.5	HBA, HBA, HBA, HBA
Hypo8	2.807	0.81	117.0	106.8	111.5	HBA, HBA, HBA, HBA
Hypo9	2.791	0.82	117.4	106.5	112.4	HBA, HBA, HBA, HBA, HBA, HYD
Hypo10	2.786	0.82	117.8	106.1	112.4	HBA, HBA, HBA, HBA, HBA, HBA

[a]Root Mean Square Deviation; [b] Squared correlation coefficient; [c]Overall cost parameter of the PH4; [d]Cost difference between Null cost and hypothesis total cost; [e] Lowest cost from 49 scrambled runs at a selected level of confidence of 98%. [f]HBA (hydrogen-bond Acceptor); HYD (Hydrophobic). The Fixed Cost=53.9with RMSD=0, Null Cost=223.9 with RMSD=4.82 and the Configuration cost=14.7
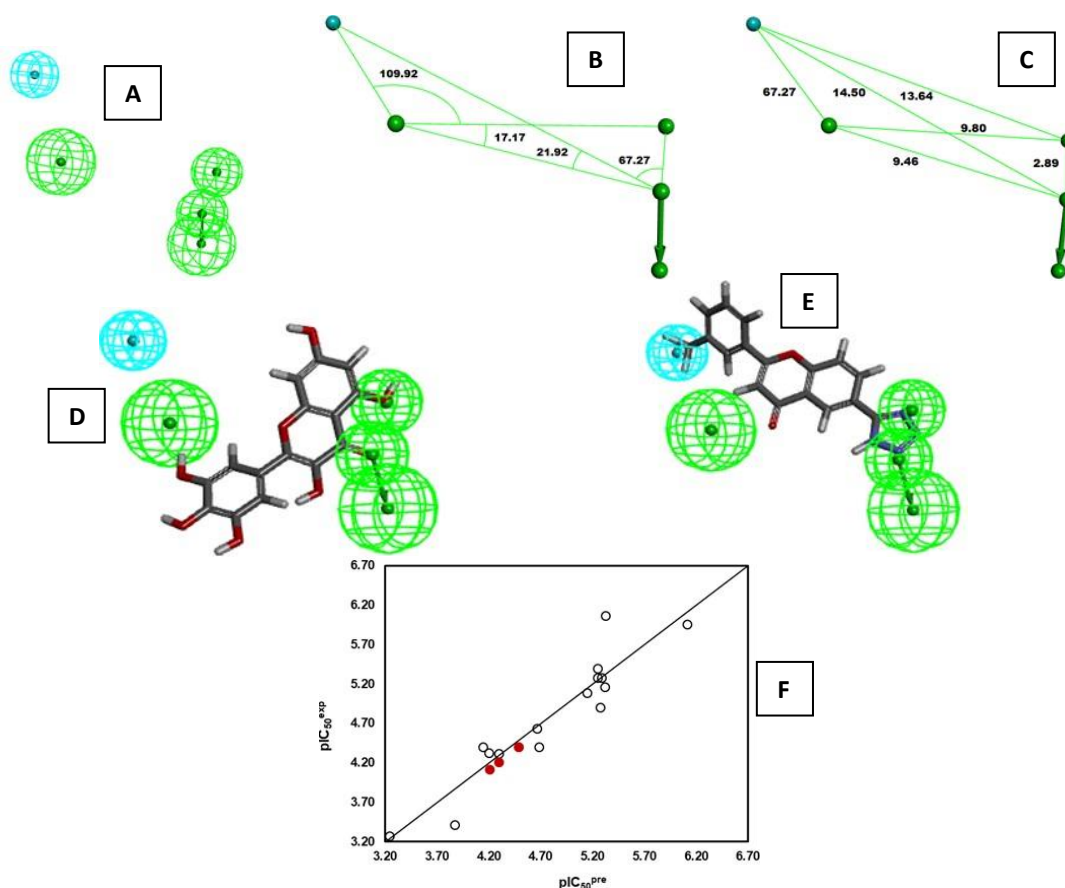
Figure 5: Features of the best PH4 model (Hypo1) of HPV16 E6 inhibitors generated by the 3D-QSAR pharmacophore module: (A)coordinates of centres: (B, C) angles and distances of centres, (D)Mapping tohypo1 features with FLAV1: (E)Mapping to the hypo1 features with FLAV2: (F)Correlation plot of experimental vs. predicted inhibitory activity.

The data for the validation set are shown in red color; the circles represent FLAVs from the training set. Feature legend: HYD=Hydrophobic (cyan), HBA=Hydrogen bond Acceptor (green).

Interaction Energy

The distribution of contributions from HPV16 E6 active site residues is useful to select substituents (R groups) able to improve the binding affinity of FLAV analogues with HPV16 E6, and thus increase their inhibitory power. Indeed, in the SAR (Structure-Activity Relationship) study by Jonathan J., Cherry *et al.*[10]. On the training setflavones it was reported that the lack of the hydroxyl on the scaffold ring A and substitution on benzene or a heterocycle in ortho position restored the activity which jumped from 23 000nM (FLAV9) to 1100 nM (FLAV2).

Table 5: R groups (fragments, substituent's) proposed for the design of the virtual library of FLAVs analogues.

#R [a, b, c, d]	1	2	3	4	5	6	7	8
R-Group	H							

#R	9	10	11	12	13	14	15
R- Group							

#R	16
R-Group	

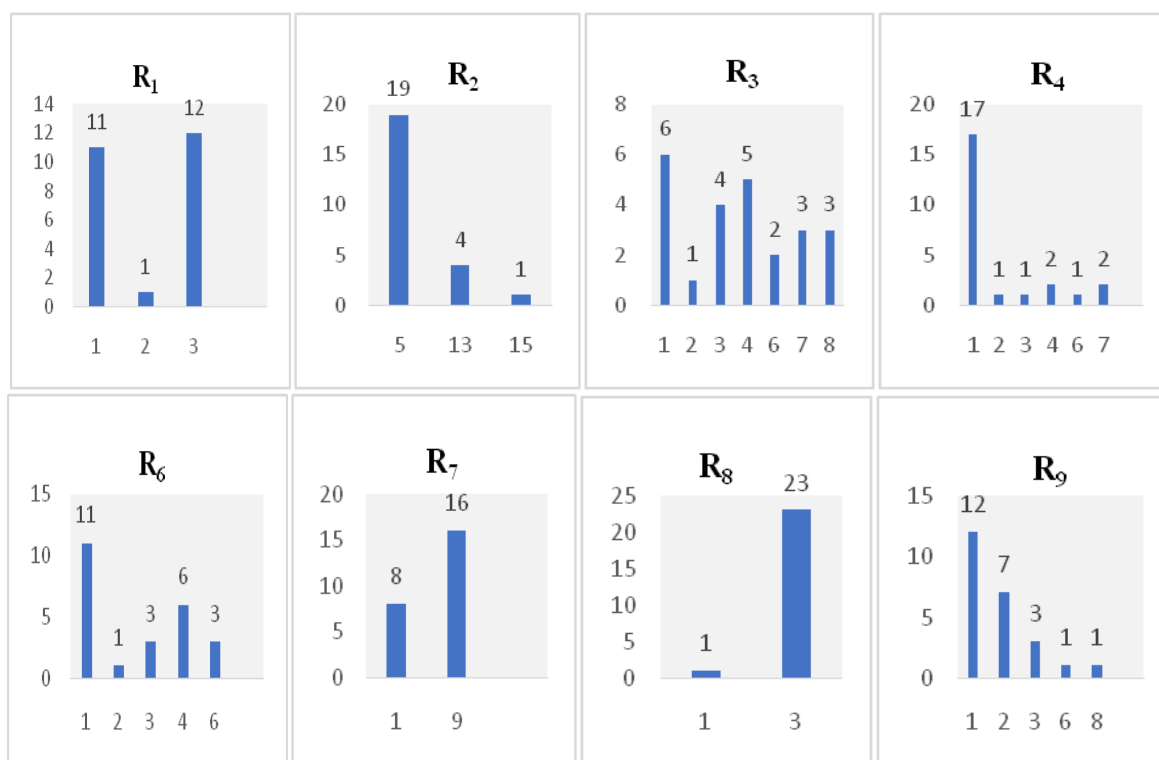**Figure 6: Histograms of the frequencies of appearance of groups R₁, R₂, R₃, R₄, R₆, R₇, R₈ and R₉ for the 24 best hits from screening by hypothesis 1.**

Table 6: GFE (ΔΔG_com) complexation and its components for 24 top-scoring FLAV virtual analogues. Analogue numbering concatenates the index of each substituent R₁ to R₉ with substituent numbers taken from Table 5.

N°	Analogues designed R₁-R₉	$M_w{}^a$ (g/mol)	$\Delta\Delta H_{MM}{}^b$ (kcal/mol)	$T\Delta\Delta S_{vib}{}^c$ (kcal/mol)	$\Delta\Delta G_{sol}{}^d$ (kcal/mol)	$\Delta\Delta G_{com}{}^e$ (kcal/mol)	$IC_{50}{}^{pre\,f}$ (nM)
	FLAV1	318	0	0	0	0	850 ᵍ
1	1-5-1-1-1-4-9-3-2	450	-4.9	-1.4	-0.8	-4.2	5
2	1-5-1-4-1-6-1-3-1	381	-3.9	0.6	2.5	-2	80
3	1-5-3-1-1-4-9-3-3	494	-3.2	0.7	2.9	-1	275
4	1-5-3-2-1-4-1-3-1	412	-0.7	-0.8	0.6	0.7	2499
5	1-5-4-1-2-3-1-3-1	412	-2.7	-2.3	0.4	0	999
6	1-5-8-1-1-2-1-3-2	386	-0.1	0.9	0.9	-0.2	828
7	1-5-8-7-1-1-1-3-2	399	-1	0.2	0.2	-1.1	254
8	1-13-1-1-2-6-9-3-8	436	1.5	-2.7	-3.1	1.1	3893
9	2-15-6-1-1-4-1-1-1	417	-9.7	-4	3.3	-2.4	50
10	1-13-4-1-1-1-1-3-2	380	0.2	2.2	2	0	1022
11	1-13-4-1-2-1-9-3-1	448	1.7	-2.5	-1.3	2.9	4165
12	1-13-7-1-1-4-9-3-1	461	-2.4	0.6	0.7	-2.2	61
13	3-5-1-3-2-1-9-3-6	481	1.7	2.8	0.5	-0.6	464
14	3-5-1-7-2-1-9-3-3	495	3.7	0.1	-0.5	3	45880
15	3-5-2-1-1-3-1-3-2	414	-0.1	1.2	-0.6	-1.9	88
16	3-5-3-1-2-1-9-3-2	482	1.2	1	0.6	0.9	3107
17	3-5-3-4-1-1-9-3-1	494	3.4	2.4	-1.3	-0.4	628
18	3-5-4-1-1-1-9-3-2	480	0.9	-1.6	0.2	2.6	28284
19	3-5-4-1-1-6-9-3-1	479	-1.3	-1.2	1.1	1.1	3830
20	3-5-6-1-2-1-9-3-1	451	5.9	2.7	-0.6	2.7	29568
21	3-5-7-1-1-4-9-3-1	493	-1.3	0.9	2.7	0.5	1944
22	3-5-7-1-2-1-9-3-3	495	5.1	2.1	0.8	3.8	116534
23	3-5-8-1-2-1-9-3-1	468	-2.5	0.9	4.2	0.8	2793
24	3-5-1-6-1-4-9-3-1	479	1.7	-0.5	2.5	1	3609

ᵃ Mw is the molar mass of the inhibitor; ᵇ ΔΔHMM is the relative enthalpic contribution to the GFE change (ΔΔGcom) of HPV16 E6 - FLAV complex formation (fordetails, see Table 2); ᶜΔΔGsol is the relative solvation GFE contribution to ΔΔGcom; ᵈTΔΔSvib is the relative entropic (vibrational) contribution to ΔΔGcom ; ᵉ ΔΔGcom is the relative change in Gibbs free energy related to the formation of the HPV16 E6 -FLAV protein-inhibitor complex ΔΔGcom ≡ ΔΔHMM - TΔΔsvib + ΔΔGsol ; ᶠ IC₅₀ᵖʳᵉ is the predicted inhibition potency towards HPV16 E6 calculated from ΔΔGcom using the correlation equation(Table 3); ᵍ IC₅₀ᵉˣᵖ is given for the reference inhibitor FLAV1 instead of IC₅₀ᵖʳᵉ.

Table 7: Predicted ADME-related properties of FLAV analogues better mapped to hypo1 and known anticancer of HPV16 E6 agents either in clinical use or currently undergoing clinical testing, as computed by QikProp[24].

FLAV$_a$	#Stars$_b$	M$_{wc}$[g.mol-1]	S$_{mol}$[Å2]	S$_{mol,hfob}$[Å2]	V$_{mol}$[Å3]	RotB$_g$	HB$_{donh}$	HB$_{acci}$	logP$_{o/wj}$	logS$_{watk}$	logK$_{HSAl}$	logB/B$_m$	BIP$_{cacon}$[nm.s-1]	#mand$_o$	IC$_{50prep}$[nM]	HOA$_q$	%HOA$_r$
1-5-1-1-1-4-9-3-2	0	318	642.9	173.3	1154.1	3.0	2.0	9.7	2.1	-4.6	-0.4	-1.3	31.6	4.0	5	2.0	66.3
1-5-1-4-1-6-1-3-1	0	381	616.4	181.2	1089.4	3.0	2.5	10.0	1.0	-3.7	-0.5	-2.0	11.0	4.0	80	2.0	51.1
1-5-3-1-1-4-9-3-3	0	494	704.1	333.8	1294.1	4.0	1.0	10.5	2.9	-4.9	-0.3	-1.1	73.2	5.0	275	3.0	77.3
1-5-8-1-1-2-1-3-2	0	386	615.4	174.4	1075.0	4.0	2.8	9.0	1.2	-4.1	-0.3	-1.5	117.0	6.0	828	3.0	71.0
1-5-8-7-1-1-1-3-2	0	399	640.8	208.7	1128.1	5.0	3.8	9.2	0.6	-3.0	-0.3	-1.4	21.3	8.0	254	2.0	54.3
2-15-6-1-1-4-1-1-1	0	417	657.6	0.0	1170.9	5.0	3.5	7.3	2.6	-4.7	-0.2	-2.8	0.8	2.0	50	1.0	39.8
1-13-7-1-1-4-9-3-1	0	461	680.9	132.8	1217.1	4.0	4.0	9.3	-0.6	-4.6	-0.2	-2.0	1.1	4.0	61	1.0	24.0
3-5-1-3-2-1-9-3-6	0	481	686.5	306.9	1261.8	5.0	2.5	10.2	2.3	-4.9	-0.1	-1.1	283.0	7.0	464	2.0	84.0
3-5-2-1-1-3-1-3-2	0	414	640.0	312.1	1155.6	5.0	2.0	10.0	1.3	-3.7	-0.3	-1.4	222.5	7.0	88	2.0	76.8
3-5-3-4-1-1-9-3-1	0	494	693.6	291.2	1273.3	4.0	1.0	10.5	3.0	-4.8	-0.3	-0.9	93.4	5.0	628	3.0	79.5
Gemcitabine	0	263	454.0	93.7	737.8	4.0	5.0	9.1	-1.5	-2.1	-0.8	-1.8	35.0	4.0		2.0	45.6
Ifosfamide	0	261	439.1	268.7	740.9	3.0	1.0	8.5	0.8	-1.7	-1.0	0.4	4086.7	2.0		3.0	96.0
Topotecan	0	421	677.4	361.5	1234.4	5.0	2.0	10.5	1.0	-2.9	-0.2	-1.1	53.0	6.0		3.0	63.8
Fluorouracile	5	130	291.0	0.0	426.6	0.0	2.0	3.5	-0.9	-1.1	-0.7	-0.8	155.8	0.0		2.0	61.0

[a] designed FLAV analogs and known anticancer agents, Table 6; [b] drug likeness, number of property descriptors (24 out of the full list of 49 descriptors of QikProp, ver. 3.7,release 14) that fall outside of the range of values for 95% of known drugs; [c] molar mass in [g.mol-1] (range for 95% of drugs: 130–725 g.mol-1) [46] ; [d] total solvent-accessible molecular surface, in [Å2] (probe radius 1.4 Å) (range for 95% of drugs: 300–1000 Å2); [e] hydrophobic portion of the solvent-accessible molecular surface, in [Å] (probe radius 1.4 Å) (range for 95% of drugs: 0–750 Å2); [f] total volume of molecule enclosed by solvent-accessible molecular surface, in [Å3] (probe radius 1.4 Å) (range for 95% of drugs: 500–2000 Å3); [g] number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds (range for 95% of drugs: 0–15) ; [h] estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over several configurations, so they can assume non-integer values (range for 95% of drugs: 0.0–6.0); [i] estimated the number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can assume non-integer values (range for 95% of drugs: 2.0–20.0); [j] logarithm of partitioning coefficient between n-octanol and water (o/w) phases (range for 95% of drugs: 2–6.5); [k] logarithm of predicted aqueous (wat) solubility, logS. S in mol dm´3 is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid (range for 95% of drugs: 6.0–0.5); [l] logarithm of predicted binding constant to human serum albumin (range for 95% of drugs: −1.5 to 1.5); [m] logarithm of predicted brain/blood partition coefficient (range for 95% of drugs: −3.0 to 1.2); [n] predicted apparent Caco-2 cell membrane permeability in Boehringer-Ingelheim scale in [nm s -1] (range for 95% of drugs: < 25 poor, > 500 nm s-1 great); [o] number of likely metabolic reactions (range for 95% of drugs: 1–8); [p] predicted constants IC_{50}^{pre} , was predicted from computed $\Delta\Delta G_{com}$ using the regression Equation (B) shown in (Table 3) ; [q] human oral absorption (1=low, 2=medium, 3=high); [r] percentage of human oral absorption in gastrointestinal tract (<25%=poor, >80%=high); * star in any column indicates that the property descriptor value of the compound falls outside the range of values for 95% of known drugs.

Flavones with those structural modifications are among the most active (1100 nM $\leq IC_{50}^{exp} \leq$ 12500 nM) of the TS compounds; others except for FLAV1 (more active ligand) are largely part of the moderately active ligands 23000 nM $\leq IC_{50}^{exp} \leq$ 48000 nM and less active 382000nM $\leq IC_{50}^{exp} \leq$ 671000nM. For more information on the different interactions between FLAVs and HPV16 E6 responsible for their inhibition potency, we first performed aQSAR model of molecular complexation allowing us to identify the active conformations of the FLAVs and derive the contributions of individual residues filling hydrophobic and hydrophilic pockets in view of the nature of the substituents of the FLAV1, FLAV2, FLAV9 and FLAV16.

The interaction energy analysis is displayed on Figure 4. The comparative analysis of the per residue energetic contribution shows that the individual IE contributions of the catalytic residues of the HPV16 E6 active site are similar for the three classes of inhibitors, except for Arg102, which shows negligible energetic contributions for FLAV9 and FLAV16 as observed on Figure 3.

In lack of structural clues from IE breakdown to enzyme active site residues contribution, we retained a combinatorial approach to the design of new analogues virtual library and screen it with the help of HPV16 E6 inhibition pharmacophore (PH4) derived from the complexation model QSAR. The best hits are evaluated subsequently by our QSAR model (pIC$_{50}^{pre}$ = - 0. 5494 ×$\Delta\Delta G_{com}$+5.9983). $\Delta\Delta G_{com}$ is the relative variation of the relative Gibbs free energy of each best hit formation.

Ligand-Based 3D-QSAR Pharmacophore Model of Inhibitory Activity
The 3D-QSAR pharmacophore protocol of the Discovery Studio molecular modeling[15] program provides the 3D-QSAR pharmacophore for HPV16 E6 inhibition which was generated from the active conformations of the 16 ligands in the training set (FLAV1-16) covering a range of experimental activities 850 nM $\leq IC_{50}^{exp} \leq$ 671 000 nM and evaluated by the other 3 in the validation set[FLAV17

(IC$_{50}$^{exp}=40 000 nM), FLAV18 (IC$_{50}$^{exp}=62 200 nM) and FLAV19 (IC$_{50}$^{exp}=77 000 nM)]. The PH4 3D-QSAR generation was carried out in three stages: constructive, subtractive and optimization (Mehodes section). During the construction phase, FLAV1 was selected as a priority because it is the unique molecule that meets the threshold criterion IC$_{50}$^{exp}≤Uncert × min (IC$_{50}$^{exp}) $^{=\frac{3}{2}}$ × 850.

It was then used to generate the characteristics of the starting pharmacophore. The uncertainty is $\frac{3}{2}$ because the molecules are not from the same laboratory, so the experimental inhibition activities will not be measured under the same conditions. "Uncert" is a value between 1 and 3. During the subtraction phase, HypoGen eliminates the characteristics present in more than half of the molecules it considers as inactive, that is those with IC$_{50}$^{exp}> 850 × 10$^{3.5}$=2 687 936 nM. Therefore, none of the ligands in the formation set were found to be inactive, and no functionality of the initial pharmacophore was removed. Finally, during the optimisation phase, the pharmacophore hypothesis scores were improved. The hypotheses selected are validated on the basis of these scores, which assess the differences in the activity estimates obtained from the regression, as well as the complexity of the pharmacophore, using a simulated annealing approach[16].

FLAV 1-5-1-1-1-4-9-3-2

FLAV 2-15-6-1-1-4-1-1-1

FLAV 1-13-7-1-1-4-9-3-1

Interactions

- □ van der Waals
- □ Conventional Hydrogen Bond
- □ Carbon Hydrogen Bond
- □ Halogen (Fluorine)
- □ Pi-Cation
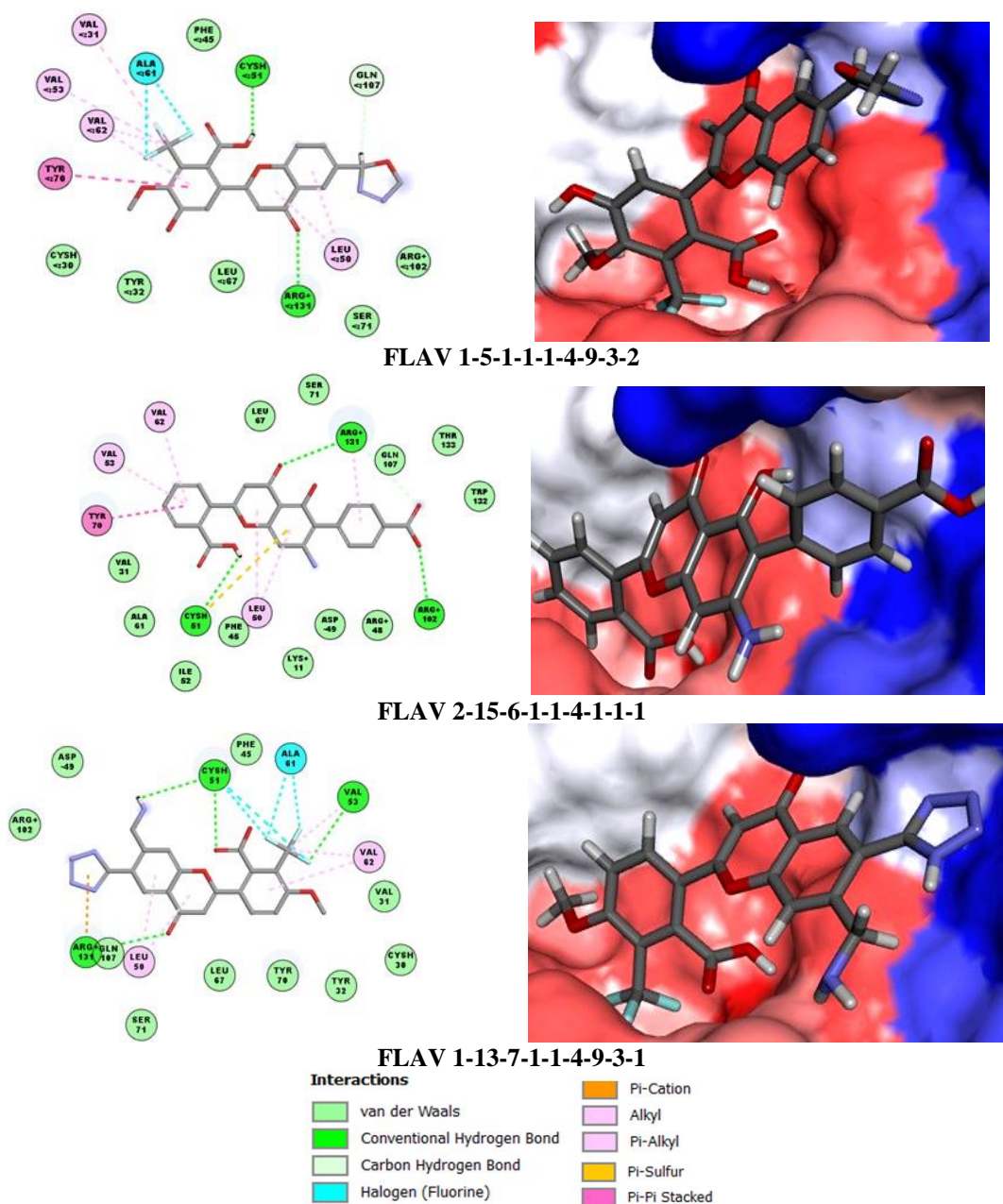- □ Alkyl
- □ Pi-Alkyl
- □ Pi-Sulfur
- □ Pi-Pi Stacked

Figure 7: (Left), Close-up of the virtual response of the three new potential developed analogues FLAV1-5-1-1-1-4-9-3-2, FLAV2-15-6-1-1-4-1-1-1 andFLAV1-13-7-1-1-4-9-3-1 in active site HPV16 E6 and (right), their respective conolly surface.

The binding site surface is colored according to residue hydrophobicity: red—hydrophobic, blue—hydrophilic, and white—intermediate.
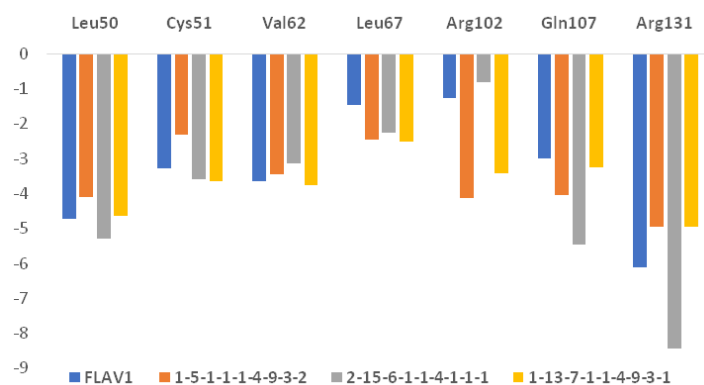
Figure 8: Molecular mechanics Intermolecular interaction energy breakdown to residue contributions, in [kcal.mol⁻¹] shown for the three best designed novel FLAV analogs and FLAV1, most active ligand of training set (the color coding refers to ligands and is given in the legend).

This algorithm applies in fact perturbations to the pharmacophore developed during the construction and subtraction phases in order to optimise the score. After optimisation, the ten best pharmacophore hypotheses were retained. The cost values, correlation coefficients, RMSDs, pharmacophore functionalities and maximum goodness-of-fit values for the ten highest-ranked hypotheses (Hypothesis 1 to Hypothesis 10) are presented in Table 3. These hypotheses were selected on the basis of important statistical criteria, such as a high correlation coefficient, low total cost and a reduced RMSD value.
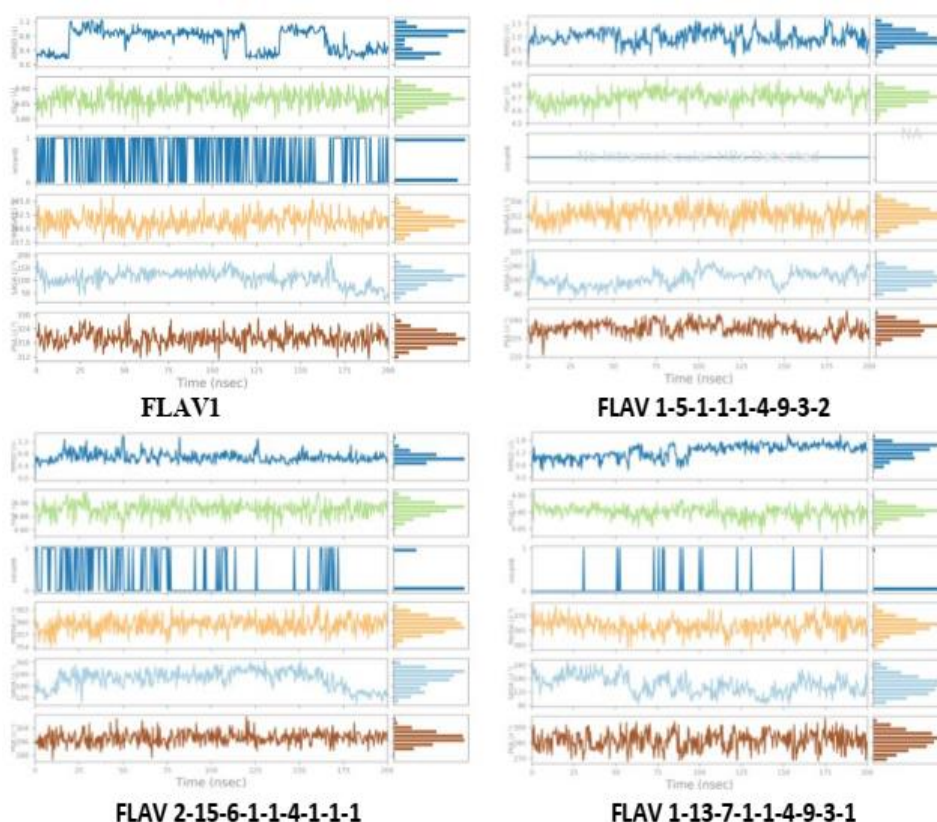

Figure 9: Time-evolution of the properties of the four HPV16 E6-FLAVs complexes during 200 ns MD simulation.

For each inhibitor, top to bottom: plot of the root mean square deviation (RMSD) with respect to the initial conformation vs. the simulation time, radius of gyration (rGyr), number of intramolecular hydrogen bonds (intraHB), molecular surface area (molSA), solvent-accessible surface area (SASA), and polar surface area (PSA).

The reliability of the generated pharmacophore models was assessed as a function of the cost parameters, which ranged from 73.3 (hypothesis 1) to 117.8 (hypothesis 16). The small variation between the cost values reflects the homogeneity of the hypotheses generated and the consistency of the training set. For this pharmacophore model, the fixed cost (53.9) is lower than the zero cost (223.9), with a difference of $\Delta=169.9$. This difference is a key indicator of the quality and predictability of the model ($\Delta>70$ indicates a high probability of more than 90% that the model represents a true correlation).
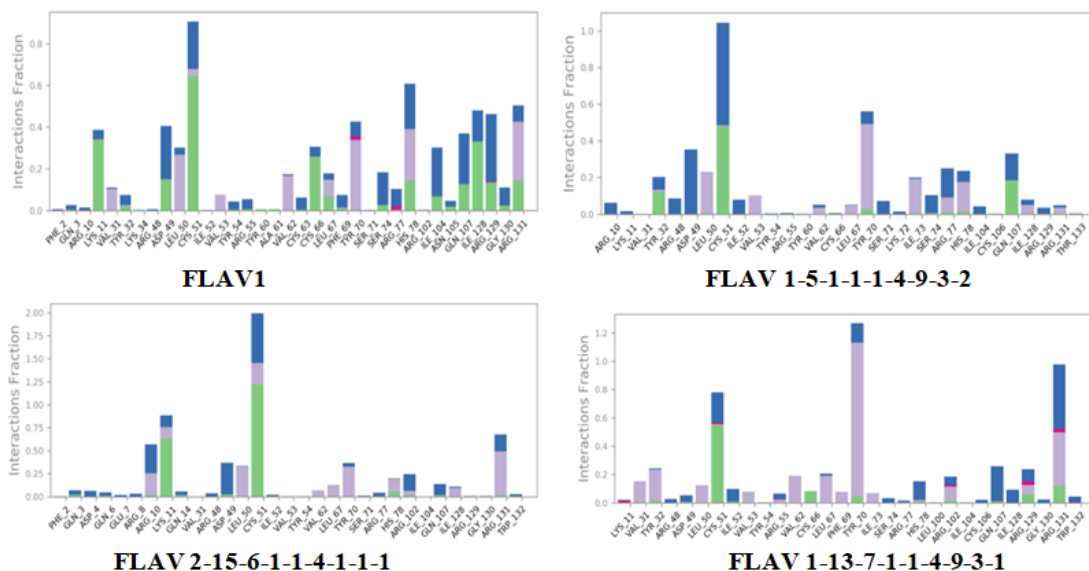
Figure 10: Contribution of individual active site residues to inhibitor binding in HPV16 E6 -FLAVs complexes present during MD simulations.

HB (green); ionic interactions (magenta); hydrophobic contacts (purple); water bridges (blue).

A hypothesis is considered statistically significant if its total cost is close to the fixed cost and far from zero cost. For the 10 hypotheses, the minimum cost difference is 106.1, which indicates the high quality of the pharmacological model. In addition, standard indicators such as the root mean square deviation (RMSD) range from 1.610 (hypothesis 1) to 3.809 (hypothesis 10) and the correlation coefficient (R^2) is between 0.82 and 0.95. The configuration cost, which is 14.68 for all hypotheses, well below 17, confirms the relevance of the model. Thus, the first hypothesis, with a cost of 73.3, close to the fixed cost (53.9) and with the best RMSD and R^2 correlation coefficient values (Table 3), was selected for the *in silico* screening of the virtual library of FLAV analogues. The additional statistical parameters for the regression of pIC_{50}^{exp} against pIC_{50}^{pre} calculated from Hypothesis 1 for the training setare: $pIC_{50}^{exp}=1.0177 \times pIC_{50}^{pre} - 0.0927$ (n=16, R^2=0.90, R^2xv=0.89, F-test=129.07, σ=0.284, α> 98%). These parameters are in accordance with the OECD QSAR guidelines[28].

Table 8: Ensemble averages of the total and potential energies of complexes HPV16 E6-FLAVfor FLAV1and selected virtual hits.

Inhibitors	Structure	E_{tot}^{a} [kcal.mol^{-1}]	E_{pot}^{b} [kcal.mol^{-1}]	IC_{50}^{pre}(nM) c
FLAV1		-69300.711	-85603.486	850^{d}
FLAV1-5-1-1-1-4-9-3-2		-70464.527	-87034.650	5
FLAV2-15-6-1-1-4-1-1-1		-79018.161	-97567.268	50
FLAV1-13-7-1-1-4-9-3-1		-74867.709	-92455.159	61

^{a}Ensemble average of the total energy E_{tot} of the system (Sum of potential energy Epot and kinetic energy E_{kin}); ^{b}Ensemble average of the potential energy E_{pot}; $^{c}IC_{50}^{pre}$ was predicted by QSAR model (Table3, correlation Equation) for the designed analogues. ^{d} Experimental IC_{50}^{exp} for the FLAV1. ***FLAV1** is the co-crystallized ligand in the crystallographic structure of HPV16E6/E6AP (PDB 4GIZ)[11]

FLAV1

FLAV 1-5-1-1-1-4-9-3-2

FLAV 2-15-6-1-1-4-1-1-1

FLAV 1-13-7-1-1-4-9-3-1

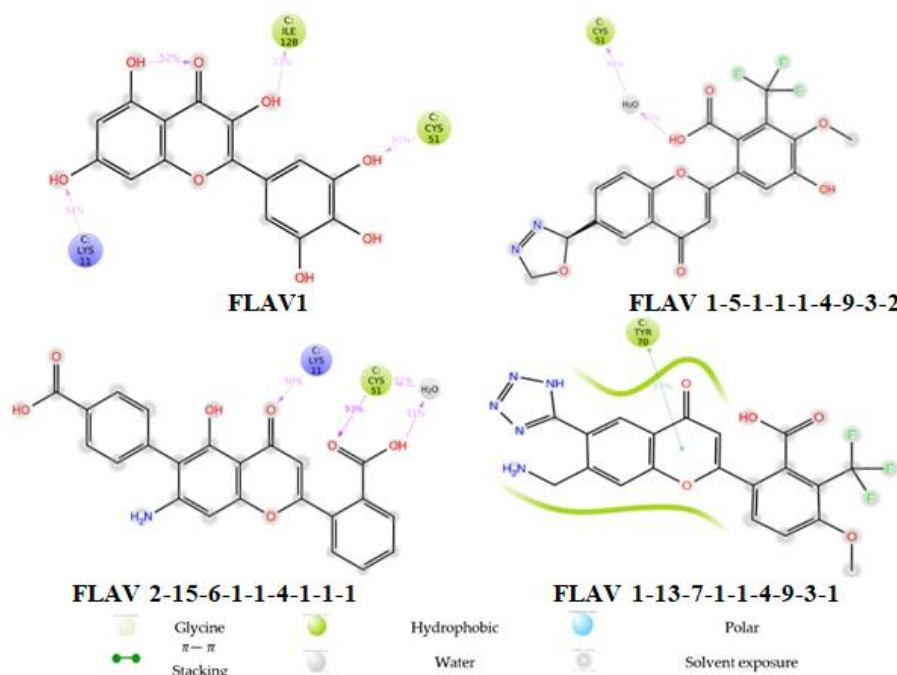|	Glycine		Hydrophobic		Polar
	$\pi - \pi$ Stacking		Water		Solvent exposure

Figure 11: 2D representation of the most populated attractive interactions between the function groups of the four inhibitors and the individual residues at the active site of HPV16 E6 that occur in at least in 1/5 of the 500 analysed frames.

To assess the predictive power of the pharmacophore model, we calculated the ratio between the activities predicted by the PH4 model and those observed experimentally ($pIC_{50}^{pre}/pIC_{50}^{exp}$) for the validation set (FLAV17-19). The following ratios were obtained FLAV17: 1.021, FLAV18: 1.022 and FLAV19: 1.022. All ratios were close to one, demonstrating the strong predictive power of this regression for the optimal PH4 model. Another evaluation of hypothesis 1 is the mapping of the PH4 binding mode in the 3D QSAR (see Figure 5) of the most active flavonols FLAV1 and FLAV2. Jonathan J., Cherry *et al*.[10], reported that the new flavones synthesised on the basis of the FLAV9 structure, notably, FLAV2, are all in a hydrophobic pocketat the interface between HPV16 E6 and ubiquitin ligase (E6AP). Figure 5 illustrates the correlation and the detailed geometry as well as the

position of the hypothesis 1 features for HPV16 E6 inhibition by FLAV1 and FLAV2.

***In silico* screening of FLAVs library**

We have constructed a targeted virtual library (TVL) of novel flavonoid compounds with the aim of contributing to a more potent oral small-molecule targeted therapy against HPV16 E6. To do this, we substituted small fragments (R_1 to R_9) for the different aromatic rings of the FLAV scaffold based on structural information gathered[10], including the absence of hydroxyls on ring A of the FLAV9 scaffold, which if replaced by a substituted benzene or heterocycle in the ortho position(R_2) would restore activity. With regard to the flavone analogues used[10], we note that the presence of fragments (OCH_3, $-CF_3$) in the ortho position would contribute to improve biological activity 1100 nM $\leq IC_{50}^{exp} \leq$ 5200 nM.

FLAV1(RMSD=0.31Å)

FLAV 1-5-1-1-1-4-9-3-2 (RMSD=9.01Å

FLAV 1-13-7-1-1-4-9-3-1 (RMSD=0.46Å)
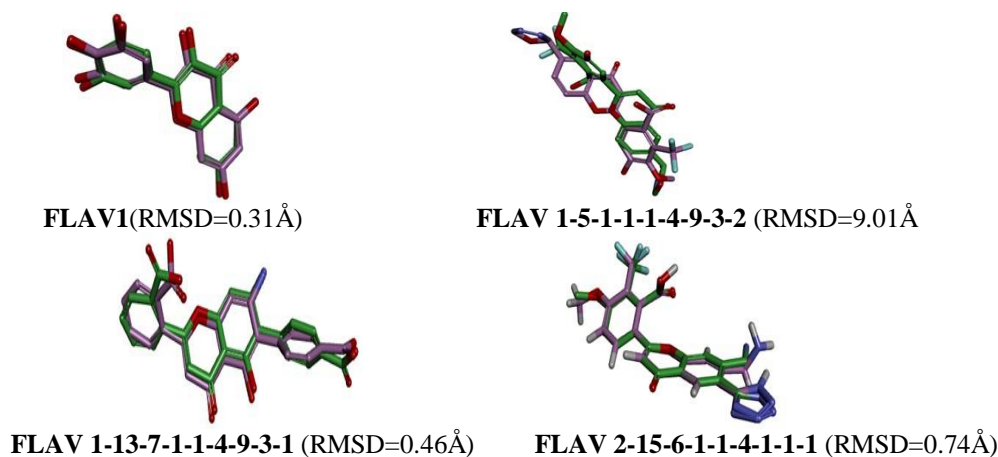
FLAV 2-15-6-1-1-4-1-1-1 (RMSD=0.74Å)

Figure 12: Superimposition of the ligand active conformation from complexes refined by molecular mechanics.

Green carbons, red oxygens, blue nitrogens) and averaged active conformations resulting from MD simulations (purple carbon atoms.

It was also proposed other substituents of a hydrophilic nature that would facilitate hydrogen bond contacts with the residues of the HPV16 E6 active site, which could improve the activity of the analogues proposed by Srikanth Kolluru *et al*[13]. Consequently, depending on the position of R-groups on the FLAVs scaffold, we have specified the type of suitable fragments. Table 5 shows the different fragments associated with the scaffold R-groups. Considering the nature of the substituents proposed on the flavonoid scaffold according to the position of the R groups, a combinatorial library of the following size was constructed: R_1 (7) $\times$ R_2 (5) $\times$ R_3 (7) $\times$ R_4 (7) $\times$ R_5 (2) $\times$ R_6 (7) $\times$ R_7 (8) $\times$ R_8 (7) $\times$ R_9 (7)=9,411,920 new analogues. This library was then filtered without violating Lipinski's rules (molecular weight greater than 500 daltons)[29][30] and Veber's rules, which take into account 5 criteria (validity, efficacy, biocompatibility, economy and reproducibility) [30]. This led us to a virtual library (VL) of 304,628 molecules, screened using the PH4 3D QSAR Hypo1 model of HPV16 E6 inhibition. The best fitting analogues (PH4 best hits) were then screened using the QSAR complexation model. The relative Gibbs free energies ($\Delta\Delta G_{com}$) calculated for HPV16 E6-FLAVs complex formationwith their components and predicted inhibitory concentrations obtained from the QSAR model complexation correlation equation (Table 3 and Table 6).

Analysis of New Inhibitors

The design of the virtual library of new analogues was guided by structural information derived from the active conformation of FLAVs, used to select appropriate substituents at positions R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8 and R_9. In order to identify substituents that could generate new inhibitor candidates with enhanced predictive biological activities against HPV16 E6, we constructed frequency histograms for substituents at positions R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8 and R_9 from the 24 top hits from PH4 screening (Figure 6).

These histograms show that substituents (1) and (3), including hydrogen atom (H) and methoxyl (OCH_3), are the most frequent in position R_1, with frequencies of appearance of 11and 12 respectively.

The fragments in position R_2, in particular fragment (5), acetamidin is the most repetitive and is followed by the substituent (13) pyrrole with frequencies of appearance of 19and 4 respectively.

As for the substituents in position R_3, the substituents with the highest frequency of occurrence are substituents 1(H) and4 (COOH), with frequencies of 6 and 5 respectively.

Concerning the fragments in position R_4, the relative histogram reveals that substituent (1) is the most important with an appearance rate of 17.

As for the frequency of appearance of the fragments in position R_6, the corresponding histogram shows that substituent 1(H) appeared most frequently in this position with a frequency of 11, followed by fragment 4 (COOH) which appeared 6 times. As for the histogram related to the substituents in position R_7, it is observed that fragment 9 (-CF_3) is in first place in this position with a significant number of occurrences of

16, followed by substituent (1) with an occurrence value of 8.

The histogram showing the frequency of appearance of substituents in position R_8 shows that fragment (3), methoxyl (OCH_3) is the most repeated in this position, with a maximum appearance score of 23.

As for the substituents in position R_9, according to the corresponding histogram, we note that the substituent with the most occurrences is hydrogen atom (1) with a frequency of occurrence of 12, and then we have hydroxyl (2) with a score of 7.

It is important to specify that the substituents at position R_5 are substituent 1 (H) or 2 (OH) with frequencies of appearance of 15 and 9, respectively.

In view of the above, the proposed 24 new FLAV analogues are mainly flavones, since the hydrogen atom containing fragment (1) appeared most frequently at position R_5 with an appearance score of 15. The hydroxyl (OH) appeared only 9 times. In addition, the fragment 5 in position R_2, the heterocycle acetamidin has a strong appearance with a maximum score of 19 for 19 analogues out of 5 remaining analogues divided into four appearances of the fragment (13) namely pyrrole and one appearance of benzoic acid (fragment15).

ADME Profiles of Designed FLAVs

The properties related to ADME such as Caco-2 cell permeability, blood-brain partition coefficient, octanol-water partition coefficient, aqueous solubility, number of likely metabolic reactions, serum protein binding and another eighteen descriptors related to absorption, distribution, metabolism, and excretion (ADME) were calculated by the QikProp program[24] for the new best FLAV analogues (Table 7). The method of Jorgensen is used by this program. Empirical data from more than 710 compounds including about 500 drugs and related hetero-cycles were used to produce regression equations correlating experimental and computed descriptors resulting in an accurate prediction of pharmacokinetic properties of molecule. The pharmacokinetic profile of the best designed FLAV analogues is favourable compared with current anticancer compounds in use or in clinical trials.

DISCUSSION

Binding Mode of FLAV

The measurements of HPV16 E6 inhibition by small molecules belonging to the flavonoid family were reported by Srikanth Kolluru *et al*[13]. The Intermolecular stacking interactions between FLAV1(IC_{50}^{exp}=850 nM)and HPV16 E6 of hydrophobic nature (pi-alkyl, pi- sulphur) and hydrogen bonds (see Figure 3) were the main determinants of the better affinity with the target. Using the SAR study of new flavones synthesised on the basis of the chemical structure of FLAV9 against HPV16 E6 by Jonathan J., Cherry *et al.*[10], we were able to extend our data set with known inhibition concentrations for the implementation of the QSAR molecular complexation model. This allowed us to identify the active conformations of the FLAV analogues used.

The specificity of the newly synthesised flavones, resulting from the absence of hydroxyls on the A-ring of FLAV9 in positions R_1 and R_3, replaced by a heterocycle or benzene substituted in the ortho position, in particular R_2, allowed the biological activity to be restored from 23,000 nM (FLAV9) to 1,100 nM (FLAV2), the most active flavone in this series of inhibitors. These molecules bind in a hydrophobic pocket between HPV16 E6 and the ubiquitin ligase (E6 AP). Using the known molecular specificity of FLAVs, we were able to build a virtual library of specific R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8 and R_9 fragments that can be accommodated inhydrophobic or hydrophilic pockets, such as hydrogen bonds.

The virtual screening of new FLAV analogues by the pharmacophore whose hypothesis1 allowed us to identify the best hits. These, evaluated by the QSAR model $(pIC_{50}^{pre} = -0.5494 \times \Delta\Delta G_{com} + 5.9983)$. Among them is a potential new flavone, FLAV 1-5-1-1-1-4-9-3-2$(\Delta\Delta G_{com} = -4.2; IC_{50}^{pre} = 5$ nM), which is substituted in R_5 by the substituent 1(H) and respects the structural modification reported by Jonathan J., Cherry *et al*. In fact, the fragment in ortho position such as R_2 is a substituted heterocycle, in particular acetamidin (fragment 5), and the substituents in positions R_1 and R_3 are all hydrogen atoms (fragment 1).

In addition to this new potential flavone analogue developed *in silico* corroborating with the work of Jonathan J., Cherry *et al*.[10], our study allowed us to discover important new flavones although the cycle A, substituted at position R_2 by a heterocycle, also has substituents in the R_3 or R_4 position, other than fragment 1 (H). Among these we can cite:

- FLAV 1-13-7-1-1-4-9-3-1$(\Delta\Delta G_{com} = -2.22; IC_{50}^{pre} = 61$ nM), substituted in position R_3 by the fragment 7($-CH_2NH_2$), which is 14 times more active than FLAV1;
- and FLAV 2-15-6-1-1-4-1-1-1 $(\Delta\Delta G_{com} = -2.38; IC_{50}^{pre} = 50$ nM) where A-cycle is substituted in the R_1 and R_3 positions respectively by fragment 2 (OH) and fragment 6 ($-NH_2$) and which is 17 times more active than FLAV1. It is important to remember that flavonol FLAV1 is the most active ligand of the training set $(IC_{50}^{exp} = 850$ nM), which is followed by the flavone FLAV2 with $IC_{50}^{exp} = 1100$ nM.

The observed intermolecular interactions (see Figure 7) between HPV16 E6 oncoprotein and new flavones are responsible for the stability of the newly developed flavones, which is reflected in the different $\Delta\Delta G_{com}$ values used to account for it.

In fact, these interactions are hydrogen bonds, van der Waals, hydrophobic and halogen. Referring to Figure 7, it is noted that the interactions of hydrophobic nature are of the type:

- Pi-sulphur, for which the A-cycle FLAV 2-15-6-1-1-4-1-1-1 $(IC_{50}^{pre} = 50$ nM) formed a Pi-sulfur interaction with the residue Cys51;
- Pi-cation, where only the fragment 13 (pyrrole) in R_2 position of cycle A of FLAV 1-13-7-1-1-4-9-3-1$(IC_{50}^{pre} = 61$ nM) formed an interaction pi- cation with Tyr 70;

- Pi-Pi stacked, for which only the C cycle of the new FLAV 1-5-1-1-1-4-9-3-2 $(IC_{50}^{pre} = 5$ nM) and FLAV2-15-6-1-1-4-1-1-1 $(IC_{50}^{pre} = 50$ nM) maintained this interaction with Tyr 70;
- Pi -Alkyl, where virtually A and B cycles of our threenew potential flavones analogues formed a Pi -Alkyl interaction with the Leu50 and their respective C cycle also formed this interaction with the residue Val 62.

The residues of amino acid Leu50, Cys51, Val 62 and Tyr 70 also make from residues belonging to the hydrophobic pocket of HPV16 E6 for different FLAVs in complex such as FLAV9, taxifolin, alizarin[30].

It is important to specify the presence of a fluorine halogen interaction observed in the active conformations of FLAV 1-5-1-1-1-4-9-3-2 $(IC_{50}^{pre} = 5$ nM) and FLAV 1-13-7-1-1-4-9-3-1$(IC_{50}^{pre} = 61$ nM) within the active site of HPV16 E6, in which respectively two atoms of fluorine substituted to the carbon atom ($-CF_3$) whose fragment 9 at position R_7 on the C cycle of the scaffold form respectively a halogen interaction with an electron donor atom of Ala 61 contrary to FLAV 1-13-7-1-1-4-9-3-1$(IC_{50}^{pre} = 61$ nM), where a fluoride atom forms a double halogen interaction with the catalytic residue Cys51 and Ala61 and the other two fluorine atoms form this type of interaction respectively with Cys51 and Ala61.

Thus, this new halogen interaction observed in the active conformations of these best new analogues proposed, marked by the presence of a halogen fragment at R_7 would be promising in the search of more potent oncoprotein HPV16 E6 inhibitors.

Considering the diversity and particularity of the structures of our three new potential analogues designed with 4.9 nM $\leq IC_{50}^{pre} \leq 60.5$ n Min relation to their *in silico* inhibition potency, these molecules are worth being synthesised and evaluated biologically in the development of small therapeutic molecules for the treatment of HPV infection and cervical cancer.

Molecular dynamics simulations

Small differences between the superposition of active FLAV inhibitor conformations modelled by in situ modification of FLAV1 using MM and those obtained via MD (Figure 12), suggest that the best modelled HPV16 E6 -FLAVs complexes are stable for FLAVs2-15-6-1-1-4-1-1-1 $(IC_{50}^{pre} = 49.7$ nM) and 1-13-7-1-1-4-9-3-1$(IC_{50}^{pre} = 60.5$ nM) respectively. These complexes respectively have RMSD values below 3Å (see Figure12), which generally indicates good stability of the complexes during simulation[32]. The FLAV 1-5-1-1-1-4-9-3-2 although by virtual screening and its evaluation by correlation equation of the QSAR model is potentially inhibiting with $IC_{50}^{pre} = 5$ nM, this one, is less stable referring to the molecular dynamics calculation (RMSD=9.01).

Consequently, the FLAVs 2-15-6-1-1-4-1-1-1 $(IC_{50}^{pre} = 50$ nM) and 1-13-7-1-1-4-9-3-1$(IC_{50}^{pre} = 61$ nM) newly designed flavones with predicted inhibitory potency 17 and 14 times that of the most active training set FLAV1 (850 nM) can be proposed for synthesis and subjected to biological evaluation[32].

Limitations of the study

The main limitation of our MM-PB study is the fact that the inhibitors and their experimental inhibition concentrations used for the QSAR model of molecular complexation do not come from the same laboratory, although the relative variation in the free enthalpy of complexation $\Delta\Delta G_{com}$ allowed us to explain the variation in the biological activity of the inhibitors used for the QSAR model of molecular complexation.

CONCLUSIONS AND RECOMMENDATIONS

The study of identification of potential flavonol pocket (FLAV1, FLAV3, FLAV10, FLAV11 and FLAV17) on HPV16 E6 and the structural SAR study of luteolin derivatives (FLAV9) as new target inhibitors, guided us in the preparation of a QSAR model for reliable complexation of HPV16 E6 inhibition which is correlated with the calculated relative Gibbs free energies to form a complex with observed HPV16 E6 inhibition potencies.

In addition, we developed a PH4 3D-QSAR model from active flavonoid conformation (FLAV) using a 16 FLAVs training set and a 3 FLAVs validation set with known experimental inhibition activities. A careful analysis of the interactions between the residues from the active site of HPV16 E6 and the FLAV has directed us to design a first virtual combinatorial library of new analogues of FLAV with multiple substitutions in position R_2 and R_7 for hydrophobic groups and hydrophilic groups in position R_1, R_3, R_4, R_5, R_6, R_8 and R_9. The library screening by matching analogues to the PH4 pharmacophore allowed for selection of a subset of FLAV in the library. This subset of the 24 best virtual results was submitted to the calculation of inhibitory potencies predicted by the QSAR complexation modelin the two digits nanomolar range of concentration.

By molecular dynamics simulations, we verified the stability of new potential analogues and retained the following FLAVs: FLAV 2-15-6-1-1-4-1-1-1 (IC_{50}^{pre}= 50 nM) and FLAV 1-13-7-1-1-4-9-3-1(IC_{50}^{pre}=61 nM) generated by computer-aided molecular design (CAMD) are recommended for synthesis and biological evaluation.

ACKNOWLEDGEMENTS

AUTHOR'SCONTRIBUTION

Bléhoué I: writing original draft, methodology, investigation. **Koné M:** editing, review. **Kéita M:** editing, review. **Fofana I**: formal analysis. **Esmel A:** writing, review, and editing. **Megnassan E:** writing, review, and editing, data curation. All authors read and approved the final manuscript for publication.

DATA AVAILABILITY

Data will be available on request to anyone from the correspondence author.

CONFLICT OFINTERESTS

None to declare.

REFERENCES

1. World Health Organization. Cervical cancer. *https://www.who.int/news-room/fact-sheands/dandail/cervical-cancer*

2. de Sanjosé S, Brotons M, Pavón MA. (2018). The natural history of human papillomavirus infection. Best Practice & Research Clinical Obstetrics Gynaecol 2018; 47: 2-13. *https://doi.org/10.1016/j.bpobgyn.2017.08.015*

3. Tommasino M. The human papillomavirus family and its role in carcinogenesis, Semin. Cancer Biol 26 (2014) 13-21.*https://doi.org/10.1016/j.semcancer.2013.11.002*

4. C. de Martel, Ferlay J, Franceschi S, *et al*. Global burden 6 of cancers attributable to infections in 2008: a review and synthetic analysis, Lancet Oncol. 13 7 (2012) 607-615 (2012) 607-615.

5. C. de Martel, Plummer M, Vignat J, Franceschi S. World-wide burden of cancer attributable to 9 HPV by site, country and HPV type. Int J Cancer 114 (2017) 664- 670.

6. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, *et al*. Structure of the E6/E6AP/p53 complex.

7. Shrestha AD, Neupane D, Vedsted P, Kallestrup P. Cervical cancer prevalence, incidence and mortality in low and middle income countries: A systematic review. Asian Pac J Cancer Prev 2018; 19:319-324.

8. Parisa Shiri Aghbash *et al*. siRNA-E6 sensitizes HPV-16-related cervical cancer through Oxaliplatin: an in vitro study on anti-cancer combination therapy. European J Med Res (2023) 28 :42, 13. *https://doi.org/10.1186/s40001-023-01014-9*

9. M. Celegato, L. Messa, L. Goracci, *et al*. A novel small-molecule inhibitor of the human papillomavirus E6-p53 interaction that reactivates p53 function and blocks cancer cells growth, Cancer Landters.

10. Jonathan J, Cherry, *et al.*, (2013) Structure based identification and characterization of flavonoids that disrupt human Papillomavirus-16 E6 Function 2013; 8:12.

11. Zanier K, Charbonnier S, Sidi AOMHO, *et al*. Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 Oncoproteins. Science 2013, 339, 694.

12. Abotaleb M, Samuel SM, Varghese E, *et al*. Flavonoids in Cancer and Apoptosis. Cancers 2019, 11, 28.

13. Srikanth Kolluru *et al*. Identification of potential binding pocket on viral oncoprotein HPV16 E6: a promising anti-cancer target for small molecule drug discovery. BMC Molecular and Cell Biology.2019. *https://doi.org/10.1186/s12860-019-0214-3*

14. Elise Emeraux. Biological properties of flavonoids: bibliographic study and evaluation of antioxidant activity. Sciences Pharmaceut 2019 *https://hal.univ-lorraine.fr/hal-03297878*

15. Discovery Studio molecular modeling and simulation program, version 2.5, Accelrys, Inc., San Diego, CA, California. USA. 2009

16. Insight II, Version 2005, Molecular Modeling package and Discover Version 2.98. Simulation package, 200x, Accelrys, Inc, San Diego, California, USA. 2005.

17. Kouassi AF, Kone M, Keita M, *et al*. Computer-aided design of orally bioavailable pyrrolidine carboxamide

inhibitors of enoyl-acyl carrier protein reductase of mycobacterium tuberculosis with favorable pharmacokinetic profiles. Int J Mol Sci 16 (12): 29744–71, 2015.

18. Kouassi AF, Kone M, Keita M, *et al*. Computer-aided design of orally bioavailable pyrrolidine carboxamide inhibitors of enoyl-acyl carrier protein reductase of mycobacterium tuberculosis with favorable pharmacokinetic profiles. Int J Mol Sci 2015; 16: 29744-29771. *https://doi.org/10.3390/ijms161226196*

19. Fogolari F, Brigo A, Molinari H. (2002). The poisson-boltzmann equation for biomolecular electrostatics: A tool for structural biology. J Mol Recog 15: 377–92.

20. Li H, Sutter J, Hoffmann R. 2000 Hypo Gen: An automated system for generating 3D predictive pharmacophore models. Güner 2000; 171–189.

21. Chemical Computing Group. MOE (Molecular Operating Environment) User Guide 2015 Edition.

22. Leach AR. Molecular modelling: Principles and applications. Pearson Education, 2016.

23. Qik Prop Version 3.7, Release 14. 2014 NY: X Schrodinger, LLC, New York.

24. Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. Bioorg Med Chem Lett 2000; 10: 1155–8. *https://doi.org/10.1016/s0960-894x(00)00172-4*

25. Jorgensen WL, Duffy EM. 2002 Prediction of drug solubility from structure. Adv Drug Deliv Rev 2002; 54:355–66. *https://doi.org/10.1016/s0169-409x(02)00008-x*

26. Duffy, Jorgensen W. 2000 Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. J Am Chem Soc 122, 78-88. *https://doi.org/10.1021/ja993663t*

27. Maestro-desmond interoperability tools. In Desmond Molecular Dynamics System; Version, 3.6; Maestro-Desmond Interoperability Tools, Schrödinger; D. E. Shaw Research: New York, NY, USA, 2021.

28. *OECD*. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD series on testing and assessment 2014; No. 69, OECD Publishing, Paris. *https://doi.org/10.1787/9789264085442-en*

29. Lipinski CA, Lombardo F, Dominy BW, Feene, PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Del Rev 2001, 46, 3–26.

30. Veber, DF, Johnson, SR, Cheng, HY, Smith, BR, Ward, KW, and Kopple, KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 2002 Jun 6;45(12):2615-23. *https:// doi.org/10.1021/jm020017n*

31. Gomes D, Yaduvanshi S, Silvestre S, *et al*. Taxifolin and Lucidin as Potential E6 Protein Inhibitors: p53 Function Re-Establishment and Apoptosis Induction in Cervical Cancer Cells. Cancers 2022; 14: 2834. *https:// doi.org/10.3390/cancers14122834*

32. Lemkul JA. From proteins to perturbed Hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package. Biophys J 2019; 117(3), 354-361.