Pharmacophore Modelling and 4D QSAR Analysis of Some Indole Glyoxamide Derivatives as HIV-1 Binding Inhibitors

Nazmiye Sabancı*

*Siirt University, Faculty of Arts and Science, Department of Chemistry, Siirt, Turkey.
nsabanci@siirt.edu.tr

Summary: The main aim of this study is to uncover the main pharmacophoric features of a series of indole glyoxamide derivatives which are known as HIV-1 attachment inhibitors and to estimate the biological activity to develop a 4D-QSAR model by using the EC-GA method.

Conformational analysis and quantum mechanical calculations were accomplished by using the Hartree Fock method with the 3-21G basis set. Based on the data produced from the quantum chemical calculations, the electron conformational matrices of congruity (ECMC)s, as the 3D- arrangement of electronic and geometric properties, were generated by the EMRE program. An individual ECMC was formed for each conformer of each indole glyoxamide derivative in the data set. Conformational flexibility was considered for each compound. Totally 1510 ECMCs were produced by EMRE software to be used in the comparison process. By analogizing the ECMCs in a predetermined tolerance value, the subset of common features matching all active compounds but not matching the low activity compounds was determined. The final ECSA was obtained as a set of nine atoms including predominantly hydrogen bond donors, hydrogen bond acceptors and lipophilic units. Key elements are mainly placed in indole nitrogen, carbonyl groups and piperazine ring. In the bioactivity prediction and variable selection, the genetic algorithm and non-linear least square techniques were employed. The obtained models were internally and externally validated by the leave-one-out cross-validation method. The resulting 4D QSAR EC-GA models were compared with the other methods and the best model with the high prediction ability was defined according to $R^2_{training}=$, $R^2_{test} =0.8$ cross-validatedated $q^2 =0.860$, $q^2_{ext1} = 0.850$ and $q^2_{ext1} = 0.850$ values.

Attained EC-GA model provides insight into the vital interaction between indole glyoxamide derivatives and the target protein. EC-GA models can be utilized as an effective and confidential tool in the design of more potent indole glyoxamide derivatives.

Introduction

The human immunodeficiency virus (HIV) is an initiative factor of acquired immunodeficiency syndrome (AIDS) which is one of the most life-threatening global health problems. When HIV enters the human body, HIV targets a particular type of white blood cells named CD4+ T-cells in the immune system [1] and progressively weakens the defense system by eventuating in CD4 cell depletion [2]. This situation enables to AIDS and its sensibility to other opportunistic infections.

The World Health Organization reported that there are 680000 HIV-related deaths in 2020 and globally 37.7 million people are living with HIV including all ages and still being an international health problem [3]. On the other hand, the transmission of HIV became a controllable long-lasting health problem by the growing availability and accessibility to the current HIV repression procedures and therapies.

Nowadays, in an effort to manage the HIV infection, the effective treatment of HIV is identified as antiretroviral therapy (ART) (as well named highly active antiretroviral therapy-HAART) and requires the use of a combination of various antiretroviral medicines.

Although it is not possible to completely cure HIV with HAART, it is possible to provide longer and healthier lives to HIV-infected people by overpowering HIV-1 replication to undeterminable levels, increasing the CD4 count and slowing down the progression of the disease [2, 4].

Drugs approved by FDA (Food and Drug Administration) to be used in the antiretroviral treatment are categorized in different classes according to their action mechanism: nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), integrase inhibitors, fusion inhibitors, chemokine receptor 5 (CCR5) inhibitors, attachment inhibitors, post-attachment inhibitors and pharmacokinetic enhancers [4, 5-7].

The requirement for new HIV-1 inhibitors acting by different mechanisms revealed new compounds including indole glyoxamide scaffold as small molecules [8]. One and the most promising of these indole based compounds is BMS-378806 (4-Benzoyl-1-[(4-methoxy-1H-pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2-(R)-methylpiperazine) which is first introduced by Wang *et*

---

*To whom all correspondence should be addressed.

al. [9-10]. BMS-378806 is the first model of a series of small-molecule HIV-1 inhibitors that facultatively bind to the viral protein and successfully prevent the attachment of the virus to CD4 receptors [9, 11, 12]. BMS-378806 was further systematically optimized by structural modifications to give improved derivatives. BMS-378806 and its current family members (like BMS-663068, also known as fostemsavir, which is currently in phase III clinical development) are considered as promising HIV-1 inhibitors [13-16]. Until now, by modifying the piperazine ring, indole/azaindole heteroring or benzamide moiety, many researchers made an effort to develop and synthesize the novel small molecule inhibitors of HIV-1 based on the indole glyoxamide derivatives developed by Bristol-Myers Squib [17-22]. On the other hand, molecular dynamic simulation study [23], docking and 3D-QSAR studies [24-25] have also been performed to predict the binding mode of BMS-related ligands for HIV-1 gp120 entry inhibitors. Several compounds reported by Meanwell and his coworkers [8, 13, 26] have been used by other researchers for the aim of docking [27] or 3D-QSAR analysis [18]. Data of indole glyoxamide derivatives given by ref [26] and also used in the current study, have been analyzed by Nirouei *et al*. to develop linear and nonlinear QSAR models with multiple linear regression, genetic algorithms and artificial neural networks [28] and by Lu *et al*. for ligand-based 3D QSAR study by CoMFA and CoMSIA methods [29]. However, there is no 4D QSAR study of related compounds in the literature.

The ultimate goal of computer-aided drug design or rational drug discovery is to establish a possible linear/nonlinear mathematical relationship between molecular features determined from the chemical structure of the drug and the biological activity [30]. During the drug design process, specifying the features dominating the biological activity and identifying the pharmacophoric components of the drug interacting with the protein is pretty important.

Three-dimensional quantitative structure-activity relationship (3D QSAR) methods correlate the biological activity of a compound with calculated descriptors produced from the spatial arrangement of the chemical structure, based on a single conformer of a ligand. 3D QSAR techniques, like CoMFA and CoMSIA, have been commonly applied to premediate the biologically active new compounds [31-34]. The detection of the bioactive conformers is a significant step affecting the reliability of 3D QSAR methods. It may be easy to determine bioactive conformation for small or rigid molecules in which there are no or only a couple of rotatable bonds. However, it is a handicap for bulky and non-rigid molecules including the notable number of single bonds. In some cases, the representation of a

molecule by a single conformation, which is chosen incorrectly, may result in less reliable models with weakened interpretability. Because parameters used in the model development are sensitive to the conformational changes of the compound. Efforts to overcome difficulties in studying with large data sets and drawbacks in considering only one conformer of a compound in 3D QSAR led up to the emergence of 4D QSAR [35, 36]. Different from the 3D QSAR methods, conformational flexibility of a ligand is taken into consideration in 4 D the QSAR methodology instead of a single conformation of a ligand. Incorporating the conformational space of a ligand as an ensemble in the QSAR model enhances the predictive power and interpretability of the model.

The current study aims to uncover the pharmacophoric features and predict the inhibitory activity of HIV-1 attachment inhibitors in the class of indole glyoxamide derivatives [26] using the hybrid 4D QSAR electron conformational-genetic algorithm (EC-GA) methodology. This approach involved in the scope of 4D QSAR allows the utilization of the conformational ensemble profile for each compound in the data set for both pharmacophore detection and bioactivity prediction [37]. In the previous studies performed by our research group, the EC-GA method was successfully applied to a wide variety of data sets and a comprehensive explanation of the methodology is presented in the literature [37-45].

In the present paper, by using the EC-GA method, we focused on the identification of the pharmacophoric elements of indole glyoxamide based HIV-1 inhibitors reported by Meanwell *et al* [26]. The details of the EC-GA method were described earlier in detail [37-44]. To control the performance of the EC-GA method for pharmacophore identification and the bioactivity estimation, we have compared our model with the results in the literature for the same data set.

Experimental

The generation of three-dimensional pharmacophore models is a highly helpful tool in ligand-based methods to comprehend the ligand-receptor interaction. According to the IUPAC, a pharmacophore is defined as "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" and defined by pharmacophoric features such as hydrogen bond donors/acceptors, hydrophobic or electronic interaction positions [46-47]. Developed pharmacophore models can be governed to improve the ligand-receptor interaction of existing drug molecules or

design novel and more potent bioactive drug candidates [48-49].

Due to the quite flexible ligand molecules, the identification of the pharmacophoric features is an intricate process. Ligands owning rotatable atomic groups around single bonds convert between different conformations. Each of these conformations has the probability to interact with the relevant receptor [50]. Therefore, a single conformer would not represent the real molecule.

The pharmacophore of a series of indole-glyoxamide derivatives which is responsible for the HIV-1 binding inhibition was identified by the EC-GA method in the scope of ligand-based pharmacophore modeling by taking into account the energetically favorable conformations for each molecule. The formation of the ECSA from the multi-conformation of compounds can be considered as a three-step action. The first is the demarcation of the conformer space for each compound by the conformational analysis performed as an independent step. The conformational space of each compound was characterized by the ensemble of conformations. The second is the generation of ECMCs by the EMRE program. The third stage is the comparison of ECMCs to extract the essential features (ECSA) for the activity. Throughout the pharmacophore identification process, two main stages need special attention are the treatment of conformational elasticity of each compound and characterization of pharmacophoric features. The former stage handles multiple conformers for each compound by conformational analysis since the global minimum energy conformer is not always equal to the biologically active conformer, the second aims to find a common feature in all compounds in the data set [51].

Comprehensive details of the used method are given in previous studies [37-45]. The chemical structures of the indole-glyoxamide derivatives and their experimental data were taken from the literature [26] and listed in Table-1. The EC_{50} values in units of nM that are given in the literature were converted to M units and then to pEC_{50} (-log EC_{50}) values.

Table-1: The skeletal structure and HIV pseudotype virus inhibitory activity (in nM) of indole glyoxamide derivatives

Compound	R1	R2	R3	R4	R5	EC_{50}	pEC_{50}
1	H	H	H	H	H	152.97	6.82
2	F	H	H	H	H	2.59	8.59
3	Cl	H	H	H	H	4.3	8.37
4	Br	H	H	H	H	4.5	8.35
5	NO_2	H	H	H	H	149.8	6.82
6	$OCOCH_3$	H	H	H	H	24857.5	4.60
7	OH	H	H	H	H	20137.4	4.70
8	OCH_3	H	H	H	H	0.52	9.28
9	OCH_2CH_3	H	H	H	H	0.45	9.35
10	H	F	H	H	H	838.3	6.08
11	H	Cl	H	H	H	395	6.40
12	H	Br	H	H	H	1090.3	5.96
13	H	CH_3	H	H	H	1548.6	5.81
14	H	OCH_3	H	H	H	21100	4.68
15	H	NO_2	H	H	H	3800	5.42
16	H	CN	H	H	H	120523.1	3.92
17	H	$OCOCH_3$	H	H	H	952.2	6.02
18	H	H	F	H	H	21.1	7.68
19	H	H	Cl	H	H	208	6.68
20	H	H	OCH_3	H	H	328.8	6.48
21	H	H	H	F	H	7.3	8.14
22	H	H	H	Cl	H	4.4	8.36
23	H	H	H	Br	H	17.4	7.76
24	H	H	H	CH_3	H	89.5	7.05
25	H	H	H	CH_3CH_2	H	24.2	7.62
26	H	H	H	OCH_3	H	6.6	8.18

27	H	H	H	OCH₂CH₃	H	0.5	9.30
28	H	H	H	OC₄H₉	H	0.14	9.85
29	H	H	H	CN	H	4.9	8.31
30	F	Br	F	H	H	23.1	7.64
31	H	Cl	Cl	H	H	500	6.30
32	H	OCH₃	OCH₃	H	H	46451.2	4.33
33	H	F	H	Br	H	73.1	7.14
34	F	H	H	F	H	0.35	9.46
35	OCH₃	H	H	OCH₃	H	0.23	9.64
36	OCH₃	H	H	Cl	H	0.07	10.15
37	OCH₃	H	H	Br	H	0.13	9.89
38	OCF₃	H	H	Br	H	42.8	7.37
39	F	H	H	Br	H	0.13	9.89
40	F	H	H	CH₃	H	0.56	9.25
41	F	H	H	OCH₃	H	0.06	10.22
42	F	H	H	OCH₂CF₃	H	0.71	9.15
43	F	H	H	CN	H	1.9	8.72
44	Br	H	H	F	H	8.2	8.09
45	F	H	F	H	H	0.42	9.38
46	F	F	F	F	H	1.7	8.77
47	H	H	H	H	CH₃	265	6.58
48	H	H	H	H	CH₃ CH₂	1450	5.84
49	H	H	H	H	CH₃(CH₂)₃	2650	5.58
50	H	H	H	H	CH₂=CH-CH₂	6760	5.17
51	H	H	H	H	PhCH₂	13750	4.86
52	OCH₃	H	H	CN	H	0.06	10.22

Conformational analysis and quantum chemical calculations of 52 indole glyoxamide derivatives were carried out at HF/3-21G level of theory in an aqueous medium by using the Spartan 10 program [52]. Structurally bulky and non-rigid molecules exist in diverse conformations depending on the number of rotatable single bonds. Existing of a compound in different conformers is especially important when the compound has pharmacological potency that is fulfilled by only a certain spatial position of atoms. Among all conformations, high-energy ones are less populated and less likely to be responsible for pharmacological activity [53]. For this reason, following the conformational analysis, conformations with a large population and lower energy than 1.5 kcal/mol were kept according to the Boltzmann distribution in order to take into account the probable biologically active conformation which possibly binds to protein. Remained conformations were aligned based on the minimum energy conformation providing that superposition of corresponding atoms as seen in Fig. 1. The superposition of conformers was done according to the indole heterocycle, piperazine ring and benzamide moiety.

In the EC-GA method, conformations with large population for each compound are defined by the electron conformational matrices of congruity (ECMCs) which are composed of electronic and geometrical parameters. Elements of the ECMCs are taken by the EMRE software [37-45] from the output files of quantum chemical calculations. In the ECMCs,
.

diagonal positions of the matrix were occupied by the Mulliken charges of the interested atoms. Non-diagonal elements of the matrix are filled by bond orders for bonded atom pairs or interatomic distances for non-bonding atom pairs [54-55]. Fig. 2 shows the visual depiction of the ECMC of the lowest energy conformer of the most potent compound 52 in the data set. By this way, an individual ECMC was formed for each conformer of each indole glyoxamide derivative in the data set. Totally 1510 ECMCs were produced by EMRE software to be used in the comparison process.
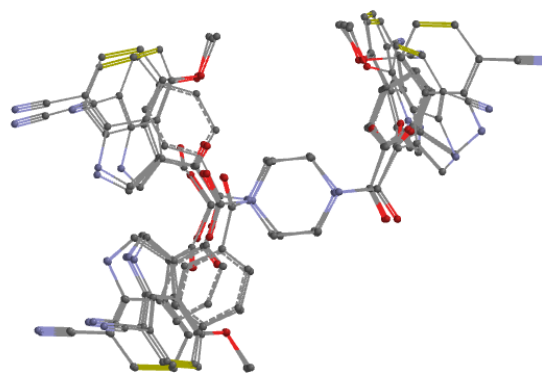

Fig. 1: Superposition of 10 conformations of the compound 52 in the indole glyoxamide derivatives.
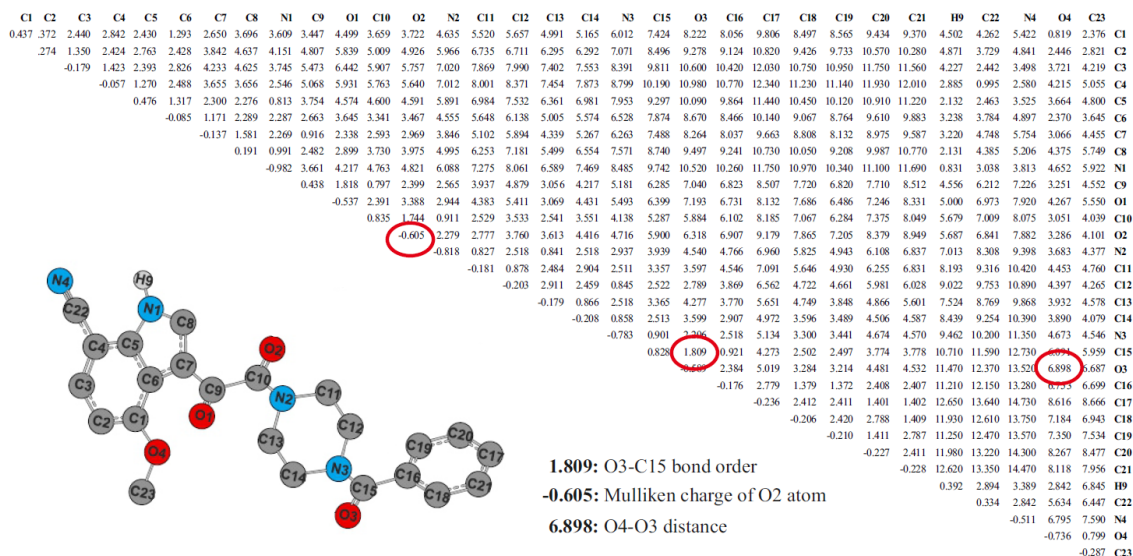
	C1	C2	C3	C4	C5	C6	C7	C8	N1	C9	O1	C10	O2	N2	C11	C12	C13	C14	N3	C15	O3	C16	C17	C18	C19	C20	C21	H9	C22	N4	O4	C23	
	0.437	.372	2.440	2.842	2.430	1.293	2.650	3.696	3.609	3.447	4.499	3.659	3.722	4.635	5.520	5.657	4.991	5.165	6.012	7.424	8.222	8.056	9.806	8.497	8.565	9.434	9.370	4.502	4.262	5.422	0.819	2.376	C1
		.274	1.350	2.424	2.763	2.428	3.842	4.637	4.151	4.807	5.839	5.009	4.926	5.966	6.735	6.711	6.295	6.292	7.071	8.496	9.278	9.124	10.820	9.426	9.733	10.570	10.280	4.871	3.729	4.841	2.446	2.821	C2
			-0.179	1.423	2.393	2.826	4.233	4.625	3.745	5.473	6.442	5.907	5.757	7.020	7.869	7.990	7.402	7.553	8.391	9.811	10.600	10.420	12.030	10.750	10.950	11.750	11.560	4.227	2.442	3.498	3.721	4.219	C3
				-0.057	1.270	2.488	3.655	3.656	2.546	5.068	5.931	5.763	5.640	7.012	8.001	8.371	7.454	7.873	8.799	10.190	10.980	10.770	12.340	11.230	11.140	11.930	12.010	2.885	0.995	2.580	4.215	5.055	C4
					0.476	1.317	2.300	2.276	0.813	3.754	4.574	4.600	4.591	5.891	6.984	7.532	6.361	6.981	7.953	9.297	10.090	9.864	11.440	10.450	10.120	10.910	11.220	2.132	2.463	3.525	3.664	4.800	C5
						-0.085	1.171	2.289	2.287	2.663	3.645	3.341	3.467	4.555	5.648	6.138	5.005	5.574	6.528	7.874	8.670	8.466	10.140	9.067	8.764	9.610	9.883	3.238	3.784	4.897	2.370	3.645	C6
							-0.137	1.581	2.269	0.916	2.338	2.593	2.969	3.846	5.102	5.894	4.339	5.267	6.263	7.488	8.264	8.037	9.663	8.808	8.132	8.975	9.587	3.220	4.748	5.754	3.066	4.455	C7
								0.191	0.991	2.482	2.899	3.730	3.975	4.995	6.253	7.181	5.499	6.554	7.571	8.740	9.241	9.497	10.730	10.050	9.208	9.987	10.770	2.131	4.385	5.206	4.375	5.749	C8
									-0.982	3.661	4.217	4.763	4.821	6.088	7.275	8.061	6.589	7.469	8.485	9.742	10.520	10.260	11.750	10.970	10.340	11.100	11.690	0.831	3.038	3.813	4.652	5.922	N1
										0.438	1.818	0.797	2.399	2.565	3.937	4.879	3.056	4.217	5.181	6.285	7.040	6.823	8.507	7.720	6.820	7.710	8.512	4.556	6.212	7.226	3.251	4.552	C9
											-0.537	2.391	3.388	2.944	4.383	5.411	3.069	4.431	5.493	6.399	7.193	6.731	8.132	7.686	6.486	7.246	8.331	5.000	6.973	7.920	4.267	5.550	O1
												0.835	1.744	0.911	2.529	3.533	2.541	3.551	4.138	5.287	5.884	6.102	8.185	7.067	6.284	7.375	8.049	5.679	7.009	8.075	3.051	4.039	C10
													-0.605	2.279	2.777	3.760	3.613	4.416	4.716	5.900	6.318	6.907	7.205	8.379	8.949			5.687	6.841	7.882	3.286	4.101	O2
														-0.818	0.827	2.518	0.841	2.518	2.937	3.939	4.540	4.766	6.960	5.825	4.943	6.108	6.837	7.013	8.308	9.398	3.683	4.377	N2
															-0.181	0.878	2.484	2.904	2.511	3.357	3.597	4.546	7.091	5.646	4.930	6.255	6.831	8.193	9.316	10.420	4.453	4.760	C11
																-0.203	2.911	2.459	0.845	2.522	2.789	3.869	6.562	4.722	4.661	5.981	6.028	9.022	9.753	10.890	4.397	4.265	C12
																	-0.179	0.866	2.518	3.365	4.277	3.770	5.651	4.749	3.848	4.866	5.601	7.524	8.769	9.868	3.932	4.578	C13
																		-0.208	0.858	2.513	3.599	2.907	4.972	3.596	3.489	4.506	4.587	8.439	9.254	10.390	3.890	4.079	C14
																			-0.783	0.901	1.206	2.518	5.134	3.300	3.441	4.674	4.570	9.462	10.200	11.350	4.673	4.546	N3
																				0.828	**1.809**	0.921	4.273	2.502	2.497	3.774	3.778	10.710	11.590	12.730		5.959	C15
																						2.384	5.019	3.284	3.214	4.481	4.532	11.470	12.370	13.520	**6.898**	6.687	O3
																						-0.176	2.779	1.379	1.372	2.408	2.407	11.210	12.150	13.280		6.699	C16
																							-0.236	2.412	2.411	1.401	1.402	12.650	13.640	14.730	8.616	8.666	C17
																								-0.206	2.420	2.788	1.409	11.930	12.610	13.750	7.184	6.943	C18
																									-0.210	1.411	2.787	11.250	12.470	13.570	7.350	7.534	C19
																										-0.227	2.411	11.980	13.220	14.300	8.267	8.477	C20
																											-0.228	12.620	13.350	14.470	8.118	7.956	C21
																												0.392	2.894	3.389	2.842	6.845	H9
																													0.334	2.842	5.634	6.447	C22
																														-0.511	6.795	7.590	N4
																															-0.736	0.799	O4
																																-0.287	C23

1.809: O3-C15 bond order

-0.605: Mulliken charge of O2 atom

6.898: O4-O3 distance

Fig. 2: The visual depiction of the lowest energy conformer of the most active compound **5.**

In order to reveal pharmacophoric elements of indole glyoxamide derivatives, compounds in the data set are classified as high activity and low activity compounds. By defining the threshold value (pEC_{50} = 7.76), compounds whose pEC_{50} values are greater than or equal to 7.76 are classified as compounds with lower activity. Other 26 compounds are considered as high activity compounds. The conformation with the minimum energy of the most potent compound 52 was considered as the reference compound and then submitted to the comparison procedure which produces the electron conformational submatrix of the activity (ECSA) in other words the pharmacophore. The diagonal and non-diagonal elements of the ECMC of the reference compound were analogized with that of the ECMCs of the all other conformations within the predefined tolerance values.

Analogizing the ECMCs to determine the subset of common features matching all active compounds but not matching the low activity compounds [37-40] produced a range of ECSAs. Among all ECSAs, the one which is most likely to be the most favorable ECSA was distinguished from others based on the two evaluation measures, P_α and α_α. Both measures are related to the possibility of the existence of the pharmacophore. But P_α indicates the evidence of pharmacophore presence in compounds with only high activity. α_α indicates the evidence of pharmacophore presence in compounds with both high and low activity [56]. These measures are given by the following expressions [37, 57]:

$$P_\alpha = (n_1 + 1)/(n_1 + n_3 + 2) \quad (1)$$

$$\alpha_a = (n_1 \times n_4 - n_2 \times n_3)/(m_1 \times m_2 \times m_3 \times m_4)^{\frac{1}{2}} \quad (2)$$

In the expressions given above, n_1 and n_2 refer to the pharmacophoric feature (ECSA)-inclusive and noninclusive high activity compounds, respectively. n_3 and n_4 are equivalent of those for the low activity compounds. m_1 and m_2 represent the number of the high and low activity compounds, respectively. $m_3 = n_1 + n_3$; $m_4 = n_2 + n_4$ [54].

In the context of 4D QSAR studies, the biological activities of the related data set was calculated by the following equation [37, 53]:

$$A_n = A_l \frac{\sum_{i=1}^{m_l} e^{-E_{li}/kT} \sum_{i=1}^{m_n} \delta_{ni}[Pha] e^{-S_{ni}} e^{-E_{ni}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ni}/kT} \sum_{i=1}^{m_l} \delta_{li}[Pha] e^{-S_{li}} e^{-E_{li}/kT}} \quad (3)$$

In this equation, δ is Kronecker delta function with two variables. This function is equal to 1 if the compound has Pha, otherwise, it is 0. A_n and A_l, refer to the activities of the nth compound and the reference compound, respectively. E_{li} is the relative energy of the ith conformation of the reference compound (in kcal mol^{-1}) while E_{ni} is the relative energy of the ith conformation of the nth compound (in kcal mol^{-1}). k (kcal mol^{-1} K^{-1}) refers to the gas constant [37]. Also, the effect of the out-of-Pha groups, named Auxiliary groups (AG) and the anti-Pha shielding (APS), were explored by the function S which is exponentially related to the activity [53] and given by equation 4:

$$S_{ni} = \sum_{j=1}^{N} \kappa_j a_{ni}^{(j)} \qquad (4)$$

Here, N is the number of selected descriptors. The constant κ_j is the relative weight of the relevant descriptor to the activity. $a_{ni}^{(j)}$ corresponds to the jth kind of property of the ith conformation of the nth compound [37]. In this way, it is possible to explore other factors influencing the activity apart from the Pha. Because compounds with identical Pha produce unlike activities. The conformational ensemble for each compound was taken into account in the Equation 3 in the extension of the 4D QSAR concept.

Using Equation 3 to estimate the quantitative activities of compounds, κ_j value of each parameter was optimized based on the non-linear least square function in the Matlab statistics toolbox [58-59]. This function tries to fit the data with the minimum difference between the calculated and the experimental activity values.

In an attempt to reveal the properties which, differentiate activities of compounds with identical Pha, first large size of parameter pool for each conformer of each compound was formed by the EMRE software [37-45] and then the most important parameters for the biological activity were determined by the GA optimization [60], probabilistic optimization technique, by discarding the unnecessary parameters. The use of GA in the calculation of the activity (Equation 3) yielded multiple subsets of parameters based on random selection (GA parameters used in the optimization; generation number: 150; population size: 100; the number of iteration: 150; crossover fraction: 80%; mutation rate: 1.5%). All potential subsets of parameters were assessed by a fitness function to decide the best one. Among the potential subsets of descriptors, the optimum subset which gives the best model was evaluated according to the following fitness function named the predictive residual sum of squares (PRESS) (Equation 5), and the fitness value of the potential subsets were determined by the leave-one-out cross-validation (LOO-CV) technique:

$$PRESS_N = \sum_{n=1}^{N} \left| A_n^{\exp} - A_n^{pred} \right|^2 \quad (5)$$

Here, A_n^{exp} and A_n^{pred} refer to the experimental and the predicted activities of the nth molecule, respectively, in the LOO-CV. N refers to the number of training molecules.

In EC-GA method, using the LOO-CV method, the effectiveness of the obtained models based on the potential subsets of descriptors were affirmed internally with training compounds and externally with test compounds. In the LOO-CV method, each compound was left out once. The biological activity of this compound was estimated by the other compounds in the training set and this procedure was repeated for each compound in the training set. Although internal validation gives information about model performance, it is insufficient to make exact estimations for unknown compounds not included in the development of the QSAR model. Therefore, external validation is indispensable to acquire the model with high predictive power [61]. Compounds apart from the training set, not included in the model development were employed to accomplish the external validation. In this study, to obtain the best QSAR model with high predictive power and statistically robust, not cross-validated correlation coefficient (q^2) but also external validation measures, q^2_{ext1} and q^2_{ext2}, proposed by Schuurmann *et al.* [61-62] were used. The final EC-GA model was revealed as a function of these parameters according to the statistical parameters R^2, q^2, q^2_{ext1} and q^2_{ext2}. Thus, the optimum EC-GA model was identified to predict the activity values of a series of indole-glyoxamide derivatives.

Results and Discussion

The pharmacophoric elements of some indole-glyoxamide derivatives as HIV-1 attachment inhibitors were revealed by the EC-GA method based on conformational ensemble for each compound in the data set. The structural scaffold of the indole-glyoxamide derivatives and substituents were given in Table 1. During the pharmacophore exploration to incorporate the conformational flexibleness, a set of conformations for each ligand is employed. Otherwise, it is possible to fail in identifying flexibility limits of ligands. Depending on the rigidity of the molecules, the entire conformational space of the related compound series was represented by totally 1510 conformers. Each conformation was represented by the individual ECMC formed by the EMRE program.

After the formation of ECMCs, a specified cut-off value ($pEC_{50} = 7.76$) separated compounds as high and low activity compounds. So, 26 compounds possessing activity values below 7.76 were marked as low activity compounds whereas the remaining 26 were high activity. Then, the reference compound, the global energy minimum conformer of the compound 52 with the minimum EC_{50} concentration (as the most

efficacious) was designated. As mentioned in the Material and Method section and related literature [37-45], by superposing the set of high and low activity compounds, and comparing them with the reference ECMC until the limit of tolerance values, common chemical features for all ligand molecules, defined as ECSA, are extracted from the entire ECMCs. The attained tolerance values allow the fittest partition of the active and low active compounds. The possible pharmacophore candidates are scored by the P_α and α_α, and the top-scoring ECSA is assigned as the best pharmacophore. Accordingly, a set of nine atoms comprised of the N1, C9, O1, C10, O2, N2, N3, C15, and O3 with the highest score of P_α and α_α was stated as the optimal ECSA for the indole-glyoxamide derivatives interested and given in Table 2. Extracted ECSA with the lowest tolerances was found out to be necessary features for the HIV-1 attachment. In Table 2, the submatrix of the pharmacophoric elements of indole-glyoxamide derivatives and the tolerance values of the high and low activity compounds are shown. As an illustration, pharmacophore atoms are displayed with white color on the reference compound in this table. The first (a) of the four submatrices in Table 2 is the ECSA components of the global energy minimum conformer of the reference compound 52. The submatrix (b) refers to the tolerance values of 26 high-activity compounds. The submatrix (c) is the tolerance matrix of the 26 low-activity compounds. Then tolerance values were set free to explore molecular elasticity limits and get the extreme values of tolerances of all conformations of all compounds. The tolerance matrix of the totally 1510 conformations of 52 compounds is shown by the submatrix (d). The general trend in the table is that high activity compounds possess lower tolerances compared to low activity compounds [63]. The atomic charge of the O1 atom has a tolerance value of ±0.0031 in high activity compounds and ±0.0072 in the low activity compounds. We can see a similar case for the charge of the N2 atom. Tolerance values of the N2 charge are ±0.035 and ±0.040 for high and low activity compounds, respectively. Analysis of the interatomic distance between the N2 and C10 atoms also proves this situation. The distance tolerance of N2-C10 atoms is ±0.020 for high-activity compounds and ±0.024 for low-activity compounds.

As it is seen in Table 2, the N1, O1, O2, N2, N3, and O3 atoms would inclined to match up with positive regions of the protein in virtue of their negative charges. The positively charged C9, C10 and C15 atoms would tend to attract to the negatively charged regions of the protein. It is possible that interaction between ligand molecule and the active site of the protein may occur between the negative regions of the ligand molecule and the positive parts of the receptor due to the majority of the negatively charged pharmacophore atoms. Besides, pharmacophoric key elements predominantly comprise the hydrogen bond acceptors and donors which play a fundamental role because of their robust effect in the interaction of ligand and receptor. The N1 atom is one of the key pharmacophoric features for indole glyoxamide derivatives and was found to be responsible for the HIV-1 inhibition by the EC-GA method. Inside of the ECSA atoms, the negatively charged N1 atom in the rigid indole heterocyclic unit act as hydrogen bond donor for the binding of the ligand molecule to the active site. According to the reported data in the literature by Meanwell *et al*. [26], the attachment of small substituents at N1-position of indole ring weakens the inhibitory potency depending on the size of the alkyl group. Bulkier alkyl moiety and rigid phenyl ring diminish increasingly the inhibitory activity. This situation can be explained by the steric hindrance of the attached substituent to the N1-position which appears to influence the spatial orientation of the ligand and block the hydrogen bonding or other non-covalent interactions. The O1 and O2 atoms in the glyoxal unit, the O3 atom in the benzylic carbonyl group, the N2 and N3 atoms in the piperazine ring are identified as hydrogen bond acceptor to be responsible for the other hydrophilic interactions.

The lipophilic ligand-protein interactions are likely to be governed by the C9, C10 and C15 atoms. We can conclude that most of the interactions with the target protein for the indole glyoxamide derivatives are type of non-covalent. Both lipophilic and hydrogen bonding interactions are crucial for the specific binding in biological recognition by the target protein.

Also, the N2 and N3 atoms in the piperazine ring are significant components for the HIV-1 binding inhibition, by functioning as a site to emplace the glyoxal and heterocyclic indole units to the binding pocket in a complementary way [8].

Table-2: (a) ECSA (Pha) of reference compound 52 (b) Tolerance matrix for 26 high activity compounds, (c) Tolerance matrix for 26 low activity compounds, (d) Tolerance matrix for 1510 conformations of 52 compounds.
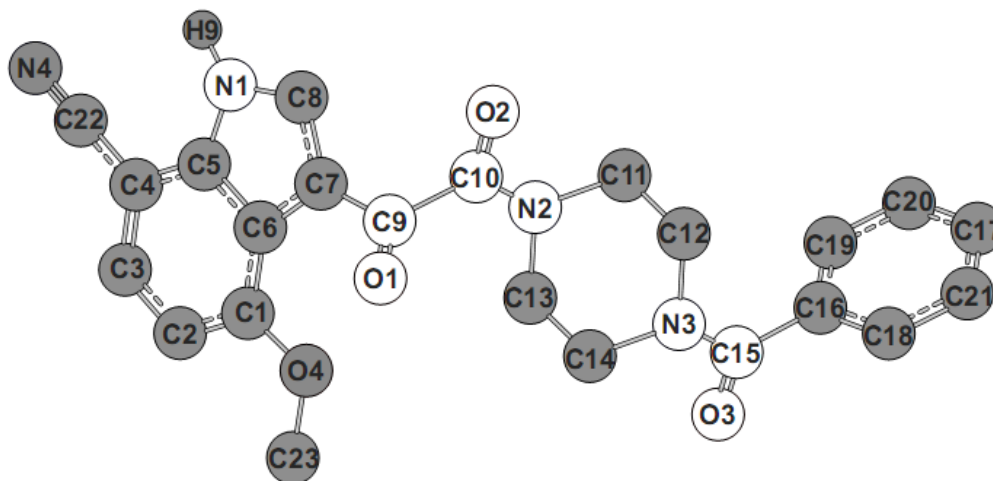
(a) ESCA (Pha) of the reference compound

N1	C9	O1	C10	O2	N2	N3	C15	O3	
-0.982	3.662	4.217	4.763	4.821	6.088	8.485	9.742	10.515	N1
	0.438	1.818	0.797	2.399	2.565	5.180	6.285	7.040	C9
		-0.537	2.391	3.388	2.944	5.493	6.399	7.193	O1
			0.835	1.744	0.911	4.138	5.287	5.884	C10
				-0.605	2.279	4.716	5.900	6.318	O2
					-0.818	2.937	3.939	4.540	N2
						-0.783	0.901	2.296	N3
							0.828	1.809	C15
								-0.589	O3

(b) Tolerance values for 26 compounds with high activity compounds

N1	C9	O1	C10	O2	N2	N3	C15	O3	
±0.034	±0.026	±0.413	±0.442	±0.514	±0.153	±1.335	±1.428	±1.917	N1
	±0.057	±0.070	±0.023	±0.123	±0.026	±0.399	±0.348	±0.641	C9
		±0.031	±0.035	±0.233	±0.352	±0.713	±0.676	±0.790	O1
			±0.047	±0.028	±0.020	±0.190	±0.442	±0.525	C10
				±0.013	±0.017	±0.325	±0.519	±0.913	O2
					±0.035	±0.065	±0.397	±0.495	N2
						±0.017	±0.012	±0.007	N3
							±0.008	±0.010	C15
								±0.006	O3

(c) Tolerance values for 26 compounds with low activity compounds

N1	C9	O1	C10	O2	N2	N3	C15	O3	
±0.035	±0.024	±0.428	±0.483	±0.705	±1.181	±2.954	±3.667	±4.304	N1
	±0.070	±0.139	±0.035	±0.015	±0.022	±0.423	±0.371	±0.602	C9
		±0.072	±0.022	±0.280	±0.358	±0.583	±0.724	±0.844	O1
			±0.048	±0.028	±0.024	±0.209	±0.442	±0.524	C10
				±0.019	±0.018	±0.291	±0.518	±0.914	O2
					±0.040	±0.064	±0.394	±0.493	N2
						±0.016	±0.011	±0.004	N3
							±0.008	±0.011	C15
								±0.010	O3

(d) Tolerance values for 1510 conformations of 52 compounds

N1	C9	O1	C10	O2	N2	N3	C15	O3	
±1.264	±0.053	±0.436	±0.483	±0.811	±1.304	±3.561	±4.800	±6.268	N1
	±0.077	±0.144	±0.060	±0.017	±0.028	±0.546	±0.653	±1.366	C9
		±0.232	±0.041	±0.332	±0.398	±0.713	±0.857	±1.050	O1
			±0.624	±0.084	±0.150	±0.345	±0.443	±0.615	C10
				±0.261	±0.025	±0.343	±0.524	±0.918	O2
					±0.760	±0.093	±0.404	±0.499	N2
						±0.721	±0.144	±0.018	N3
							±0.522	±0.116	C15
								±0.243	O3

Extracted pharmacophoric features for the gp120-CD4 binding inhibition of indole glyoxamide derivatives by the EC-GA method are compatible with similar studies available in the literature [18, 25, 26]. The final pharmacophore model obtained by the EC-GA method would play an essential role in the optimization and discovery of novel lead HIV-1 binding inhibitors. It is possible to find out more potent new lead indole glyoxamide derivatives with different pharmacological features. EC-GA method is capable of handling ligand flexibility and bioactive conformation handicap.

As the second part of the 4D QSAR EC-GA modelling to produce the estimated inhibitory activities, the data of indole glyoxamide derivatives was randomly partitioned into the training set to model construction and the test set to model validation. In this step, by including the conformational ensemble profiles of each compound, different models were constructed based on the different sizes of the training and test sets to compare with the literature. Variable selection was performed with the GA in Matlab to keep the less significant variables out and to hold the most effective ones representing the best model. The relative weights of the parameters (Equation 4), κ_j, were adjusted using Equation 5. Then calculated κ_j values were used in the prediction of inhibitory activities of the related data set. By multiple GA runs, various variable subsets were obtained and scored regarding the fitness function. The one with the best score was accepted as the optimum subset. However, the optimal number of parameters in the subset is uncertain. We tried to obtain the best and the most predictive model with the minimum number of descriptors. First, training and test sets were randomly generated. By using the generated training and test sets as fixed, we examined the parameter numbers to determine the optimal number of parameters.

In this study, we generated different models to define the most reliable and powerful 4D-QSAR model for indole glyoxamide derivatives by the EC-GA method and to measure the model performance compared to other QSAR techniques in the literature. Although a QSAR model with high enough internal prediction capacity, it does not make certain the same external prediction power. In order to prove the model performance of the EC-GA method and develop a better model with high predictivity, we developed three models given by Model 1- 3. In Model 1, we adopted the model reported by Lu *et al.*[29], based on 41 training and 11 test compounds. To be able to make a comparison, the same compounds given by Lu *et al.* [29] were selected for training and test sets. Results of the GA runs for Model 1 were plotted in Fig. 3 as the number of parameters vs. statistical regression parameters. In Fig 2, it is seen that statistical regression coefficients approach 1 by increasing the number of parameters. But saturation point is reached

with 8 parameters. For this model, using parameters more than 8 would not influence the model performance much. As a general rule, the ratio between the number of compounds and the selected parameter is 5:1 [64]. Model 2 was generated by the use of the same size of training and test sets in Model 2, but with randomly selected compounds. By changing the size of training and test data (35 training and 17 test compounds), we obtained Model 3 with randomly selected compounds. Fig.2 and 3 represent the correlation between the number of parameters and some of the statistical regression coefficients. According to the. 2 and 3, the optimum numbers of parameters for Model 2 and 3 are 8 and 7, respectively
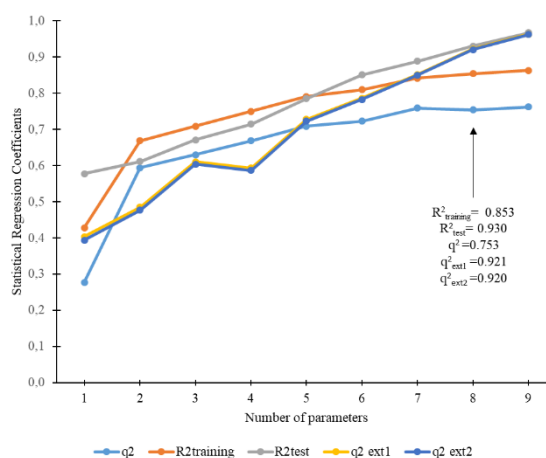

Fig. 3:   Correlation between the number of parameters and the statistical regression coefficients for Model 1.
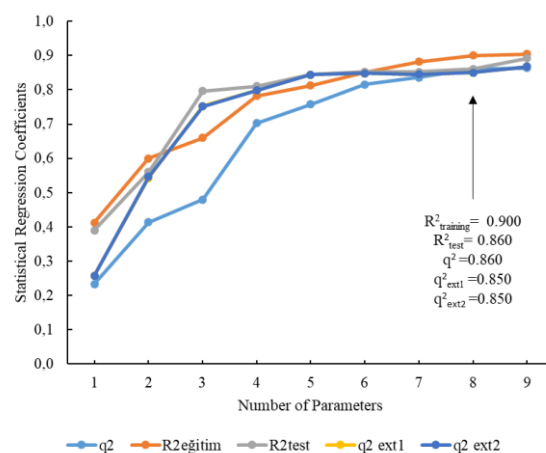

Fig. 4:   Correlation between the number of parameters and the statistical regression coefficients for Model 2.
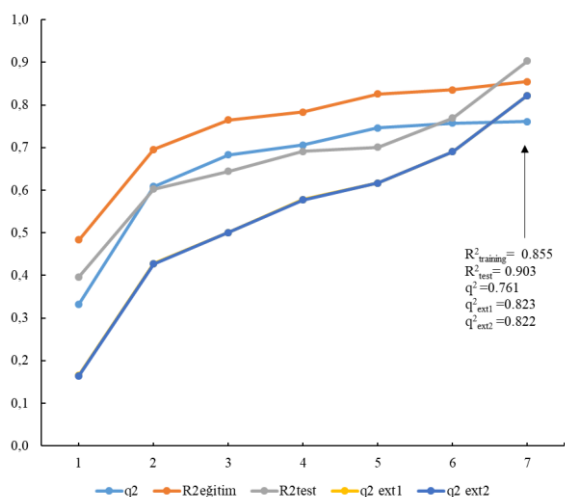
Fig. 5: Correlation between the number of parameters and the statistical regression coefficients for Model 3.

A comparison of models obtained by the EC-GA method with other models published in the literature was presented in Table 3. From Table 3, we can see that all of the EC-GA models give satisfactory statistical results both internally and externally in general. Considering Model 1 based on the same training and test compounds as reported by Lu *et al* [29], the EC-GA method gives better results than CoMFA and CoMSIA methods. The EC-GA method has the higher cross-validated correlation coefficient (q^2) and non-cross-validated correlation coefficients (R^2) for the test set than Ref. 29. In Model 2, randomly chosen 41 training compounds and 11 test compounds were used. If $R^2 > 0.7$ and $q^2 > 0.6$ for a QSAR model, this model is accepted as robust. q^2_{ext1} and q^2_{ext1} determine the capability of a model to make good external estimations (q^2_{ext1} and $q^2_{ext1} > 0.6$) [65]. Since the model fulfills both the internal and external validation requirements described in the literature, Model 2 is robust, capable of making good external predictions, and has high prediction power. Model 3 also has relatively high R^2, q^2, q^2_{ext1} and q^2_{ext1} values indicating the high prediction performance. From Table 3, all EC-GA models have satisfactory internal

and external validation results. Although CoMFA and CoMSIA models have high R^2 values, it does not indicate the high external prediction. Compared to EC-GA models, they have lower q^2 values. In Table 3, the results of the study by Nirouei *et al* were given for comparison. In their study, they developed linear and non-linear QSAR models of 40 indole glyoxamide derivatives by using different combinations of multiple linear regression (MLR), genetic algorithms (GA) and artificial neural networks (ANN). But in the mentioned study which uses the same indole glyoxamide derivatives, the number of compounds is different from the present study. Due to the different sizes of the entire data, this situation does not allow a proper comparison between the methods. However, it can be concluded that EC-GA models are better than models in Ref. 28. Among the EC-GA model, even though q^2_{ext1} and q^2_{ext1} values are higher the than other two models, it has relatively low R^2 and q^2 values. Because of this reason, we can conclude that Model 2 represents better predictivity than the others. It can be effectively employed in the development of novel indole glyoxamide derivatives as HIV-1 attachment inhibitors.

The predicted and experimental activity values of compounds are listed in Table 4. Training and test compounds in Table 4 were randomly chosen from the complete data. Test compounds were indicated with a """. As seen in the table, the residual value between the experimental and the predicted activity values are lower than 1 in general. This situation proves that the model has reliability and high prediction ability. The correlation between the experimental and the predicted activities of compounds was plotted in Fig. 6 for Model 2. Considering the results for Model 2, the optimum EC-GA model was achieved as a function of 8 parameters and had the following statistical results: $R^2_{training}=$, $R^2_{test} = 0.8$, cross-validated $q^2 = 0.860$, $q^2_{ext1} = 0.850$ and $q^2_{ext1} = 0.850$. Accordingly, the final model overcompensates all statistical threshold values in the literature by the use of conformational ensembles for each compound.

Table-3: Comparison of model performance.

Reference-Method	Training/Test size	q^2	$R^2_{training}$	R^2_{test}	q^2_{ext1}	q^2_{ext2}
Ref. 29- CoMFA	41/11	0.589	0.963	0.875	NA	NA
Ref. 29-CoMSIA	41/11	0.621	0.972	0.960	NA	NA
Ref. 28-MLR-MLR	32/8	NA	0.85	0.55	NA	NA
Ref. 28-MLR-ANN	32/8	NA	0.95	0.58	NA	NA
Ref. 28-GA-ANN	32/8	NA	0.96	0.75	NA	NA
Model 1 (EC-GA)	41/11	0.753	0.853	0.930	0.921	0.920
Model 2 (EC-GA)	41/11	0.860	0.900	0.860	0.850	0.850
Model 3 (EC-GA)	35/17	0.761	0.855	0.903	0.823	0.822

Table-4: Experimental and predicted activity values of 52 indole glyoxamide derivatives for Model **2**.

Compound	Experimental pEC$_{50}$	Predicted pEC$_{50}$	Compound	Experimental pEC$_{50}$	Predicted pEC$_{50}$
1	6.815	6.957	27 [t]	9.301	8.521
2	8.587	8.472	28	9.854	10.093
3	8.366	8.296	29 [t]	8.310	8.512
4	8.347	7.972	30	7.636	9.301
5	6.824	6.930	31	6.301	6.997
6 [t]	4.604	5.439	32	4.333	4.149
7	4.696	4.902	33	7.136	7.011
8	9.284	9.128	34	9.456	9.158
9	9.347	8.670	35	9.638	10.774
10	6.077	6.181	36	10.155	10.082
11 [t]	6.403	6.800	37 [t]	9.886	10.487
12 [t]	5.962	7.121	38	7.369	6.887
13 [t]	5.810	4.829	39	9.886	9.090
14 [t]	4.676	3.921	40 [t]	9.252	9.992
15	5.420	5.364	41	10.222	10.530
16	3.919	4.020	42 [t]	9.148	9.602
17	6.021	6.406	43	8.721	9.407
18	7.676	6.904	44	8.086	8.795
19	6.682	6.819	45 [t]	9.377	9.034
20	6.483	6.292	46	8.769	8.1945
21	8.137	7.123	47 [t]	6.577	6.4023
22	8.356	7.321	48	5.838	5.932
23	7.760	7.388	49	5.577	5.350
24	7.048	6.994	50	5.170	5.787
25	7.616	7.026	51	4.862	4.922
26	8.180	7.876	52	10.222	10.222

[t] Test set

Table-5: Definition of the optimal parameters for Model **2**.

$a_{nl}^{(j)}$	Molecular Parameters
$a^{(1)}$	**Orthogonal distance of H9 atom to O2-C10-C9 plane (Å)**
$a^{(2)}$	**Orthogonal distance of H2 atom to O1-N2-N3 plane (Å)**
$a^{(3)}$	**Orthogonal distance of N1 atom to O2-C10-N2 plane+ Van der Waals radius (Å)**
$a^{(4)}$	**Orthogonal distance of the nearest atom to O2-N2-N3 plane +Van der Waals radius (Å)**
$a^{(5)}$	**Orthogonal distance of the farthest atom to O2-N2-N3 plane + Van der Waals radius (Å)**
$a^{(6)}$	**Bond order of N1-H9 bond**
$a^{(7)}$	**Mulliken charge of the C2 atom**
$a^{(8)}$	**Interatomic distance between C4-N2 atoms (Å)**
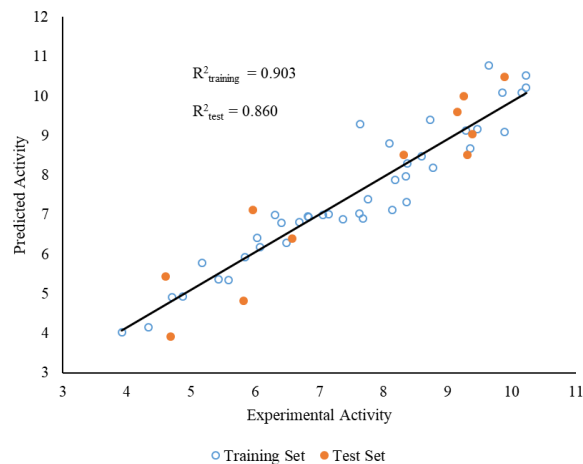
Fig. 6: Correlation between the experimental and the predicted activities for training and test data in Model 2.

The final model is constructed based on the 8 parameters which are the most significant for bioactivity. These parameters are given in Table 5. Selected parameters giving the best model with high statistical results are generally composed of geometrical and electronic parameters. The most influential parameters include especially pharmacophoric atoms which are responsible for the activity. The predominance of geometrical parameters shows that spatial orientations of atoms leading to conformations are of vital importance for the interaction between the drug molecules and bioreceptor.

Conclusion

In the current study, key pharmacophoric units of the 52 indole glyoxamide derivatives were revealed and the inhibitory activities were calculated by EC-GA method to gain insight about the ligand binding to the gp120 glycoprotein. Conformational ensembles were utilized to take into consideration the flexibility of ligands during the ligand-protein interaction. As well as docking studies are available for the different indole glyoxamide or azaindole glyoxamide derivative, only a few QSAR study was reported on the same data set examined in this paper. Because there is not a reported 4D QSAR study on indole glyoxamide derivatives in the literature, the current 4D QSAR study is original.

The pharmacophore elements were identified as a set of nine atoms proving that hydrogen bond donor and acceptors sites are the most favorable elements for the activity of inhibition and molecular recognition. The nitrogen atom placed in the indole ring is capable of making hydrogen bonds as a donor. Substituents at N1-position would prevent the formation of hydrogen bonds towards the H-bond acceptor site of the CD4 protein. Three carbonyl H-bond acceptors and two nitrogens in the piperazine are the other master key components of the hydrophilic interactions due to their relatively flexible positions orienting towards their complemental matches in the protein surface. When comparing the identified pharmacophoric features with the results in the literature, it is clearly seen that EC-GA method successfully determines the activity features for the HIV-1 attachment inhibition and produces compatible results with the similar studies in the literature [18,25-27]. Different 4D QSAR EC-GA models for indole glyoxamide derivatives were set up to compare the outcomes of other methods in the literature. All of the models were submitted to internal and external validation procedure. The final model was constructed with 8 parameters which give the optimal values for statistical parameters. This model disclose the importance of the utilization of multiple conformers for each compound. The optimum EC-GA model (Model 2) was found as a function of 8 parameters and had the following statistical values: $R^2_{training}$=, R^2_{test} =0.860, q^2 =0.860, q^2_{ext1} = 0.850 and q^2_{ext1} = 0.850. Accordingly, the best 4D QSAR model show high prediction power and robustness, which can be employed in the future design of new indole glyoxamide compounds. Attained pharmacophore and QSAR model provides insight into the vital interaction between HIV-1 binding inhibitory ligands and CD4 cell and can be effectively used for the design and development of the novel HIV-1 attachment inhibitor agents. This study will be the basis for the future 4D-QSAR model development study for indole glyoxamide derivatives with desirably increased potency.

References

1. A. A. Okoye and L. J. Picker, CD4+ T-Cell Depletion In Hiv Infection: Mechanisms Of Immunological Failure, *Immunol. Rev*., **254**, 54 (2013).

2. A. B. Bhatti, M. Usman, and V. Kandi, Current Scenario of HIV/AIDS, Treatment Options, and Major Challenges with Compliance to Antiretroviral Therapy, *Cureus,* **8**, 1 (2016).

3. World Health Organization, HIV data and statistics. https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics. Accessed 16 November 2021.

4. E. J. Arts and D. J. Hazuda, HIV-1 antiretroviral drug therapy, *Cold Spring Harb. Perspect. Med.*, **2**, a007161, (2012).

5. U. S. D. of H. and H. S. U.S. Food and Drug Administration, *HIV and AIDS: Medicines to Help You* (2019).

6. D. Warnke, J. Barreto, and Z. Temesgen, Therapeutic review: Antiretroviral drugs, *J. Clin. Pharmacol.*, **47**, 1570 (2007).

7. V. Summa, A. Petrocchi, F. Bonelli, B. Crescenzi, M. Donghi, M. Ferrara, F. Fiore, C. Gardelli, O. G. Paz, D. J. Hazuda, P. Jones, O. Kinzel, R. Laufer, E. Monteagudo, E. Muraglia, E. Nizi, F. Orvieto, P. Pace, G. Pescatore, R. Scarpelli, K. Stillmock, M. V. Witmer, and M. Rowley, Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection, *J. Med. Chem.*, **51**, 5843 (2008).

8. T. Wang, J. F. Kadow, Z. Zhang, Z. Yin, Q. Gao, D. Wu, D. D. G. Parker, Z. Yang, L. Zadjura, B. A. Robinson, Y. F. Gong, W. S. Blair, P. Y. Shi, G. Yamanaka, P. F. Lin, and N. A. Meanwell, Inhibitors of HIV-1 attachment. Part 4: A study of the effect of piperazine substitution patterns on antiviral potency in the context of indole-based derivatives, *Bioorganic Med. Chem. Lett.*, **19**, 5140 (2009).

9. 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. H. Wang, W. S. Blair, P. Y. Shi, P. F. Lin, R. J. Colonno, and N. A. Meanwell, Discovery of 4-benzoyl-1-[(4-methoxy-1H-pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2-(R)-methylpiperazine (BMS-378806): A novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions, *J. Med. Chem.*, **46**, 4236 (2003).

10. P. F. Lin, W. Blair, T. Wang, T. Spicer, Q. Guo, N. Zhou, Y. F. Gong, H. G. H. Wang, R. Rose, G. Yamanaka, B. Robinson, C. Ben Li, R. Fridell, C. Deminie, G. Demers, Z. Yang, L. Zadjura, N. Meanwell, and R. Colonno, A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding, *Proc. Natl. Acad. Sci. U. S. A.,* **100**, 11013 (2003).

11. H.-G. Wang, R. Williams, and P.-F. Lin, A Novel Class of HIV-1 Inhibitors that Targets the Viral Envelope and Inhibits CD4 Receptor Binding, *Curr. Pharm. Des.*, **10**, 1785 (2005).

12. Q. Guo, H.-T. Ho, I. Dicker, L. Fan, N. Zhou, J. Friborg, T. Wang, B. V. McAuliffe, H. H. Wang, R. E. Rose, H. Fang, H. T. Scarnati, D. R. Langley, N. A. Meanwell, R. Abraham, R. J. Colonno, and P. Lin, Biochemical and Genetic Characterizations of a Novel Human Immunodeficiency Virus Type 1 Inhibitor That Blocks gp120-CD4 Interactions, *J. Virol.*, **77**, 10528 (2003).

13. T. Wang, Z. Yin, Z. Zhang, J. A. Bender, Z. Yang, G. Johnson, Z. Yang, L. M. Zadjura, C. J. D'Arienzo, D. D. G. Parker, C. Gesenberg, G. A. Yamanaka, Y. F. Gong, H. T. Ho, H. Fang, N. Zhou, B. V. McAuliffe, B. J. Eggers, L. Fan, B. Nowicka-Sans, I. B. Dicker, Q. Gao, R. J. Colonno, P. F. Lin, N. A. Meanwell, and J. F. Kadow, Inhibitors of human immunodeficiency virus type 1 (HIV-1) attachment. 5. An evolution from indole to azaindoles leading to the discovery of 1-(4-benzoylpiperazin-1-yl)-2-(4,7-dimethoxy-1H-pyrrolo[2,3-c]-pyridin-3-yl) ethane-1,2-dione (BMS-488043), a drug candidate that demonstrates antiviral activity in HIV-1-infected subjects, *J. Med. Chem.*, **52**, 7778 (2009).

14. M. Pancera, Y. T. Lai, T. Bylund, A. Druz, S. Narpala, S. O'Dell, A. Schön, R. T. Bailer, G. Y. Chuang, H. Geng, M. K. Louder, R. Rawi, D. I. Soumana, A. Finzi, A. Herschhorn, N. Madani, J. Sodroski, E. Freire, D. R. Langley, J. R. Mascola, A. B. McDermott, and P. D. Kwong, Crystal structures of trimeric HIV envelope with entry inhibitors BMS-378806 and BMS-626529, *Nat. Chem. Biol.*, **13**, 1115 (2017).

15. N. A. Meanwell, M. R. Krystal, B. Nowicka-Sans, D. R. Langley, D. A. Conlon, M. D. Eastgate, D. M. Grasela, P. Timmins, T. Wang, and J. F. Kadow, Inhibitors of HIV-1 Attachment: The Discovery and Development of Temsavir and its Prodrug Fostemsavir, *J. Med. Chem.*, **61**, 62 (2018).

16. C. Lagishetty, K. Moore, P. Ackerman, C. Llamoso, and M. Magee, Effects of Temsavir, Active Moiety of Antiretroviral Agent Fostemsavir, on QT Interval: Results From a Phase I Study and an Exposure–Response Analysis, *Clin. Transl. Sci.*, **13**, 769 (2020).

17. M. E. Meuser, A. A. Rashad, G. Ozorowski, A. Dick, A. B. Ward, and S. Cocklin, Field-based affinity optimization of a novel azabicyclohexane scaffold HIV-1 entry inhibitor, *Molecules*, **24**, 1581 (2019).

18. R. Vangala, S. K. Sivan, S. R. Peddi, and V. Manga, Computational design, synthesis and evaluation of new sulphonamide derivatives targeting HIV-1 gp120, *J. Comput. Aided. Mol. Des.*, **34**, 39 (2020).

19. J. F. Kadow, J. Bender, A. Regueiro-Ren, Y. Ueda, T. Wang, K.-S. Yeung, and N. A. Meanwell, In *Antivir. Drugs*, John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 149–162, (2011).

20. T. Wang, Z. Yang, Z. Zhang, Y. F. Gong, K. A. Riccardi, P. F. Lin, D. D. Parker, S. Rahematpura, M. Mathew, M. Zheng, N. A. Meanwell, J. F. Kadow, and J. A. Bender, Inhibitors of HIV-1 attachment. Part 10. The discovery and structure-activity relationships of 4-azaindole cores, *Bioorganic Med. Chem. Lett.*, **23**, 213 (2013).

21. E. V. Nosova, G. N. Lipunova, V. N. Charushin, and O. N. Chupakhin, Fluorine-containing indoles: Synthesis and biological activity, *J. Fluor. Chem.* **212**, 51 (2018).

22. T. Liu, B. Huang, P. Zhan, E. De Clercq, and X. Liu, Discovery of small molecular inhibitors targeting HIV-1 gp120-CD4 interaction drived from BMS-378806, *Eur. J. Med. Chem.*, **86**, 481 (2014).

23. R. Kong, J. J. Tan, X. H. Ma, W. Z. Chen, and C. X. Wang, Prediction of the binding mode between BMS-378806 and HIV-1 gp120 by docking and molecular dynamics simulation, *Biochim. Biophys. Acta - Proteins Proteomics*, **1764**, 766 (2006).

24. C. Teixeira, N. Serradji, F. Maurel, and F. Barbault, Docking and 3D-QSAR studies of BMS-806 analogs as HIV-1 gp120 entry inhibitors, *Eur. J. Med. Chem.*, **44**, 3524 (2009).

25. C. G. Gadhe, G. Kothandan, T. Madhavan, and S. J. Cho, Molecular modeling study of HIV-1 gp120 attachment inhibitors, *Med. Chem. Res.*, **21**, 1892 (2012).

26. N. A. Meanwell, O. B. Wallace, H. Fang, H. Wang, M. Deshpande, T. Wang, Z. Yin, Z. Zhang, B. C. Pearce, J. James, K. S. Yeung, Z. Qiu, J. J. Kim Wright, Z. Yang, L. Zadjura, D. L. Tweedie, S. Yeola, F. Zhao, S. Ranadive, B. A. Robinson, Y. F. Gong, H. G. H. Wang, W. S. Blair, P. Y. Shi, R. J. Colonno, and P. fang Lin, Inhibitors of HIV-1 attachment. Part 2: An initial survey of indole substitution patterns, *Bioorganic Med. Chem. Lett.*, **19**, 1977 (2009).

27. A. Hosny, M. Ashton, Y. Gong, and K. McGarry, The development of a predictive model to identify potential HIV-1 attachment inhibitors, *Comput. Biol. Med.*, **120**, 103743 (2020).

28. M. Nirouei, G. Ghasemi, P. Abdolmaleki, A. Tavakoli, and S. Shariati, Linear and non-linear quantitative structure-activity relationship models on indole substitution patterns as inhibitors of HIV-1 attachment, *Indian J. Biochem. Biophys.*, **49**, 202 (2012).

29. P. Lu, X. Wei, and R. Zhang, CoMFA and CoMSIA studies on HIV-1 attachment inhibitors, *Eur. J. Med. Chem.*, **45**, 1792 (2010).

30. G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, Computational methods in drug discovery, *Pharmacol. Rev.*, **66**, 334 (2014).

31. S. Pirhadi and J. B. Ghasemi, 3D-QSAR analysis of human immunodeficiency virus entry-1 inhibitors by CoMFA and CoMSIA, *Eur. J. Med. Chem.*, **45**, 4897 (2010).

32. J. K. Buolamwini and H. Assefa, CoMFA and CoMSIA 3D QSAR and docking studies on conformationally-restrained cinnamoyl HIV-1 integrase inhibitors: Exploration of a binding mode at the active site, *J. Med. Chem.*, **45**, 841 (2002).

33. J. Verma, V. Khedkar, and E. Coutinho, 3D-QSAR in Drug Design - A Review, *Curr. Top. Med. Chem.*, **10**, 95 (2010).

34. S. C. Peter, J. K. Dhanjal, V. Malik, N. Radhakrishnan, M. Jayakanthan, D. Sundar, and D. Sundar, In *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, *Quantitative structure-activity relationship (QSAR): Modeling approaches to biological applications*, Elsevier, pp. 661–676, (2018).

35. C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, and A. J. Hopfinger, 4D-QSAR: Perspectives in drug design, *Molecules*, **15**, 3281 (2010).

36. A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, and C. Duraiswami, Construction of 3D-QSAR models using the 4D-QSAR analysis formalism, *J. Am. Chem. Soc.*, **119**, 10509 (1997).

37. E. Sarıpınar, N. Geçen, K. Şahin, and E. Yanmaz, Pharmacophore identification and bioactivity prediction for triaminotriazine derivatives by electron conformational-genetic algorithm QSAR method, *Eur. J. Med. Chem.*, **45**, 4157 (2010).

38. E. Yanmaz, E. Saripinar, K. Şahin, N. Geçen, and F. Çopur, 4D-QSAR analysis and pharmacophore modeling: Electron conformational- genetic algorithm approach for penicillins, *Bioorganic Med. Chem.*, **19**, 2199 (2011).

39. N. Geçen, E. Sarıpınar, E. Yanmaz, and K. Şahin, Application of electron conformational-genetic algorithm approach to 1,4-dihydropyridines as calcium channel antagonists: Pharmacophore identification and bioactivity prediction, *J. Mol. Model.*, **18**, 65 (2012).

40. L. Akyüz and E. Saripinar, Conformation depends on 4D-QSAR analysis using EC-GA method: pharmacophore identification and bioactivity prediction of TIBOs as non-nucleoside reverse transcriptase inhibitors, *J. Enzym. Inhib. Med. Chem.*, **28**, 776 (2013).

41. A. Özalp, S. Ç. Yavuz, N. Sabancı *et al.*, 4D-QSAR investigation and pharmacophore identification of pyrrolo[2,1-c][1,4]benzodiazepines using electron conformational–genetic algorithm method, *SAR QSAR Environ. Res.*, **27**, 317 (2016).

42. S. C. Yavuz, N. Sabanci, and E. Saripinar, Pharmacophore Modelling and 4D-QSAR Study of Ruthenium(II) Arene Complexes as Anticancer Agents (Inhibitors) by Electron Conformational-Genetic Algorithm Method, *Curr. Comput. Aided. Drug Des.*, **14**, 79 (2017).

43. S. Çatalkaya, N. Sabancı, S. Ç. Yavuz, and E. Sarıpınar, The effect of stereoisomerism on the 4D-QSAR study of some dipeptidyl boron derivatives, *Comput. Biol. Chem.*, **84**, 107190 (2020).

44. S. Kopru, F. O. Küp, N. Sabanci, M. Çadir, D. C. Bulut, F. Duman, I. O. İlhan, and E. Saripinar, DNA Cleavage Properties, Antimicrobial and Cytotoxic Activity and 4D-QSAR Analysis of Some Pyrazole Derivatives, *Lett. Drug Des. Discov.*, **16**, 904 (2019).

45. B. Tüzün and E. Saripinar, Molecular docking and 4D-QSAR model of methanone derivatives by electron conformational-genetic algorithm method, *J. Iran. Chem. Soc.*, **17**, 985 (2020).

46. C. G. Wermuth, C. R. Ganellin, P. Lindberg, and L. A. Mitscher, International Union of Pure and Applied Chemistry Chemistry and Human Health Division Medicinal Chemistry Section Glossary of Terms Used in Medicinal Chemistry, *Pure Appl. Chem.*, **70**, 1129 (1998).

47. O. F. Güner and J. P. Bowen, Setting the record straight: The origin of the pharmacophore concept, *J. Chem. Inf. Model.*, **54**, 1269 (2014).

48. S. Pickett, In *Protein-Ligand Interact. From Mol. Recognit. to Drug Des., The biophore concept. In: Protein-Ligand Interactions: From Molecular Recognition to Drug Design*, Wiley-VCH, Verlag, pp. 73–105 (2005).

49. S.-K. Lin, In *Development and Use in Drug Design. Pharmacophore Perception, Ed. O. F. Güner, Molecules*, **5**, 987-989, (2000).

50. O. Dror, A. Shulman-Peleg, R. Nussinov, and H. Wolfson, Predicting Molecular Interactions in silico: I. A Guide to Pharmacophore Identification and its Applications to Drug Design, *Curr. Med. Chem.*, **11**, 71 (2005).

51. B. Chandrasekaran, N. Agrawal, and S. Kaushik, in *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma, Pharmacophore development*, Elsevier, pp. 677–687, (2018).

52. Spartan'10; Wavefunction, Inc.: Irvine, CA, 2011.

53. I. Bersuker, Pharmacophore Identification and Quantitative Bioactivity Prediction Using the

Electron-Conformational Method, *Curr. Pharm. Des.*, **9**, 1575 (2005).

54. E. Saripinar, Y. Guzel, S. Patat, I. Yildirim, Y. Akcamur, and A.S. Dimoglo, Electron-topological investigation of the structure-antitubercular activity relationship of thiosemicarbazone derivatives, *Arzneimittel-Forschung/Drug Res.*, **46,** 824 (1996).

55. A. S. Dimoglo, N. M. Shvets, I. V. Tetko, and D. J. Livingstone, Electronic-Topological Investigation of the Structure - Acetylcholinesterase Inhibitor Activity Relationship in the Series of N-Benzylpiperidine Derivatives, *Quant. Struct. Relationships*, **20**, 31 (2001).

56. A. Altun, M. Kumru, and A. S. Dimoglo, Study of electronic and structural features of thiosemicarbazone and thiosemicarbazide derivatives demonstrating anti-HSV-1 activity, *J. Mol. Struc.-Theochem.*, **535**, 235 (2001).

57. A. S. Dimoglo, P. F. Vlad, N. M. Shvets, M. N. Coltsa, Y. Güzel, M. Saracoglu, E. Sarıpınar, and S. Patat, Electronic-topological investigations of the relationship between chemical structure and ambergris odor, *New J. Chem.*, **19**, 1217 (1995).

58. MATLAB (ver 7.0), The MathWorks Inc, 3 Apple Hill Drive, Natick, MA 01760-2098.

59. B. Tuzun, S.C. Yavuz, N. Sabanci and E. Saripinar, 4D-QSAR study of some pyrazole pyridine carboxylic acid derivatives by electron conformational-genetic algorithm method, *Curr Comput Aided Drug Des.*, **14**(4), 370 (2018).

60. J. Devillers, 1996. Genetic Algorithms in Molecular Modeling, 1st ed. Academic Press Inc, San Diego.

61. Organisation for Economic Co-operation and Development, Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models, OECD Series on Testing and Assessment No. 69, ENV/JM/MONO(2007)2, pg. 55-65.

62. G. Schuurmann, R.U. Ebert, J. Chen, B. Wang and R. Kuhne, External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean, *J. Chem. Inf. Model.*, **48**, 2140 (2008).

63. I. B. Bersuker, QSAR without arbitrary descriptors: The electron-conformational method, *J. Comput. Aided. Mol. Des.*, **22**, 423 (2008).

64. J.G. Topliss and R.P. Edwards, Chance factors in studies of quantitative structure-activity relationships, *J. Med. Chem.*, **22**, 1238 (1979).

65. N. Chirico, P. Gramatica, Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, *J. Chem. Inf. Model.*, **51**, 2320 (2011).