International Journal of Chemistry and Technology

http://dergipark.org.tr/ijct

Research Article

Confirmation by molecular docking of the pharmacophore defined as 4D-QSAR using the MCET method for the 2-hydroxydiarylamide derivatives

Burçin TÜRKMENOĞLU[1], Yahya GÜZEL[2*]

[1]Department of Pharmaceutical Basic Sciences, Faculty of Pharmacy, Erzincan Binali Yıldırım University,

Erzincan 24100, Turkey

[2]Department of Chemistry, Faculty of Science, Erciyes University, 38039, Kayseri, Turkey,

ABSTRACT

For the 2-hydroxydiarylamide derivative series that inhibits serine proteases by mediating various events related to the basic processes of tumor invasion and metastasis in cancer, the pharmacophore (Pha) responsible for the activity was estimated by the Molecular Conformer Electron Topological (MCET) method. The atoms in the common core structure of the molecules and template are aligned so that the remaining oriented atoms superimposition with the maximum number. Pha; in the 4D-QSAR analysis, it is selected and used to represent the molecule that can interact best by the receptor among all conformers. The model proposed in the training set was verified in the external test set and the electronic identifying values of the Pha atoms in both sets were considered. With Leave One Out-Cross Validation (LOO-CV) the model proposed according to 33 molecules in the training set ($q^2 = 0.998$) was validated by using 8 molecules ($r^2 = 0.993$) in the external test set. As a result of the MCET method, the proposed pharmacophore structure was confirmed using molecular docking.
.

Keywords: 2-hydroxydiarylamide derivatives, Molecular docking, 4D-QSAR, MCET, Klopman index.

2-hidroksidiarilamid türevleri için 4D-QSAR'da MCET yöntemi kullanılarak belirlenen farmakoforun moleküler docking ile doğrulanması

ÖZ

Pb (Kanserde tümör istilası ve metastazının temel süreçleri ile ilgili çeşitli olaylara aracılık ederek serin proteazları inhibe eden 2-hidroksidiarilamid türevi serisi için, aktiviteden sorumlu farmakofor (Pha), Moleküler Konformer Elektron Topolojik (MCET) yöntemi ile tahmin edilmiştir. Moleküllerin ve şablonun ortak çekirdek yapısındaki atomlar ile diğer yönlendirilmiş atomlar maksimum sayı olacak şekilde üst üste hizalanır. 4D-QSAR analizinde, Pha grubu tüm konformerler arasında reseptör tarafından en iyi etkileşime girebilen molekülü temsil etmek için seçilir ve kullanılır. Eğitim setinde önerilen model harici test setinde doğrulanmış ve her iki setteki Pha atomlarının elektronik tanımlayıcı değerleri dikkate alınmıştır. Bir veri dışarıda bırakılarak çapraz doğrulama (LOO-CV) ile eğitim setindeki 33 moleküle göre önerilen model ($q^2 = 0.998$) harici test setinde 8 molekül ($r^2 = 0.993$) kullanılarak doğrulanmıştır. MCET yönteminin bir sonucu olarak, önerilen farmakofor yapısı moleküler docking kullanılarak doğrulanmıştır.

Anahtar Kelimeler: 2-hidroksidiarilamid türevleri, Moleküler docking, 4D-QSAR, MCET, Klopman indeksi.

1. INTRODUCTION

Cancer is one of the most important health problems of our time. The aim of cancer treatment is to kill tumor cells in a manner that minimizes damage to the surrounding cells. Apoptosis (programmed cell death) is a physiological process that is commonly used in the treatment of cancer, allowing cells to be killed without producing an inflammatory response.[1] 2-hydroxydiarylamide family has been reported to have a variety of interesting biological properties, including fungicidal, antiviral and bactericidal activities.[2] However, the inhibitory activity of the 2-hydroxydiarylamide series against any serine protease in

cancer has not been previously reported in the literature. Despite this, it has been reported to be the strongest preventive activity against over-expression of colon cancer cells. Transmembrane protease/serine 4 (TMPRSS4) has been reportedin the literature.[2] TMPRSS4 is a type II transmembrane serine protease (TTSP), is a member of the family and has been found to be highly expressed in many cancers. TMPRSS4 is a new type II transmembrane serine protease that is highly in vivo on the cell surface in pancreatic, thyroid, colon and other cancerous tissues.[3] Recently, type II transmembrane serine proteases (TTSP) have been recognized as a novel sub-family of serine proteases that have an extracellular proteolytic region, a transmembrane region, and a short cytoplasmic region.[4-7] In previous studies, TMPRSS4 was found to be associated with cell migration, invasion, epithelial mesenchymal transition and stage progression in colon, metastasis and colorectal cancer cells.[8-10] The 2-hydroxydiarylamide derivative series and TMPRSS4 synthesized by researchers have been shown experimentally for a good inhibitory activity, evaluated to inhibit serine protease activity and suppress cancer cell invasion.[2]

Advances in computational chemistry in drug design have provided an important basis for the discovery of new drug candidates for biological activity.[11] In addition to the structure and function of the target, both ligand and structure-based studies are conducted to understand the mechanism by which it interacts with potential drugs.[12] Both methods play a vital role in modern drug design because it is a much cheaper and faster alternative to discovery.

Recently, both the 4D-Quantative Structure Activity Relationship (4D-QSAR) method and the molecular docking approach for multiple conformers have been proposed to overcome drug discovery challenges.[13]

The primary goal of this study is to propose the pharmacophore (Pha) model of the newly synthesized 2-hydroxydiarylamide compounds[2] by Klopman index 4D-QSAR analysis developed by us for the first time and to use the molecular docking method to confirm this model.

2. METHODS

The methods used in this study are given below according to the order of operation.

- SPARTAN'10 model molecular modeling program (Wavefunction, Irvine, CA, USA) for quantum chemical calculations; "Molecular Modeling" is used to represent or imitate molecules' lifelike behavior with all theoretical and computational methods.
- ETM program to convert quantum chemical results to Electron Topological Matrix (ETM); To

characterize molecular systems in 3D, atomic characters originating from molecular modeling methods were stored in matrix form for each conformer.

- The interaction points of the pharmacophore between the Ligand-Receptor (L-R) were determined by the descriptor type with the MCET program in the QSAR method, which is the most used in Ligand-based drug design calculations.
- FlexX docking program was used to calculate and observe binding interaction between L-R.

2.1. Quantum chemical calculations

Molecular conformational optimization was performed using Spartan'10 (Wavefunction, Irvine, CA, USA) using molecular mechanical field approach and quantum chemical calculations with B3LYP/6-31G level method. The conformers of a compound that are like each other are aligned to a common structure and recognized to avoid reuse. In the same compound, the high energy conformers were canceled due to their low population according to the Boltzmann distribution, and those with the lowest energy were included in the study. Of these conformers, only one that is most compatible with the receptor structure is considered responsible for the activity to represent the molecule. 41 compounds consisting of 2-hydroxydiarylamide derivatives as the anti-metastasis agent under investigation is taken from the literature and given in Table 1.[2]

2.2. Electron topological matrix program (ETMP)

The Electron Topological Matrix (ETM) of each conformer is a mathematical indicator where the conformer is represented by spatial and atomic descriptor numerical values. After the results obtained from the calculations in the quantum chemistry are recorded as Cnfrmr_No.txt file, the electronic values of atoms are Natural, Mulliken and Electrostatic charges, HOMO/LUMO orbital coefficients, Fukui index (f +, nucleophilic and f-, electrophilic functions) distances are converted into ETM with our in house software ETM program. The rows and columns of this matrix are ordered with the same index atoms. While the topological distance between different atoms is the descriptor electronic values between the same atoms, the upper triangular matrix is given in the diagonal region of the ETM. The non-diagonal values of the matrix, the distance between the atoms and the length of the bond will not change, so the conformer's space structure will remain the same. Accordingly, the descriptors of different genera used in diagonal values are represented by the corresponding values of atoms in the upper lines of the diagonal matrix and the descriptor name at the end of the line. Thus, the two-dimensional ETM is transformed into a three-dimensional matrix with the same distances and different descriptors as in Figure 1.

Int. J. Chem. Technol. **2021**, 5 (1), 11-25

Türkmenoğlu and Güzel

Table 1. Compound structures of 2-hydroxydiarylamide derivative series

Mol. No	Mol. Structures	Mol. No	Mol. Structures
n01		n22	
n02		n23	
n03		n24*	
n04*		n25	
n05		n26	
n06*		n27	
n07		n28	

Table1. Continued
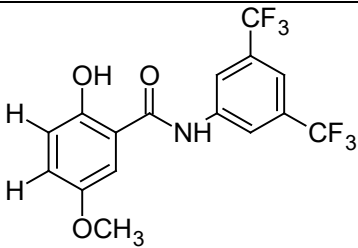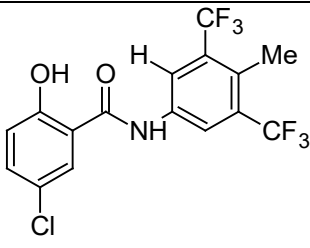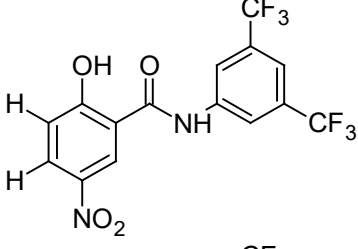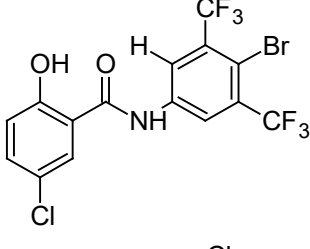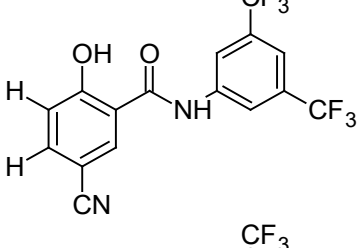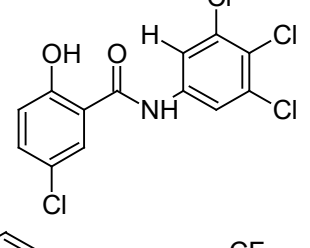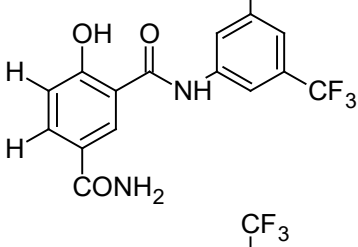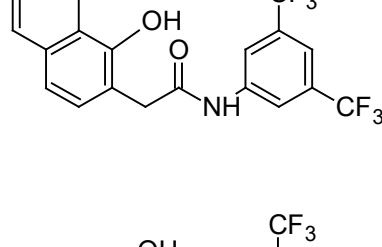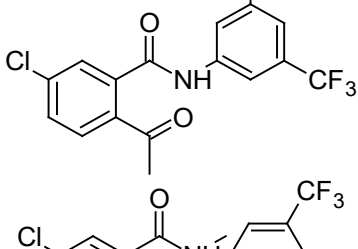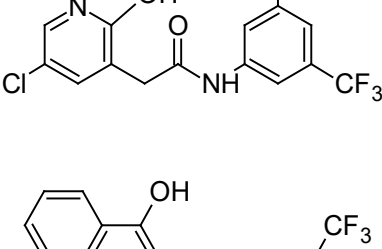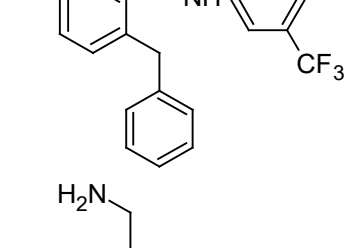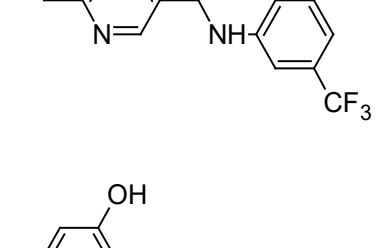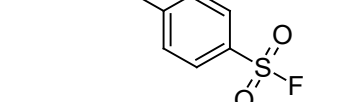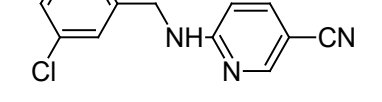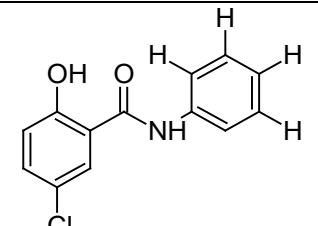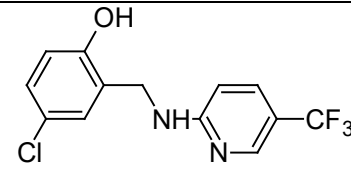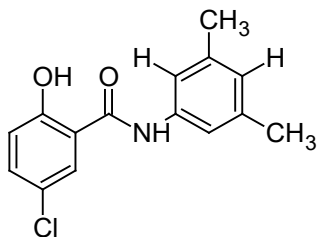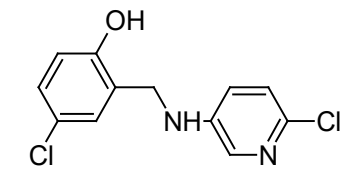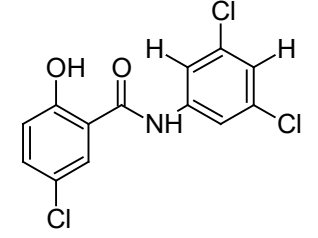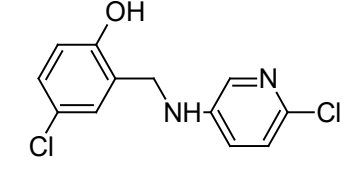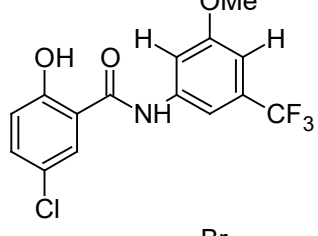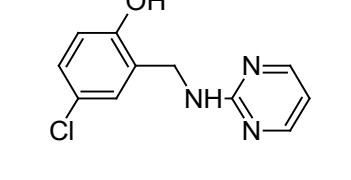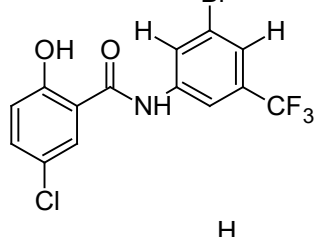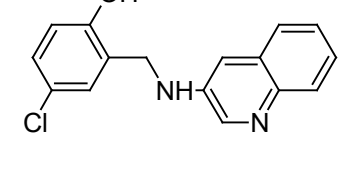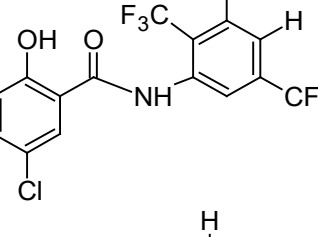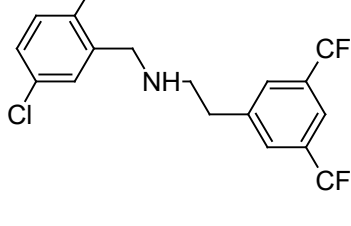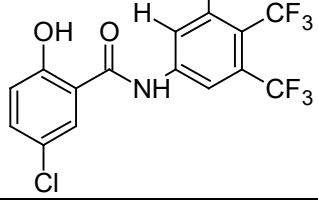
n08		n29	
n09		n30	
n10*		n31	
n11		n32	
n12*		n33	
n13		n34	
n14		n35*	

Table1. Continued

Table 2. Cartesian coordinate of the conformer n01_01.

Atom	X	y	Z
O2	0	0	0
N1	2.2841	0	0
C5	1.0195	4.9372	0
C6	-0.135	4.2161	0.246
O1	-1.299	2.1881	0.4297
C3	2.2411	2.8757	-0.325
C1	-0.133	2.8101	0.2077
C7	1.067	0.636-	0.022
C4	2.2088	4.259	-0.297
C2	1.0822	2.1212	-0.051
C8	2.5627	-1.38	0.0101
C9	3.3013	-4.093	0.0224
C10	1.5791	-2.375	-0.094
C11	3.9094	-1.752	0.1204
C12	4.2709	-3.097	0.1245
C13	1.9631	-3.715	-0.084
C14	0.9014	-4.786	-0.13
F1	-0.229	-4.353	-0.723
F2	0.5691	-5.199	1.113
F3	1.3288	-5.874	-0.806
C15	5.732	-3.464	0.17
F4	5.9284	-4.682	0.7121
F5	6.2711	-3.485	-1.069
F6	6.4455	-2.572	0.8936
H7	3.0965	0.5909	0.1005
H1	-1.138	1.2184	0.2988
Cl1	3.6678	5.1748	-0.65

Although the coefficients of the boundary orbitals in the atomic charges, Fukui index and ETM are optionally changed, the non-cross values do not change because the conformer's skeletal structure is constant. In the two-dimensional ETM shown in Figure 1, diagonal values are given by electrostatic charges and non-diagonal values, bond length or distance between atoms. If different diagonal values are used instead of the electrostatic charge value, a new two-dimensional ETM is created while the non-diagonal values remain constant. A three-dimensional matrix is formed with each successive ETM. It is possible to show this three-dimensional matrix in two dimensions, as shown in Figure 1, when the non-cross values are unchanged, and the cross values are given in separate lines.

Property	O2	N1	C5	C6	O1	C3	C1	C7	C4	C2	C8	C9	C10	C11	C12	C13	C14	F1	F2	F3	C15	F4	F5	F6	H7	H1	Cl1
Natural charge	-0.649	-0.6	-0.208	-0.266	-0.681	-0.218	0.385	0.696	-0.072	-0.221	0.166	-0.199	-0.217	-0.224	-0.15	-0.145	1.135	-0.36	-0.359	-0.363	1.134	-0.359	-0.359	-0.364	0.424	0.52	0.001
Mulliken charge	-0.571	-0.755	-0.125	-0.156	-0.641	-0.205	0.314	0.632	-0.077	0.03	0.374	-0.192	-0.181	-0.232	-0.012	-0.029	0.795	-0.263	-0.256	-0.267	0.792	-0.263	-0.255	-0.269	0.351	0.452	-0.016
Electrostatic charge	-0.5	-0.57	-0.104	-0.252	-0.545	-0.149	0.411	0.602	0.092	-0.265	0.449	-0.292	-0.315	-0.416	0.175	0.132	0.367	-0.14	-0.149	-0.143	0.372	-0.141	-0.15	-0.149	0.352	0.428	-0.114
f+, nucleophilic	0.094	0.127	0.175	0.118	0.001	0.012	0.01	0.289	0.086	0.092	0.009	0.037	0.005	0.017	0.026	0.02	0.008	0.001	0.005	0.004	0.01	0.001	0.007	0.001	0.056	0.001	-0.236
f-, electrophilic	0.004	0.036	0.018	0.017	0.001	0.051	0.008	0.012	0.464	0.007	0.007	-0.003	0.001	0.01	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.038	0.001	0.155
HOMO	0.003	0.029	0.015	0.014	0.001	0.041	0.007	0.01	0.371	0.006	0.006	0.001	0.001	0.008	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.03	0.001	0.324
LUMO	0.075	0.102	0.14	0.095	0.001	0.01	0.008	0.231	0.068	0.073	0.007	0.034	0.004	0.014	0.021	0.016	0.006	0.001	0.004	0.003	0.008	0.001	0.006	0.001	0.044	0.001	0.011
Interaction index	0.544	0.151	-0.066	0.057	0.662	-0.1	0.032	0.349	0.259	0.089	-0.17	0.109	-0.415	0.138	-0.191	0.113	0.174	0.296	0.571	0.977	0.009	0.511	0.815	0.18	-0.018	0.043	0.61

	O2	N1	C5	C6	O1	C3	C1	C7	C4	C2	C8	C9	C10	C11	C12	C13	C14	F1	F2	F3	C15	F4	F5	F6	H7	H1	Cl1	
O2	-0.5	2.284	5.041	4.225	2.58	3.66	2.821	1.242	4.806	2.381	2.91	5.258	2.853	4.286	5.277	4.202	4.872	4.419	5.348	6.077	6.699	7.588	7.253	6.997	3.154	1.694	6.376	O2
N1		-0.57	5.096	4.867	4.22	2.894	3.713	1.373	4.27	2.438	1.408	4.217	2.479	2.393	3.681	3.73	4.983	5.079	5.587	6.006	4.89	5.976	5.402	4.973	1.009	3.645	5.396	N1
C5			-0.104	1.383	3.622	2.418	2.428	4.301	1.401	2.817	6.503	9.314	7.334	7.288	8.668	8.704	9.725	9.402	10.2	10.84	9.634	10.82	9.983	9.307	4.818	4.309	2.737	C5
C6				-0.252	2.345	2.788	1.406	3.786	2.407	2.441	6.217	8.994	6.819	7.211	8.539	8.211	9.069	8.625	9.482	10.25	9.665	10.77	10.1	9.476	4.859	3.161	4.023	C6
O1					-0.545	3.685	1.339	2.866	4.138	2.43	5.275	7.796	5.42	6.539	7.685	6.764	7.335	6.728	7.651	8.57	9.025	9.976	9.578	9.102	4.688	0.991	5.895	O1
C3						-0.149	2.435	2.546	1.383	1.409	4.281	7.057	5.297	4.94	6.324	6.601	7.78	7.65	8.371	8.811	7.254	8.473	7.567	6.988	2.476	3.815	2.725	C3
C1							0.411	2.494	2.8	1.421	4.987	7.713	5.469	6.097	7.369	6.86	7.674	7.224	8.091	8.865	8.589	9.651	9.071	8.528	3.92	1.884	4.558	C1
C7								0.602	2.8	1.485	2.51	5.23	3.055	3.715	4.921	4.443	5.426	5.202	5.966	6.563	6.213	7.242	6.72	6.329	2.033	2.303	5.268	C7
C4									0.092	2.428	5.658	8.429	6.666	6.261	7.651	7.98	9.14	8.961	9.703	10.18	8.501	9.736	8.779	8.125	3.794	4.561	1.758	C4
C2										-0.265	3.802	6.599	4.523	4.799	6.118	5.902	6.91	6.64	7.431	8.035	7.27	8.388	7.706	7.189	2.534	2.422	4.045	C2
C8											0.449	2.811	1.402	1.401	2.424	2.412	3.792	4.144	4.447	4.731	3.796	4.767	4.398	4.156	2.044	4.531	6.68	C8
C9												-0.292	2.435	2.419	1.393	1.394	2.502	3.617	3.143	2.784	2.515	2.779	3.222	3.599	4.689	6.928	9.299	C9
C10													-0.315	2.421	2.795	1.394	2.505	2.753	3.233	3.58	4.301	4.989	4.919	4.969	3.337	4.522	7.853	C10
C11														-0.416	1.392	2.771	4.279	4.96	4.901	4.95	2.5	3.606	3.161	2.774	2.48	5.86	6.974	C11
C12															0.175	2.398	3.777	4.748	4.37	4.151	1.507	2.367	2.361	2.365	3.87	6.922	8.33	C12
C13																0.132	1.508	2.371	2.362	2.363	3.785	4.158	4.425	4.728	4.456	5.84	9.069	C13
C14																	0.367	1.348	1.351	1.35	5.017	5.098	5.604	6.057	5.812	6.356	10.35	C14
F1																		-0.14	2.174	2.178	6.093	6.331	6.567	7.095	6.015	5.737	10.29	F1
F2																			-0.149	2.171	5.527	5.399	6.341	6.441	6.399	6.691	10.97	F2
F3																				-0.143	5.114	4.988	5.495	6.322	6.764	7.591	11.29	F3
C15																					0.372	1.347	1.352	1.352	4.836	8.315	8.919	C15
F4																						-0.141	2.173	2.18	6.016	9.215	10.2	F4
F5																							-0.15	2.172	5.297	8.882	9.052	F5
F6																								-0.149	4.674	8.499	8.373	F6
H7																									0.352	4.285	4.68	H7
H1																										0.428	6.297	H1
Cl1																											-0.114	Cl1

Mulliken charge of N1= -0.57
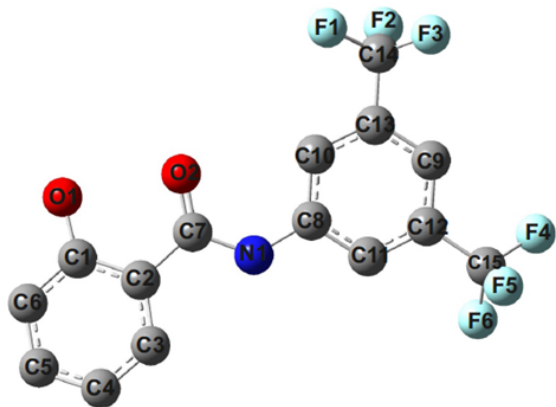**O2-C7 bond length=1 .242 Å**
F1-C10 distance= 2.753 Å

Figure 1. Three-dimensional ETM is represented by the same non-diagonal and different diagonal values.

For each conformer, the three-dimensional upper triangular matrix of the remaining atoms is found, ETM.txt, and for the sake of simplicity, except for functional atoms such as O, N, S in the molecular skeleton, the hydrogen atoms bound to carbon atoms are ignored. Then, for each conformer, the three-dimensional upper triangular matrix is as ETM_txt, and the Cartesian coordinates keep the x-, y-, z-values (Table 2) as Cart_txt in the files. Coordinates of the first three atoms; Other coordinate values are arranged as in Table 2 with reference to (0,0,0), (x, 0,0) and (0, y, z). For each Pha experiment, this process is applied to the coordinates of the remaining atoms, referring to the first three atoms of the nucleus structure being made similar 'zero'. Thus, all atoms of each ligand are brought to the appropriate and common positions relative to the receptor with maximum overlap and real 3D interaction is provided

2.3. MCET Method

MCET program [14-21] written by GÜZEL is used to operate 3D/4D-QSAR methods. Like CoMFA/CoMSIA, it works on the principle of aligning and matching atoms in ligands. common core structure in all molecules that are part of the pharmacophore can form the beginning of pairing. The electronic and geometric similarities of conformer atoms in the molecules align according to the proposed core structure. According to the core structure, the conformer with the maximum number of atoms matching the template is chosen to represent its molecule. However, it is possible to select representatives of different conformer structures for each of the core structures derived from the template conformer. First, the electronic and geometric similarity of the atoms in the molecular nucleus structure, depending on their tolerance values, is a good start for both the aggregation of the atoms in the molecules and the determination of the pharmacophore structure. Significant difference of our method from the above two methods:
i. Molecules are placed in the grid by overlapping with the template conformer instead of using the reference volume tolerance dτ value instead of grid spacing.

ii. It is extremely useful to know the active conformation of the drug receptor complex for drug design to be effective. Therefore, in order to explain the interactions between L-R, it is important to predict this conformer which is responsible for the activity for each molecule. Since there is a dynamic balance between conformers, the reduction in the amount of compatible conformer held by the receptor is compensated by others, and the penetration rates of the molecules remain the same according to the Boltzmann distribution. As discussed herein, the interaction of multiple conformers with the receptor results in 4D-QSAR.

iii. The conformer compatible with the receptor may not always be at the lowest energy level. One of the conformers of any molecule that best matches the template is used to determine Pha, as it will be in the optimal position for interaction with the receptor.

iv. Recently, different local reactive descriptor (LRD) has been used as chemical properties responsible for ligand-receptor interaction to predict a ligand-based model. The energy resulting from the interaction points between the L-R varies depending on the type and value of the LRD selected in the program interface. Mulliken / Natural / Electrostatic atomic charges, HOMO / LUMO coefficients, Fukui index and Klopman index are used as LRD type atoms in the interface of the MCET program. [16, 17, 22] With the descriptors used in the MCET method, research can be conducted with different series showing preventive properties, especially cancer. In other words, with these LRDs, the compounds studied on anticancer can be screened, as well as the activities of the compounds that will inhibit other diseases can be calculated in detail. The aggregation of atoms depends on the LRD types. Since the LRD type, an independent variable in the ligands, changes, an aggregation of different clusters will occur due to different alignment and different overlap. In order to determine the structure of the binding site of the 3D receptor more realistically, the LRD type should be chosen well. Therefore, the established model may be close to reality. A good model cannot be created using a wrong identifier. A good LRD selection is required to get a better Pha. In determining the Pha of a molecule, different types of probes are used as descriptors in CoMFA/CoMSIA, while the electronic properties of atoms in MCET are used in ETM with different LRD types. One of the LRDs we use as the descriptor is more effective than the probe used in CoMFA / CoMSIA. Because multiple probes are required to define an interaction point, a single descriptor (e.g. Klopman index) used as LRD in MCET is ionic (positive / negative), hydrogen bond (donor / acceptor) is examined by many features such as non-polar interactions. More realistic results can be achieved with the more comprehensive and general descriptor used in our software MCET method.

The steps of the algorithm operated in MCET are summarized as follows:[16, 17, 22]

- The simplest and least compatible molecule are selected as a template because it is considered one of the most compatible with the receptor.
- Groups consisting of combinations of 3, 4, 5 or 6 atoms containing at least one functional atom are the starting or core structure (CS) of the active structures.
- At least one active molecule conformer should contain this CS. This is determined by the tolerance given by the defining values of the atoms in the formula and the coordinate values in space. Accordingly, it is ordered from the number of molecules to the minimum number of CS.
- Using a tolerance slightly greater than the spatial geometric tolerance applied to CS, the positions of

the oriented atoms in the conformers of all molecules are overlapped with the atomic positions of the template conformer. The tolerances used for both CS and spatial positions vary in specific increments, so overlaps are checked for suitability of all molecules. The optimum tolerance value is determined by the best overlap.

- The conformer having the largest number of atoms matching the atoms in the template conformer is selected to represent other conformers in its molecule.

- The reference positions are taken from the atoms of at most and least active molecules. The conformers of the selected molecules containing the atoms in these positions are aligned to be superposition.

- At the positions in space, the receptor-side interaction parameter values corresponding to this position are calculated using the descriptor values of the atoms of the clustered molecules. Here, the descriptor types of atoms are natural, Mulliken, electrostatic atomic charges; the coefficients or Fukui indexes in HOMO/LUMO have used, as well as the Klopman index, which was first introduced by us to the literature. As expected in this study, Klopman index gave better results. The reason for this is that a more realistic analysis can be performed because the Klopman index not only involves the atomic charge but also the coefficients in the boundary orbitals. An important detail to consider when using the Klopman index is; if the HOMO coefficients of the ligand are used, the LUMO coefficients of the receptor are used and the receptor that provides the electron in the ligand serves as the hydrogen. HOMO or LUMO of the ligand can be selected as the descriptor, or one of the atomic charges (natural, Mulliken and electrostatic) can be selected. The Klopman index, natural-HOMO, natural-LUMO, Mulliken-HOMO, etc., from each combination of these selections different descriptors can be created.

- Positions which may be interaction points are determined by Genetic Algorithm (GA). The model is matured by adding new positions to the CSs discussed with this algorithm. Whether a new position can be added to the position of the model is determined by a significant increase in the statistical value. The positions in the resulting model show the superposition of the ligands, including the CSs, but reveal the pharmacophore structure of the receptor as a negative image of these positions. The CS in the superposition constitutes a common structure in which each molecule interacts with the receptor, while the remaining directed atoms act as enhancer or reducing activity. Each molecule is determined to increase or decrease the activity relative to its atomic descriptor value at that position.

- By aligning the CSs of the molecules, the greatest number of matching atoms of the conformer that

will interact best with the receptor helps to select the most compatible one. Therefore, the most accurate structure of the CS is found by accurately overlapping the conformers.

- Those involved in the interaction between the clusters of superimposed atoms can give rise to superposition and may be independent variables of the model. The terms align, superimpose and superposition used in QSAR investigations are expressed separately in MCET although they are used interchangeably.

- Adding a new position within the models is determined by someone who has the best statistical result with PLS measurements. PLS calculations are treated as NLLS because the interaction energy resulting from the ligand-receptor pharmacophore structure is used as a model in non-linear equality. For each CS and NLLS calculations of clusters of oriented atoms, the best statistical results are obtained for the most appropriate tolerance values. The tolerance values during the clustering of each CS and molecules carrying atoms and atoms around it, the limit that is possible is rated from minimum to maximum. NLLS calculations are performed using the Levenberg-Marquardt algorithm.

- Molecules are divided into two as training and test set by Gaussian process. This process ensures that the molecules in the test set are best represented by those in the training set. In order to do this, there must be an agreement between the independent variables of the molecules in both sets. A machine learning algorithm including the Gauss process uses the similarity between points to estimate the value of an unseen point in the test set, utilizing the training data. The activity of each molecule is validated according to q^2 values with the model proposed by using LOO-CV in the training set. The activities of the molecules in the excluded test set are predicted by using the same model. [16, 17, 22]

As a result, the steps processed within the MCET method; selection of template conformer, alignment of CSs and selection of molecular conformers from superimposed atoms around it, clustering of molecules, separation of training and test sets according to GP, suggestion of the model with GA, superposition of calculation of activities with non-linear equality and determining q^2 and r^2 values by using PLS in LOO-CV. [16, 17, 22]

Energy calculations in biological interaction are a long and difficult question in practice. To solve this problem, the energy contribution of L-R interaction and the selected conformal energy contribution of molecular conformers are handled together. Conversion of conformers of the same molecule in a selected conformer can be achieved by energy intake or application. The receptor binding energy of conformers should be higher during conversion between conformers. Thus, a

conformer may represent other conformers within the molecule to interact with the receptor. One of the most important features that distinguish MCET from others is the interaction of a single conformer with the most suitable receptor. Another feature of our method is that it can use LRDs such as atomic charge, Fukui index and HOMO/LUMO coefficients as well as the Klopman index introduced by you. They provide information about the electronic structures of molecules that interact with the receptor.

2.4. Partial least squares (PLS) analysis

The partial least squares statistical method used in the derivation of 4D-QSAR models is an extension of the multiple regression analysis where the original variables are replaced by a small set of linear combinations. In this study, the PLS method with Leave One Out-Cross Validation (LOO-CV) was used to determine the optimal number of components using the cross-validation coefficient q^2. External validation of the various models was performed using a test set of five molecules. The final analysis was performed using the optimal number of components obtained from to obtain the correlation coefficient (r^2). The value of q^2 determines the internal predictability of the model, while the value of r^2 evaluates the internal consistency of the model. Thus, the best QSAR model was chosen based on the values of q^2 and r^2.

2.5. Klopman Index

We first introduced the Klopman-Salem equation used to describe the regional/stereo selective reactions as a new local reactive descriptor and used the equation [23-26] to demonstrate ligand-receptor interaction. Therefore, we have given the name of "Klopman's Index" in a similar way to another LRD. A new molecular orbital, ψ_{AB}, is formed from the linear combinations ψ_A and ψ_B orbitals by using c_a and c_b weight coefficient between two separate (A and B) molecules. For a simple Lewis acid-base (A + B → AB) reaction, how the wave function ψ_{AB} is calculated in Klopman and Jensen perturbation molecular orbital (MO) theory is given in Eq.(1).

$$\psi_{AB} = c_a\psi_A + c_b\psi_B \tag{1}$$

According to the Klopman-Salem equation, the energy gain (or loss) can be calculated when the corresponding atoms of the two separate compounds interact with electrostatic and covalent, separately or both. The detailed form of the Eq.(2) is given below.[23-26]

$$\Delta E = -\sum_{a\in A, b\in B}^n (q_a + q_b)\beta_{ab} S_{ab} + \sum_{k\in A, l\in B}^n \left(\frac{Q_k Q_l}{\varepsilon R_{kl}}\right) +$$
$$\sum_{r\in A}^{occ} \sum_{s\in B}^{unocc} \cdot - \sum_{s\in B}^{occ} \sum_{r\in A}^{unocc} \cdot \frac{2\left(\sum_{a\in A, b\in B}^n c_{ra} c_{sb}\beta_{cb}\right)^2}{E_r - E_s} \tag{2}$$

A and B are interacting molecules n is number of interaction points between two molecules q_a is the electron population in the atomic orbital a_i, i =1, 2, 3…n β and S are resonance and overlap integrals Q_k is the total charge on atom k_i ε is the local dielectric constant R_{kl} is the distance between the atoms k_i and l_i, c_{ra} is the coefficient of atomic orbital a_i in molecular orbital r, (r refers to the molecular orbitals on one molecule). E_r is the energy of molecular orbital r.

$$\Delta E \approx \sum_i \left(\frac{Q_{nuc} Q_{elect}}{\varepsilon R_i}\right) -$$
$$2\sum_i^n \frac{\left(c_{nuc,i} c_{elect,i}\beta_i\right)^2}{E_{LUMO(elec)} - E_{HOMO(nuc)}} \tag{3}$$

For the interaction between the ligand and the receptor as nucleophiles and electrophiles in the binary system, the first term of Eq. (3) means Coulombic repulsion or attraction, while the second term in Eq. (3) refers to the contribution in the frontier orbital term. Q_{nuc} and Q_{elec} represent atomic charges, c_{nuc} and c_{elec} orbital coefficients, R distance between atoms, ε local dielectric constant, β resonance integrals, E_{HOMO}/E_{LUMO} cross-frontier energy values.

During the interaction of the ligand and the receptor, the distance between the ligand and the receptor, the dielectric constant, and the values Kappa (κ) and Xi (ξ) of the position points in the receptor field are continuously given in both terms, and Kappa and Xi can be summed up as constants respectively.

The output of the model is a non-linear function arising from atomic descriptors on the ligand and adjustable constants in the receptor domain. In order to find a model that gives a good statistical result, a significant correlation of the adjusted constants calculated according to the observed activity values should be ensured. The statistical results of experimental and predicted efficiency values were calculated by NLLS method. Estimated activity values can be determined after the adjustable constants of all positions in the receptor area are calculated at the same time.

2.6. Molecular docking

Molecular docking techniques are often used to investigate how drugs, drug candidates, enzymes, nucleic acids, and receptor proteins in computer-aided rational drug design fit together.[27, 28] In these docking studies, the binding energy of the three-dimensional receptor can be determined and the binding site between the ligand-receptor can be visualized. This can be useful for understanding the type of attachment and for designing smaller, more compatible ligands targeting proteins.[29]

Table 3. Observed and predicted activities of 41 2-hydroxydiarylamide derivatives.

Mol.No	Obs.(pKi)	Pred.(pKi)	Mol.No	Obs. (pKi)	Pred.(pKi)
n01	1.960	1.960	n22	1.520	1.495
n02	1.000	1.098	n23	1.460	1.471
n03	1.890	1.876	n24*	1.460	1.452
n04*	1.540	1.489	n25	1.550	1.545
n05	1.920	1.867	n26	1.800	1.829
n06*	1.920	1.913	n27	1.890	1.884
n07	1.550	1.582	n28	1.120	1.064
n08	1.540	1.529	n29	1.920	1.934
n09	1.920	1.920	n30	2.000	2.020
n10*	1.720	1.730	n31	2.220	2.217
n11	1.000	1.019	n32	1.000	1.005
n12*	1.000	1.014	n33	1.000	1.007
n13	1.000	1.005	n34	1.000	1.015
n14	1.410	1.406	n35*	1.000	0.951
n15	1.300	1.379	n36*	1.100	1.141
n16	1.410	1.421	n37	1.000	1.009
n17	1.720	1.666	n38	1.000	0.990
n18	1.800	1.823	n39	1.000	1.009
n19*	1.920	1.875	n40	1.000	1.000
n20	1.700	1.653	n41	1.070	1.052
n21	1.890	1.891	*Compounds in the test set		

Molecular docking was applied to investigate molecular scavenging using a patented search engine to place a scoring function and ligands in the binding site of a protein. The crystal structure of the TMPRSS4 receptor from the RCSB Protein Data Bank (PDB ID: 5CE1). The ligands enter the bonding zone with the FlexX docking. All the water molecules in 5CE1 were deleted and polar hydrogen atoms were added.

In this article we implemented automatic docking. The 5CE1 structure was used in subsequent docking experiments without energy reduction. FlexX docking was used to visualize the binding mode between protein and ligand.

3. RESULTS AND DISCUSSION

The data set is divided into two groups, 33 compounds are selected as training set and 8 compounds as test set. Observed and predicted activities of all compounds in training and test sets are given in Table 3. This dataset was used to construct the 4D-QSAR (MCET) model and to analyze its physicochemical properties. In the molecular docking calculations were made according to the molecule.

3.1. MCET results

TTMPRSS4 has been reported to be used as an active target in cancer treatment by minimizing cancer spread in the body.[2] Established the role of TMPRSS4 serine proteolytic effect in tumor cell invasion, it has been shown to be a novel approach in the treatment of malignant tumors.[2] In order to use pKi values in activity calculations, inhibitor Ki (nM) values were arranged as log (1 / Ki) on a negative logarithmic scale. In the study,

data sets of 41 compounds with structural diversity were divided into training test sets with 33 and 8 compounds, respectively. The test set consists of approximately 20% of the whole set, ensuring that it contains representative samples of the training set and the range of activity values of the training set. For both the training and test sets, the activities calculated by the MCET method were compared to the experimental results obtained from the article. Accordingly, activity calculations, Klopman index, Fukui index, Natural, Mulliken and Electrostatic charges were made with different descriptors and the statistical results are shown in Table 4.

Table 4. The values of q^2 and r^2 according to different descriptors. (The descriptor with the best results is marked in bold.)

Descriptors	q^2	r^2
Natural Charge	0.822	0.465
Mulliken Charge	0.870	0.757
Electrostatic Charge	0.832	0.685
N_Fukui (f+)	0.863	0.824
P_Fukui (f-)	0.981	0.995
HOMO	0.920	0.913
LUMO	0.932	0.867
Natural HOMO Klopman	0.996	0.986
Mulliken HOMO Klopman	0.995	0.991
Electrostatic HOMO Klopman	**0.998**	**0.993**
Natural LUMO Klopman	0.995	0.991
Mulliken LUMO Klopman	0.997	0.996
Electrostatic LUMO Klopman	0.993	0.994

Pha information of the descriptor which gives the best statistical result is given in Table 5.

Table 5. Klopman index using predicted Pha atoms and parameter constants (The atomic positions of CS are shown in italics).

Mol. No	Atom No	x, y, z Cartesian Coor.			Position letters	Kappa-(κ) Values	Xi-(ξ) Values
		x	**Y**	**z**			
n01_01	*O2*	*0*	*0*	*0*	*a*	-0.059	-0.2120
n01_01	*O1*	*2.5810*	*0*	*0*	*b*	-0.186	-1.0120
n01_01	*C12*	*-4.7550*	*2.2890*	*0*	*c*	-0.110	-5.9390
n02_04	C15	3.7070	1.9950	0.8160	d	0.4930	39.9290
n02_04	*N1*	*-1.2090*	*2.0160*	*0.7430*	*e*	0.2150	11.1940
n02_04	C11	1.2080	1.9760	0.6950	f	0.5020	-4.2760
n02_04	C1	-3.9010	1.0920	-1.6680	g	0.4570	27.3150
n02_04	F1	1.088	-2.037	-2.149	h	0.066	7.0660
n12_04	F6	4.418	-2.329	1.297	i	1.339	-0.8660
n13_02	C3	1.907	1.862	-2.233	j	0.449	0.9980
n13_02	C13	2.054	-3.352	1.609	k	-1.079	-2.3240
n02_04	C3	-1.602	1.652	-2.210	l	-0.231	-8.3280
n02_04	F3	2.337	-2.716	-0.504	m	0.941	3.558
n11_01	C11	-2.800	3.890	-0.115	n	-0.163	15.459
n12_04	C17	-2.139	-0.136	-1.006	o	-0.618	-5.754
n12_04	C10	4.389	2.103	-0.959	p	0.713	-28.826
n01_01	F4	-6.835	3.236	-0.615	r	0.026	-34.854
n29_03	H1	1.066	-1.466	1.622	s	-0.340	-8.486
n02_04	F2	0.164	-2.765	-0.321	t	0.323	-5.353
n11_01	C15	-4.569	5.625	-0.391	u	0.513	-5.473
n12_04	F5	6.526	-2.014	0.844	v	-0.676	-1.508
n32_03	O2	-1.486	-0.033	1.914	y	0.198	-0.138

Herein, according to the MCET method, the positions of the reference molecules from which Pha is obtained are marked with a, b, c and other letters according to the cartesian coordinate values of the corresponding atoms. These letters are the geometric information of the interaction areas of Pha and the Kappa and Xi parameters of the corresponding receptor are calculated. Although the Kappa constant contains the charge value of the corresponding atom in the receptor, the constant Xi is the coefficient of the atom. At these points given as a, b, c and so on between L-R, either the pulling force or the pushing force occurs due to negative or positive markings of both the ligand and the receptor. As can be seen from

Eq, (3), in the first term, for atomic charges, if both sides are the same marked, the pulling force arises if the thrust is marked differently. On the other hand, if the coefficients in the second term are in the same sign, if the tensile force is in the different sign, the thrust force occurs. These forces have a greater effect due to the greater absolute value of the numbers in Kappa and Xi.
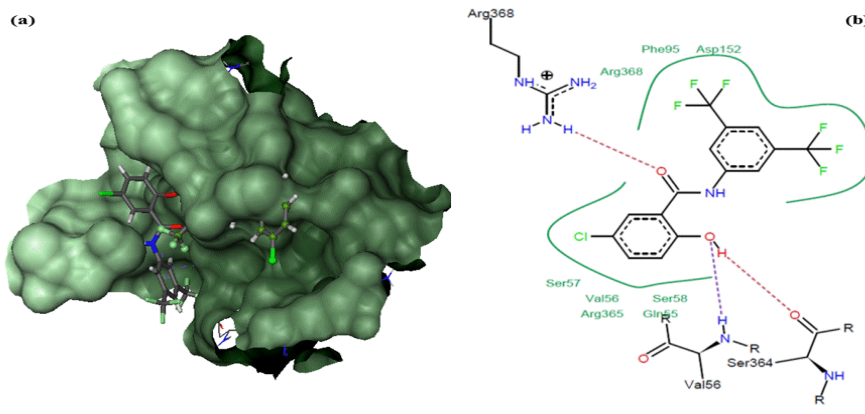
L-R complex parameters; the descriptor is the structure and receptor of CS with distance tolerance values. The structure of CS, which is one of the interaction points shown in Figure 2, is given by points a, b, c and e.
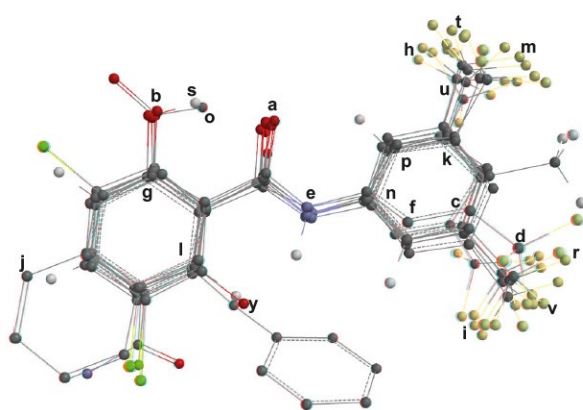
Figure 2. Superimposition of **n01**, **n02**, **n11**, **n12**, **n13**, **n29** and **n32** molecules determination of Pha.

The first three atoms of each candidate CS (called a, b and c) are aligned with the starting points (0,0,0), (x, 0,0) and (x, y, 0). The positions of the remaining atoms in the conformation are rearranged according to the Cartesian coordinate values of x, y and z. Thus, the atoms of all ligands having coordinate values in space may interact with the directed points of the receptor position.

Using two different tolerance values, only the CS atoms are aligned with respect to the position tolerance value in the remaining set of directed atoms. The similarity and identity of CS atoms is the source of interaction with the receptor. Direct and indirect combinations are made between clustered positions. First, r^2 (regression coefficient for all ligands) values are calculated for the first three atoms and for the best of CS. It is then debatable which of the clustered positions should be chosen as the fourth argument. For this purpose, each position is used as a candidate and is designated as the fourth position leading to the greatest increase in r^2. Similarly, the fifth, sixth and subsequent interaction points with the best r^2 values were included in the model. In this study, the points of interaction consisting of 22 positions with the letters a, b, c are given in Figure 2.

Based on MCET descriptors, a 4D-QSAR model was proposed to explain and predict the effects of substituents on the metastatic activity of 41 compounds on hydrophobic, electrostatic, steric, donor and receptor domains. The predicted activities of the molecules in the training and test set were shown with statistical values, and the accuracy of the established model was proven to be a measure of these values. A model was confirmed with statistical results for 41 molecules using Klopman index values ($q^2 = 0.998$, $r^2 = 0.993$). The predicted strength of the model found by the MCET method is confirmed in an external test set and is shown in Table 3 with the molecules marked "*". During modeling, the compounds in the training set were used outside the test set. The 3D model, validated by the activities of the ligands in the internal test set, was used to estimate the ligands in the external test set. The graph values between the estimated and observed activities using the model for both sets are shown in Figure 3.
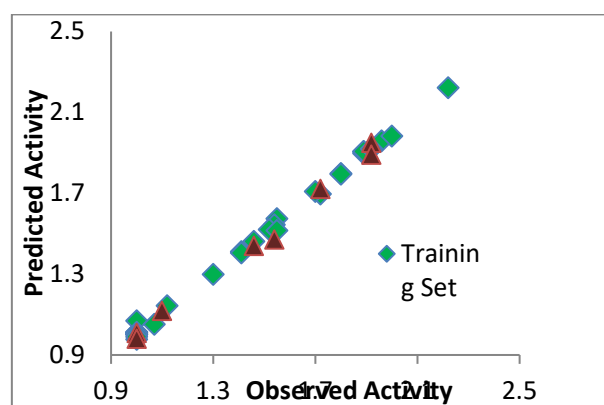
Figure 3. Observed and predicted activity graph of training and test sets.

3.2. Molecular docking results

Molecular docking was performed to investigate the stability of the previously formed 4D-QSAR model and to the interaction between the molecule (n30) and the TMPRSS4 enzyme (PDB ID: 5CE1) and was shown in Figure 4.

Figure 4. (a) Interaction between receptor-ligand. (b) Interaction diagram of the selected conformation selected for the highest activity molecule in the active site of the TRSSP enzyme.

4. CONCLUSIONS

4D-QSAR was used to investigate the structure-activity relationship of new 2-hydroxydiarylamide derivatives as a metastasis inhibitor. 4D-QSAR calculations have been developed by using a receptor-affected and representing others among multiple conformers. The Klopman index, which was first used as a new LRD descriptor in QSAR calculations, could be applied in the MCET method as a simplified formula. This new descriptor shows higher statistical results than other typical LRD descriptors. This descriptor, which includes both atomic charges and atomic coefficients in boundary orbitals, can be safely used in L-R interactions. Not only statistically high results are enough, but also exactly confirmed by docking. Molecular docking was performed according to the serine protease enzyme. Primarily from the protein data bank (PDB: 5CE1) TMPRSS4, the proteins acting on were tested one by one starting with the lowest RMSD value. The binding affinity with the FlexX docking program was found to be $\Delta G = -11.42$ kcal/mol.

ACKNOWLEDGEMENTS

Conflict of interests

Authors declare that there is no a conflict of interest with any person, institute, company, etc.

REFERENCES

1. World Health Organization. WHO "Cancer" February 2006 Retrieved on 2009-02-12

2. Kang, S.; Min, H. J.; Kang, M. S.; Jung, M. G.; Kim, S. *Bioorg. Med. Chem. Lett.* **2013,** *23*, 1748-1751.

3. Min, H. J.; Lee, M. K.; Lee, J. W.; Kim, S. *Biochem. Biophys. Res. Commun.* **2014,** *446*, 1-7.

4. Hooper, J. D.; Clements, J. A.; Quigley, J. P.; Antalis, T. M. *J. Biol. Chem.* **2001,** *276*, 857-860.

5. Netzel-Arnett, S.; Hooper, J. D.; Szabo, R.; Madison, E. L.; Quigley, J. P.; Bugge, T. H.; Antalis, A. M. *Cancer Metastasis Rev.* **2003,** *22*, 237-258.

6. Szabo, R.; Wu, Q. Y.; Dickson, R. B.; Netzel-Arnett, S.; Antalis, T. M.; Bugge, T. H. *Thromb. Haemost.* **2003,** *90*, 185-193.

7. Szabo, R.; Bugge, T. H. *Int. J. Biochem. Cell Biol.* **2008,** *40*, 1297-1316.

8. Kim, S.; Kang, H. Y.; Nam, E. H.; Choi, M. S.; Zhao, X. F.; Hong, C. S.; Lee, J. W.; Lee, J. H.; Park, Y. K. *Carcinogenesis* **2010,** *31*, 597-606.

9. Larzabal, L.; Nguewa, P. A.; Pio, R.; Blanco, D.; Sanchez, B.; Rodriguez, M. J.; Pajares, M. J.; Catena, R.; Montuenga, L. M.; Calvo, A. *Br. J. Cancer.* **2011,** *105*, 1608-1614.

10. Jung, H.; Lee, K. P.; Park, S. J.; Park, J. H.; Jang, Y. S.; Choi, S. Y.; Jung, J. G.; Jo, K.; Park, D. Y.; Yoon, J. H.; Park, J. H.; Lim, D. S.; Hong, G. R.; Choi, C.; Park, Y. K.; Lee, J. W.; Hong, H. J.; Kim, S.; Park, Y. W. *Oncogene.* **2008,** *27*, 2635-2647.

11. Ou-Yang, S. S.; Lu, J. Y.; Kong, X. Q.; Liang, Z. J.; Luo, C.; Jiang, H. L. *Acta Pharmacol. Sin.* **2012,** *33*, 1131-1140.

12. Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z. *Curr. Top. Med. Chem.* **2014,** *14*, 1923-1938.

13. Andrade, C. H.; Pasqualoto, K. F. M.; Ferreira, E. I.; Hopfinger, A. J. *Molecules* **2010,** *15*, 3281-3294.

14. Kizilcan, D. S.; Turkmenoglu, B.; Guzel, Y. *Struct Chem* **2020**, 31, 1339–1351

15. Alp Tokat, T., Türkmenoğlu, B., Kızılcan, D. Ş., Güzel, Y. *J. Mol. Model.* **2019,** *25*, 1-13.

16. Turkmenoglu, B.; Guzel, Y. *Comput. Biol. Chem.* **2018,** *76*, 327-337.

17. Guzel, Y.; Aslan, E.; Turkmenoglu, B.; Su, E. M. *Curr. Comput-Aid. Drug.* **2018,** *14*, 207-220.

18. Yilmaz, H.; Boz, M.; Turkmenoglu, B.; Guzel, Y. *Trop. J. Pharm. Res.* **2014,** *13*, 117-126.

19. Yilmaz, H.; Guzel, Y.; Onal, Z.; Altiparmak, G.; Kocakaya, S. O. *Bull. Korean Chem. Soc.* **2011,** *32*, 4352-4360.

20. Türkmenoğlu, B., Yılmaz, H., Su, E.M., Alp Tokat, T. Güzel, Y. *Int.J. Chem. Technol.* **2017,** *1*, 14-23.

21. Su, E. M., Turkmenoglu, B., Guzel, Y. Int. J. Innov. Stud. Sci. Eng. Technol. **2016,** *02*, 67-75.

22. Turkmenoglu, B.; Guzel, Y.; Su, E.M; Kizilcan, D. S. Mater. Today Commun. **2020,** *25*, 101583

23. Klopman, G. *J. Am. Chem. Soc.* **1968,** *90*, 223-243.

24. Salem, L. *J. Am. Chem. Soc.* **1968,** *90*, 543-552.

25. Salem, L. *J. Am. Chem. Soc* **1968,** *90*, 553-566.

26. Fujimoto, H., Fukui, K. *In Chemical Reactivity and Reaction Paths*, John Wiley&Sons: New York, 1974.

27. Jain, A. N. *J. Comput. Aid. Mol. Des.* **2007,** *21*, 281-306.

28. Spitzer, G. M.; Wellenzohn, B.; Laggner, C.; Langer, T.; Liedl, K. R. *J. Chem. Inf.* **2007,** *47*, 1580-1589.

29. Holt, P. A.; Chaires, J. B.; Trent, J. O. *J. Chem. Inf. Model.* **2008,** *48*, 1602-16158. Ghorbani, M.; Seyedin, O.; Aghamohammadhassan, M. *J. Environ. Manage.* **2020**, 254, 109814-109919.