

Molecular Docking Studies towards Development of Novel Gly-Phe Analogs for Potential Inhibition of Cathepsin C (dipeptidyl peptidase I)

U. Kalathiya^{1,2,*}, M. Padariya^{1,2}, M. Jewginski³, M. Baginski³

¹Biotechnology (Bioinformatics) Department, Wroclaw University of Technology, Poland

²Department of Pharmaceutical Technology and Biochemistry, Gdansk University of Technology, Poland

³Department Bioorganic Chemistry, Wroclaw University of Technology, Poland

Article Info

Article history:

Received Nov 13th, 2013

Revised April 8th, 2014

Accepted April 8th, 2014

Keyword:

Dipeptidyl Peptidase I
Cathepsin C
Molecular Modeling
Scoring function
Force Field

ABSTRACT

Cathepsin C is a cysteine protease required for activation of various pro-inflammatory serine proteases and, essentially, is of interest as a therapeutic target. Cathepsin C coordinate system was employed as a model to study the interaction of some already available inhibitors of Cathepsin C. Compounds containing Gly-Phe fragment with functional groups at its ends were designed by knowledge based approach. Using AutoDock and Discovery Studio Client 3.1 software packages, binding energy of different conformations and ten scoring functions (LigScore1, LigScore2, PLP1, PLP2, JAIN, PMF, PMF04, LUDI_1, LUDI_2 and LUDI_3) were calculated for newly designed compounds. These docking studies revealed favorable energy scores which also helps to understand interaction of ligands with enzyme.

Copyright © 2014 *International Journal for Computational Biology*,
<http://www.ijcb.in>, All rights reserved.

Corresponding Author:

Umesh Kalathiya,
Department of Pharmaceutical
Technology and Biochemistry, Gdansk
University of Technology, Poland
Email: umekalat@gmail.com



How to Cite:

U. Kalathiya *et. al.* Molecular docking studies towards development of novel Gly-Phe analogs for potential inhibition of Cathepsin C (dipeptidyl peptidase I). *IJCB*. 2014; Volume 3 (Issue 1): 03-26.

1. INTRODUCTION

Cathepsin C (DPPI or dipeptidyl peptidase I; EC 3.4.14.1) is a lysosomal cysteine protease, which sequentially removes dipeptides from the N-termini of protein and peptide substrates. DPPI homologues have been identified from gene cloning, biochemical characterization or sequence comparison in various species [1] [2] [3] [4]. Continuously growing evidence of the key role of hDPPI (human DPPI) in various diseases, for instance sepsis [5], arthritis [6] and other inflammatory disorders, applying animal models has drawn attention to the potential of DPPI as a drug target.

First time Cathepsin C was described by Guttman and Frutin in 1948 [7] [8], but the cDNA of human enzyme was published in 1995 [9]. The cDNAs encoding rat, human, murine, bovine, dog and two schistosome Cathepsin C have been cloned and sequenced which shows that Cathepsin C enzyme is highly conserved [10]. The human and rat Cathepsin C cDNAs encode precursors (prepro-cathepsin C) consisting signal peptides of 24 residues, pro-region of 205 (rat Cathepsin C) or 206 (human Cathepsin C) residues and are 30-40% identical to the mature amino acid sequences of papain and a number of other cathepsins involving cathepsins, B, H, K, L, and S [11].

Cathepsin C is the only member of C1 family that needed a tetrameric form and requires halide ions for its activity [9]. It can also activate zymogens of several proteases [12] [13] [14] [15] [16]. DPPI activates several pro-inflammatory serine proteases by removal of an aminoterminal inhibitory dipeptide [17]. DPPI in acidic lysosomal milieu is primarily an amino dipeptidase cleaving two-residue units from the N-terminal of a polypeptide chain, although it can also acts as a transferase and catalyze the reverse reaction [9] [14].

Although it is processing enzyme of many proteases, DPPI is not capable of auto activation and requires other proteases such as Cathepsin L or S [9]. DPPI is distributed in a variety of tissues and it is expressed in lung, kidney, liver and spleen [7]. It is widespread in mammals (humans and other apes, cows dogs, rabbits, rats and mice), human parasites (*Plasmodium* and *Schistosoma*), fish (rainbow trout and killifish), reptiles (frog) and birds (Chicken) [18] [19] [20] [21] [22], that indicates an important role widespread of Cathepsin C in nature and is present in high levels in cytotoxic lymphocytes, and mature myeloid cells [23]. Cathepsin C has two distinct roles: First, it contributes to general protein catabolism within the lysosome. Second, the enzyme is found in secretory granules of immune effector cells (cytotoxic lymphocytes, mast cells and neutrophils) where it activates granular serine proteases by removing an inhibitory N-terminal dipeptide sequence [24] [25] [26]. DPPI also functions as a key enzymes in activation of granule serine protease in cytotoxic T lymphocytes and natural killer cells (granzymes A and B) [27] [28], mast cells (chymase and tryptase) [29] [30] [31] [32] and neutrophils (Cathepsin G, proteinase 3 and elastase). DPPI and granzymes are found in granules of cytotoxic lymphocytes [23] and DPPI is responsible for processing and activating progranzymes [15].

DPPI is unique proteases within the papain super family, as it has an oligomeric structure; indeed it has a unique structure. It also has a unique mechanism compared with other oligomeric proteolytic complexes such as proteasome, bleomycin hydrolase and tryptase, which all have their active site located inside the structure. Active site of DPPI is located on outside of the structure, which has the capability of binding four molecules of cystatin per oligomer [9].

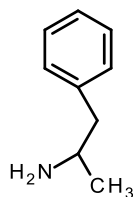
Human DPPI is 200 kDa (Figure 1) which consists four identical subunits and each subunit is composed from three different polypeptide chains: a heavy chain, a light chain, and an exclusion domain. N-terminal fragment also known as residual propeptide domain was suggested to be involved in formation of tetramer [9]. The interface of heavy chain and light chain mainly consists of hydrophobic interactions. DPPI has four active site clefts. The catalytic residues Cys 234 and His 381 are situated above the oxyanion hole [15].

The four active site clefts are positioned approximately at tetrahedral corner of the molecule, ~50-60 Å apart and are exposed to solvent. Each active site (Figure 2) cleft is formed by feature of all three domains of a functional monomer of DPPI; the papain-like domains form the sides of monomer, which is closed at one end by exclusion domain [9].

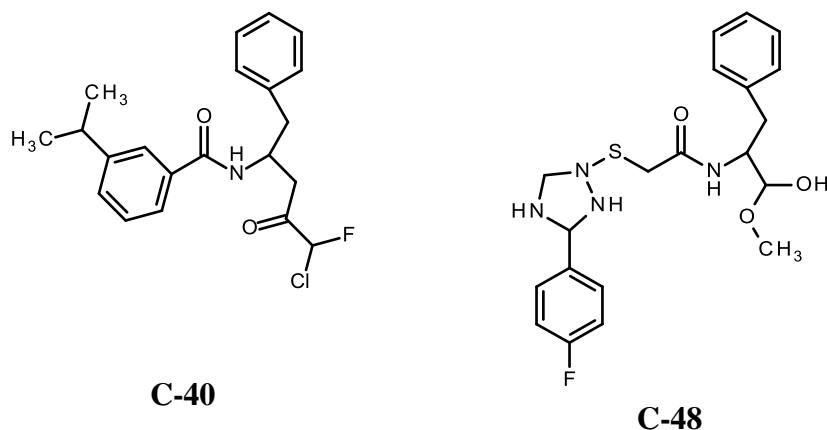
Some available inhibitors of Cathepsin C, which are considered for designing new compounds are: Dipeptide Diazomethyl Ketenes [33] [34] [35], A Dipeptide Nitrile [36] [37], Dipeptide Vinyl Sulfones [38], Dipeptide Acyloxy-Methyl And Fluoromethyl-Ketones [38], Dipeptide O-Acyl Hydroxamic Acid [39], Arginine-Based Peptides [40], Cyanamide-Based Inhibitors [13] and Phoshinic Tripeptides [41]. Natural DPPI inhibitor is cystatins or E-64 (Trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane) [33] [42].

Gly-Phe-CHN₂ is one of the most efficient inhibitor for suppressing activity of endogenous Cathepsin C enzyme [15], it binds to S1 and S2 binding sites and it is covalently linked to Cys 234 residue, making a thioether bond. All hydrogen bond donors and acceptors from the inhibitor form hydrogen bonds with the enzyme and forms strongest interaction with Asp [33].

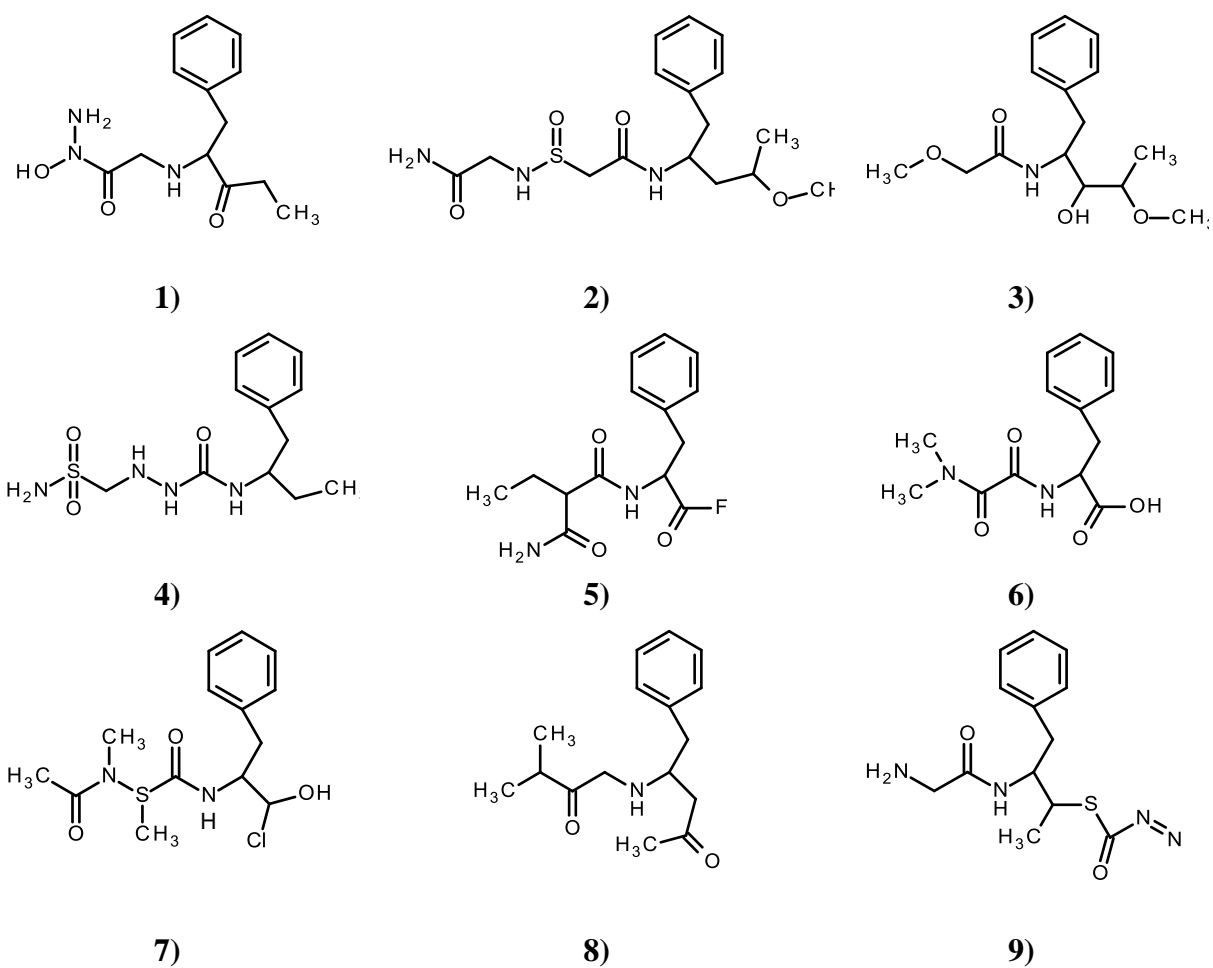
Studying active site of Cathepsin C and activity of some already available inhibitor, different compounds were designed by knowledge based approach, containing Gly-Phe fragment with functional groups at its ends. Using AutoDock and Discovery Studio Client 3.1 software packages, binding energy of different conformations and ten scoring functions (LigScore1, LigScore2, PLP1, PLP2, JAIN, PMF, PMF04, LUDI_1, LUDI_2 and LUDI_3) were calculated for newly designed compounds, which helps to understand interaction of ligands with enzyme.

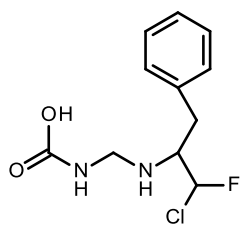


Scheme 1 1-phenylpropane-2-amine (Gly-Phe) fragments



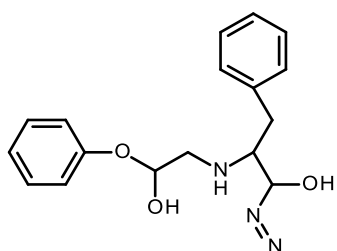
Scheme 2 Chemical structure of compound C-40 and C-48



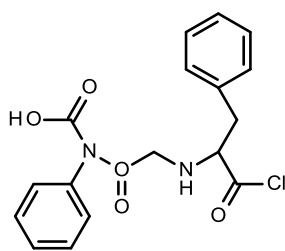


10)

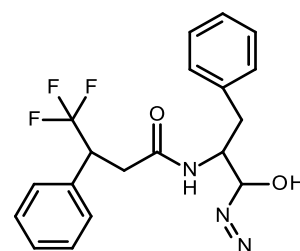
Scheme 3 Compounds containing 1-phenylpropane-2-amine (Gly-Phe) fragment with non cyclic derivatives



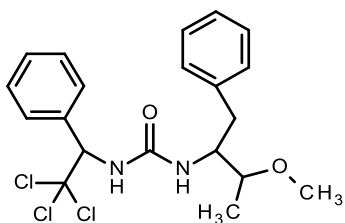
11)



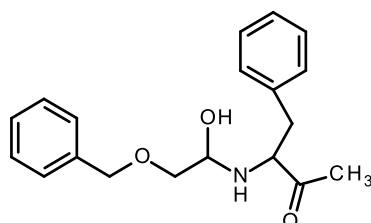
12)



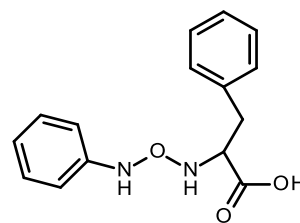
13)



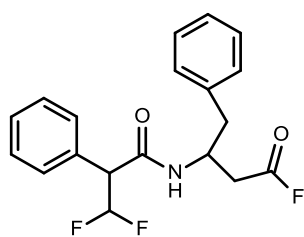
14)



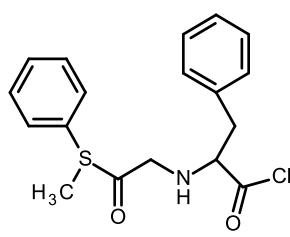
15)



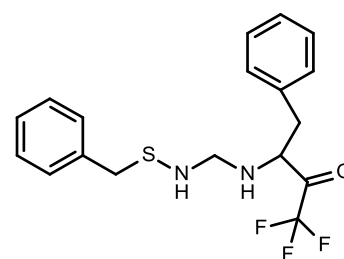
16)



17)

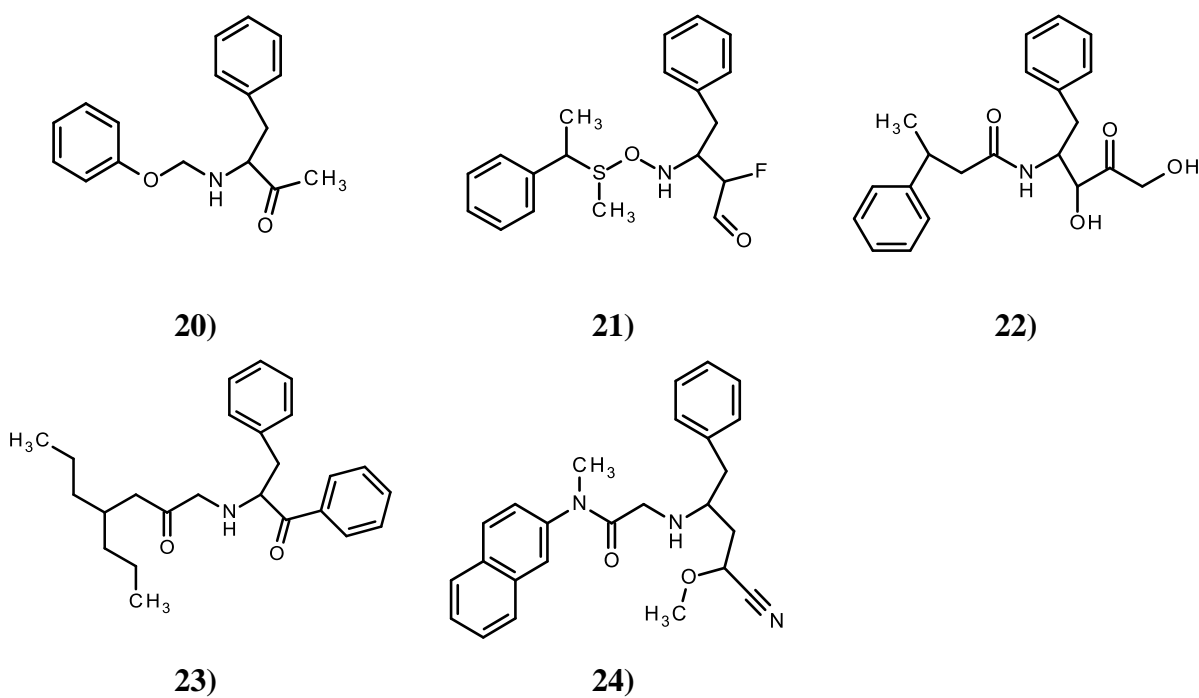


18)

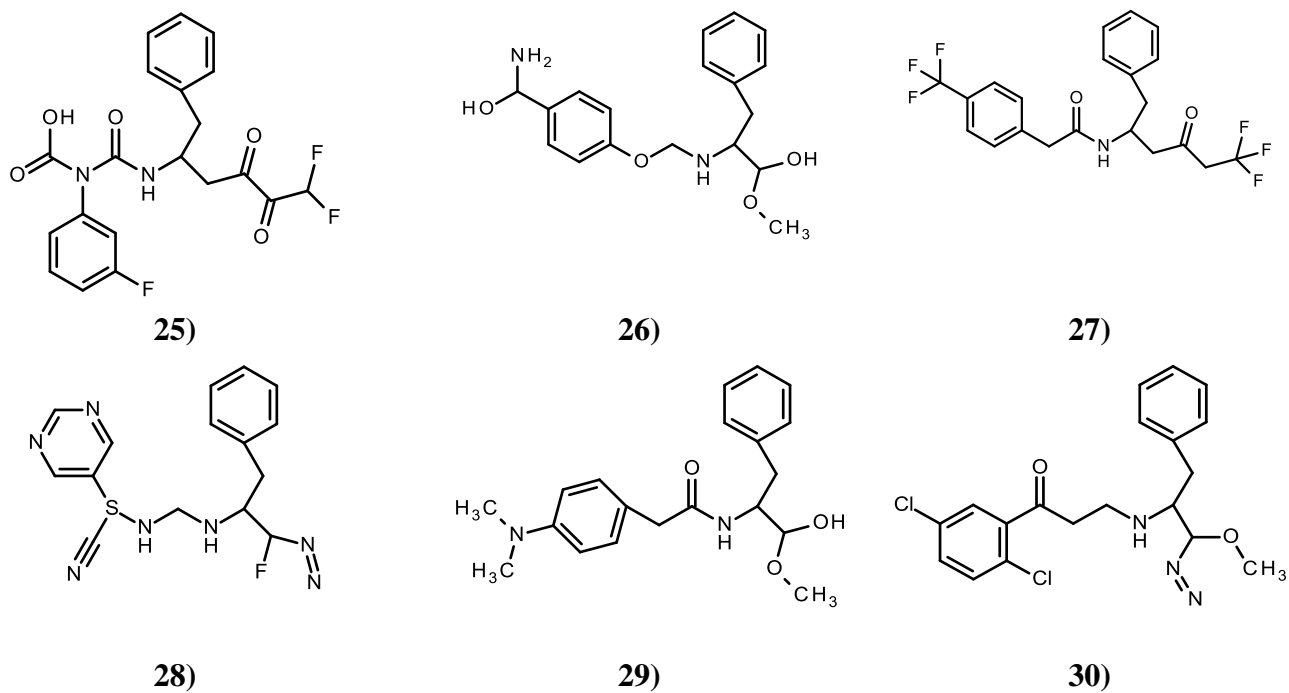


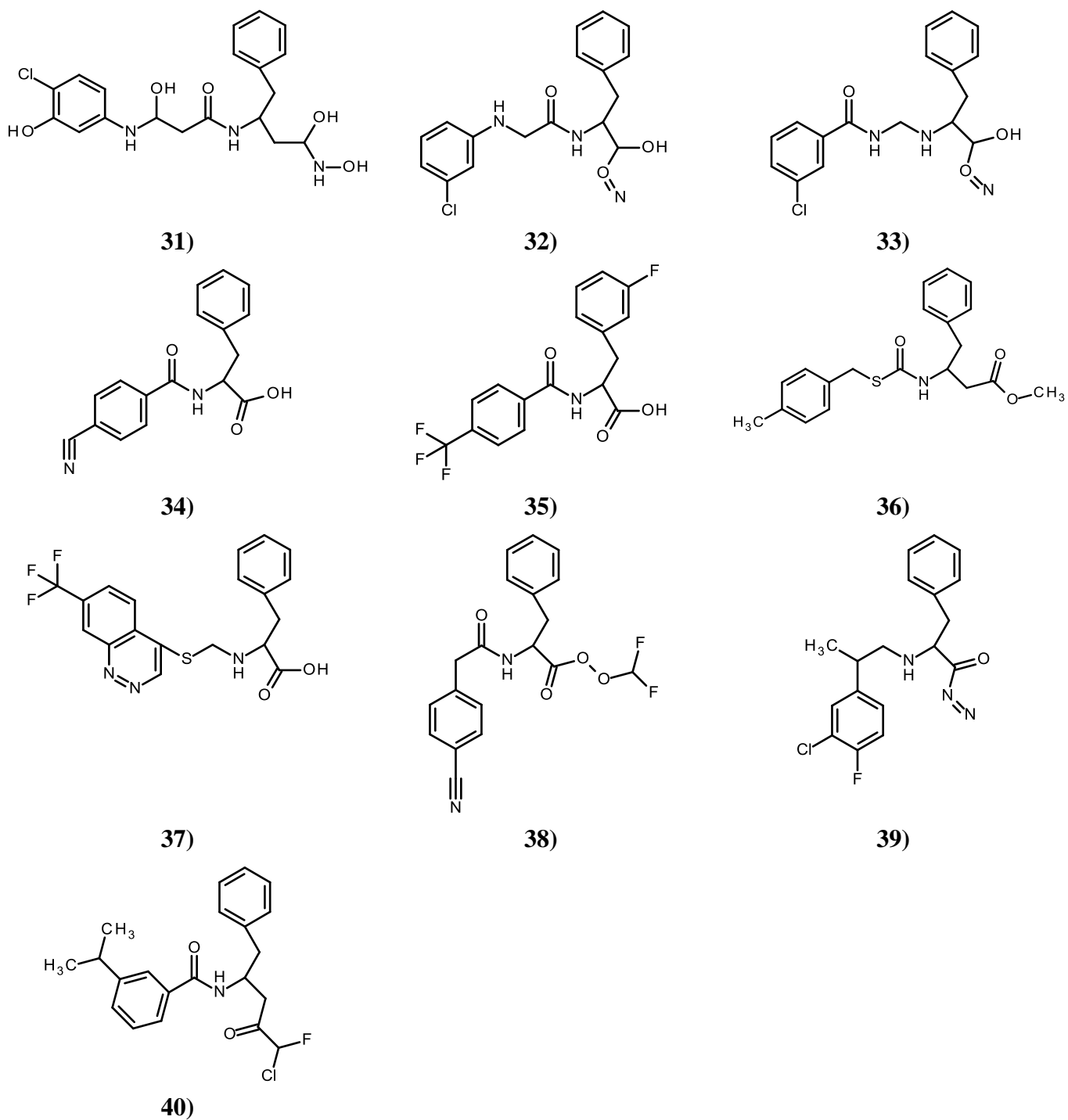
19)





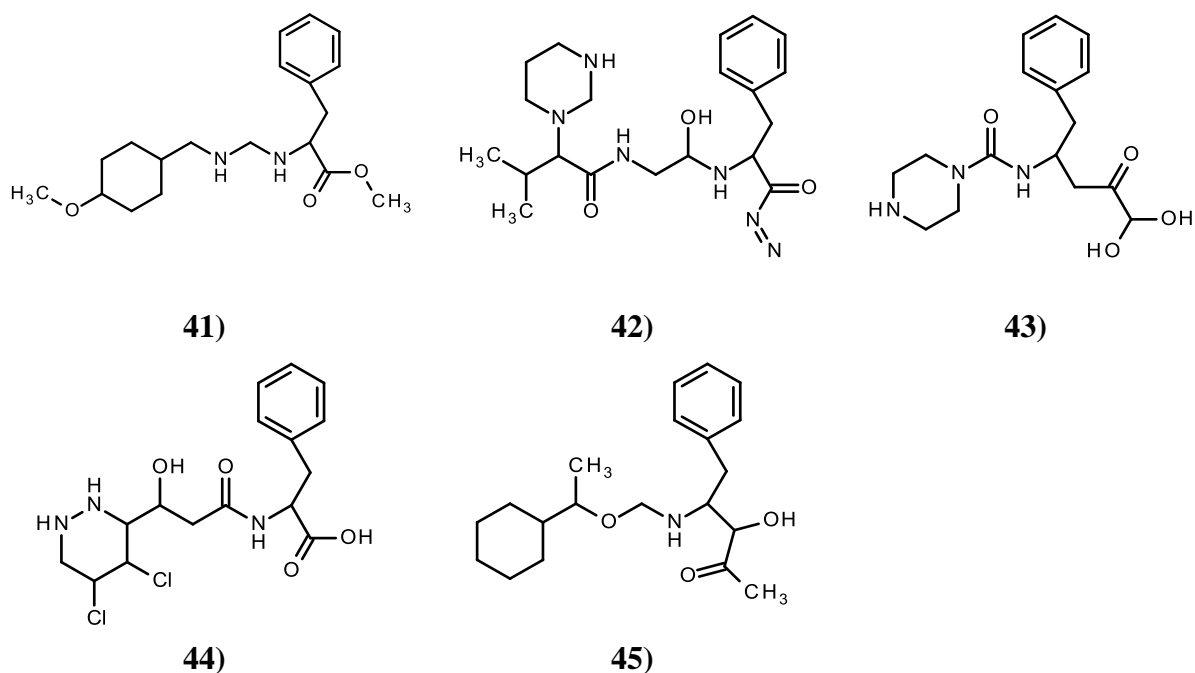
Scheme 4 Compounds containing 1-phenylpropane-2-amine (Gly-Phe) fragment with benzene



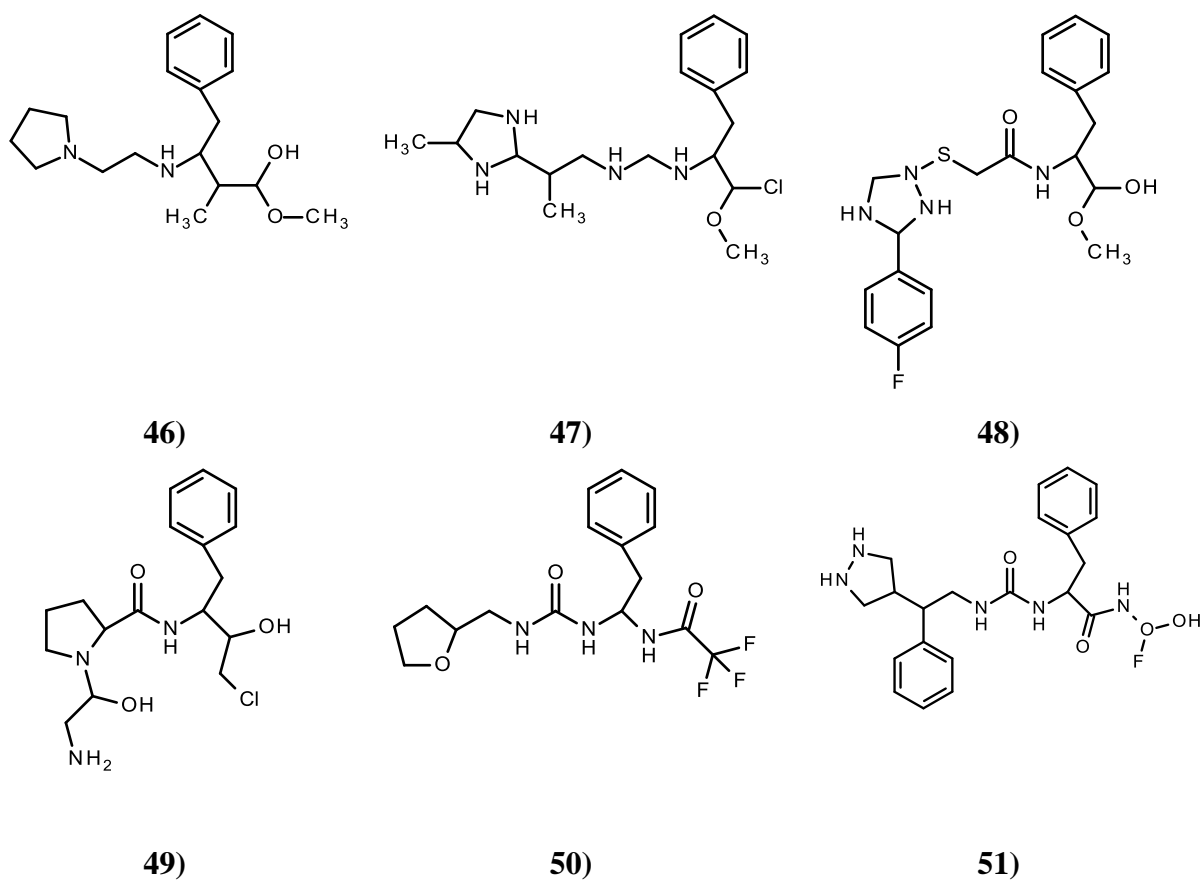


Scheme 5 Compounds containing 1-phenylpropane-2-amine (Gly-Phe) fragment with benzene modifications

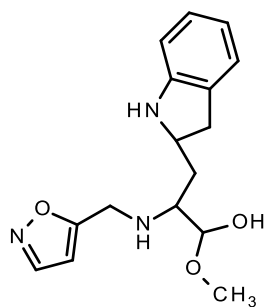




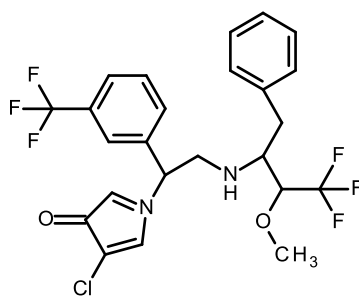
Scheme 6 Compounds containing 1-phenylpropane-2-amine (Gly-Phe) fragment with saturated 6-member ring



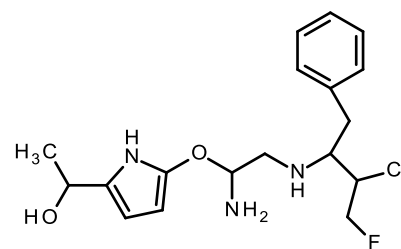
Scheme 7 Compounds containing 1-phenylpropane-2-amine (Gly-Phe) fragment with saturated 5-member ring.



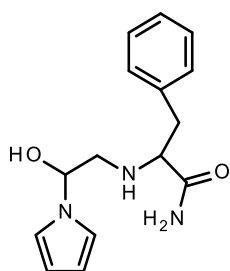
52)



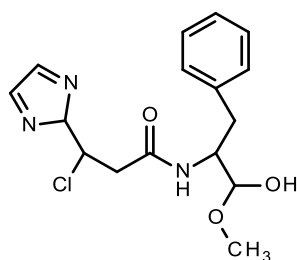
53)



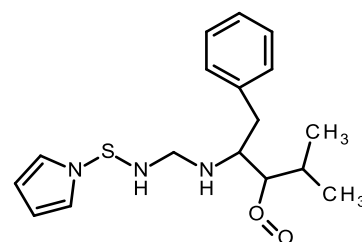
54)



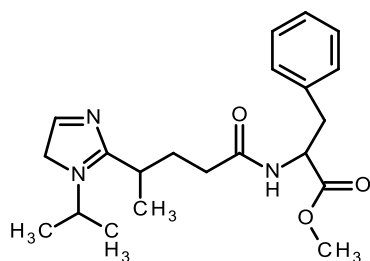
55)



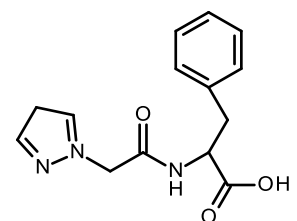
56)



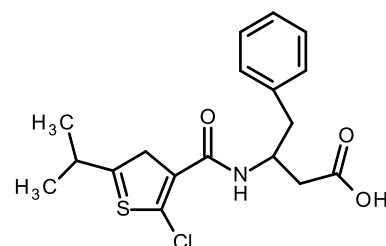
57)



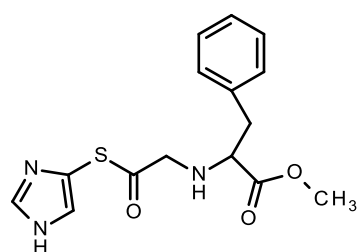
58)



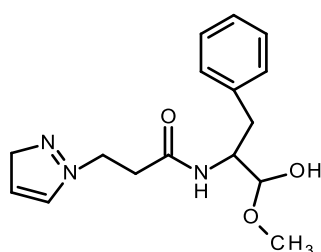
59)



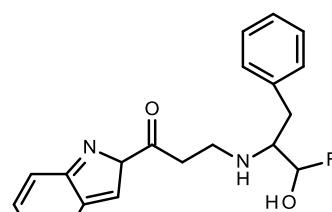
60)



61)

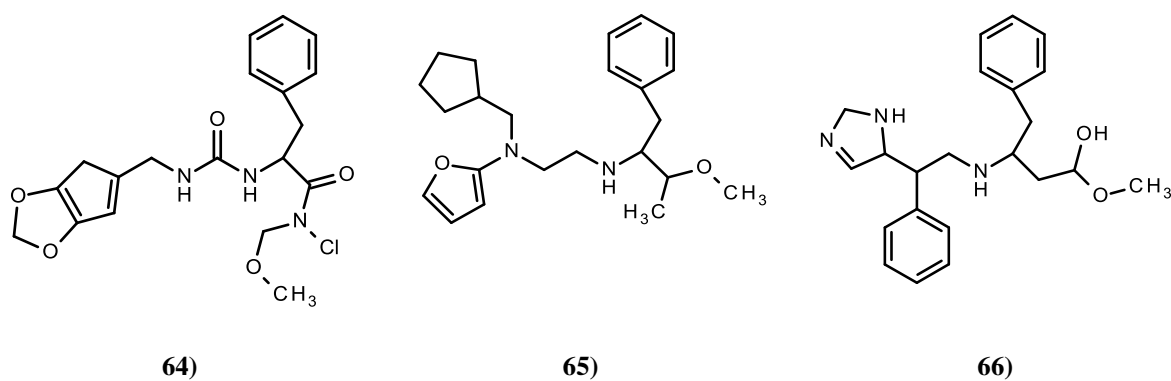


62)



63)





Scheme 8 Compounds containing 1-phenylpropane-2-amine (Gly-Phe) fragment with unsaturated 5-member ring.

In all the potential inhibitors of Cathepsin C, it has been observed that most of them contain 1-phenylpropane-2-amine (Gly-Phe) fragments (Scheme 1) in their structure. Studying the activity of Gly-Phe-CHN2 inhibitor, it has been seen that this fragment penetrates into deep pocket formed by active site; interaction with enzyme is formed with both ends of inhibitors.

To design more effective inhibitor for Cathepsin C, peptides were added at ends of 1-phenylpropane-2-amine fragments, one end is kept short as there is not much space between ligand and receptor, when Gly-Phe penetrates into the cavity and other end is kept long. Various compounds were designed in this way and different functional groups were added to them considering the active site residues of the enzyme. Designed compounds were docked individually and their interactions with active site residues of enzyme were studied and considered for designing further new inhibitor. Almost all newly designed compounds contains 1-phenylpropane-2-amine (Gly-Phe) fragment with: non cyclic derivatives (Scheme 3), benzene (Scheme 4), benzene with modifications (Scheme 5), saturated 6-member ring (Scheme 6), saturated 5-member ring (Scheme 7) and unsaturated 5-member ring (Scheme 8).

Computing interactions of all compounds with enzyme by using Autodock and Discovery Studio Client 3.1 software, it has been observed that compounds with non cyclic derivatives shows less interaction with enzyme as compared to compounds of other groups. Benzene with modifications and unsaturated 5-member ring compounds show quite strong interactions with enzyme and steady conformations in comparison to other compounds.

1.2 Knowledge Based Approach

Knowledge based design is one of the well known method for prediction of new small molecule or ligands, which can be more effective than already available molecules [43] [44]. It is also found that knowledge based designed structure have the same biological activities, as they can exert related effects similar to available ligands [45].

Designed new molecules with knowledge based are having some small modifications in structure, but this can change the functionality of molecule and also affects its molecular property [46]. Modifications can be done by adding and replacing peripheral functional groups, in this way knowledge based method helps to design molecules that are easier to synthesize and more efficient. The nature of binding pocket of the target has a considerable impact on designing of new small molecules. It is useful to understand the type of interaction, for instance, lipophilic or hydrophilic interaction [47].

There are several methods proposed to measure ligand-binding efficiency [48] [49] that helps to evaluate how well the functional groups or heavy atoms are used in binding. It is found that less-complex ligands have a better efficiency of binding to target [50] [51]. Another element of ligand quality is structural complexity of the ligand, especially the features that are associated with its scaffold. It has been shown that 50% of known drugs share only 32 molecular frameworks [52], implying an optimal level of skeletal complexity is an intrinsic character of drug likeness. Recent analysis revealed that a certain level of molecular complexity is necessary to achieve desired biological activity [53]. In this work knowledge based approach, was considered for designing and modeling new inhibitors which could be the potential target for inhibiting the activity of the Cathepsin C.

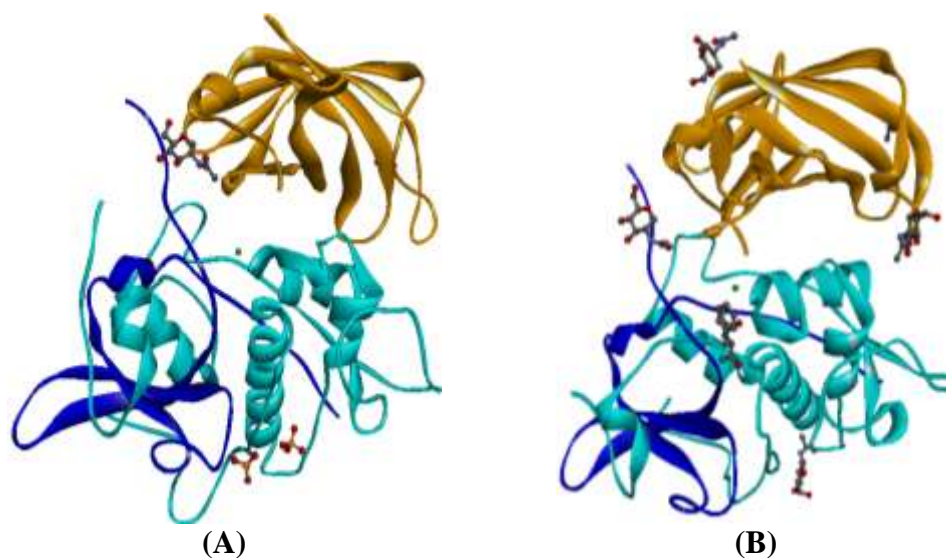


Fig. 1 (A) Crystal structure of Human Dipeptidyl Peptidase I (Cathepsin C): Exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases (1K3B) [9]. (B) Crystal structure of Human Dipeptidyl Peptidase I (Cathepsin C) in complex with inhibitor Gly-Phe-CHN₂ (2DJF) [33]. *Color Scheme:* Exclusion domain (A chain) is represented in orange, heavy B chain is colored in light blue and light C chain in blue. A chlorine atom is represented in green in center.

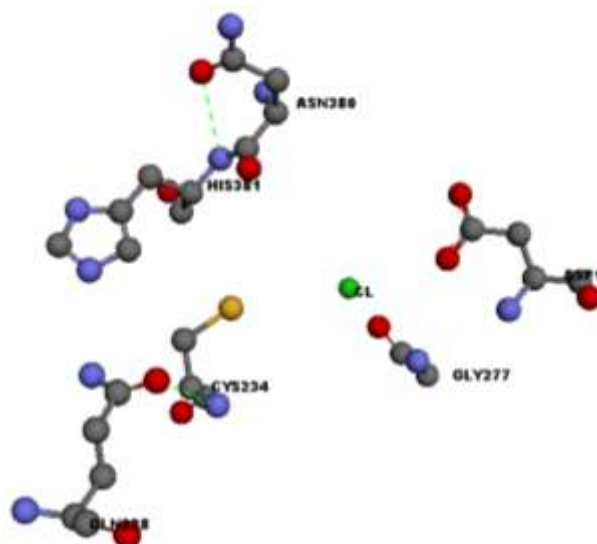


Fig. 2 Active site with hydrogen bonds, amino acid residues that are present in active site of Cathepsin C enzyme are: Asp1, Gln 228, Cys 234, Gly 277, Asn 380 and His 381 and they take part in catalytic mechanism or substrate binding [9] [33]. *Color Scheme:* Carbon atom are colored in grey, oxygen atom in red, nitrogen atom in blue, sulfur atom in orange, chlorine atom in green and hydrogen bonds are shown in green dashed lines.

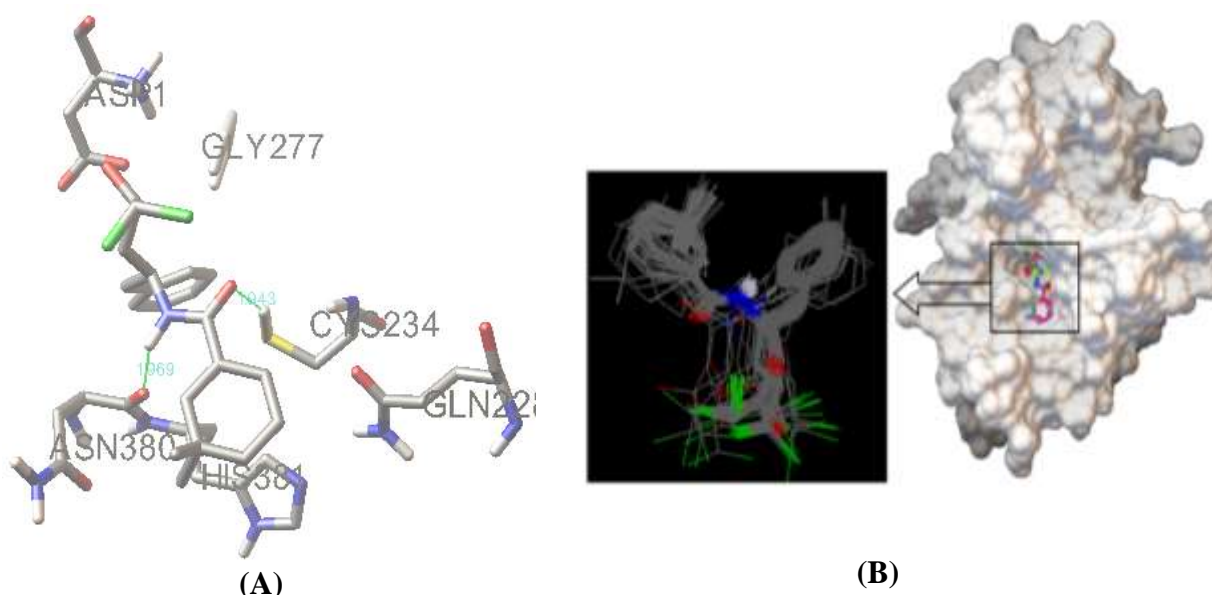


Fig. 3 AutoDock result of compound **C-40**. (A) Modeled structure of compound **C-40** along with active site of Cathepsin C (1K3B) [12] showing interactions with active site residues. (B) 104 conformations generated when compound **C-40** interacts with Cathepsin C enzyme, which demonstrate the steady state of the ligand interaction with the receptor.

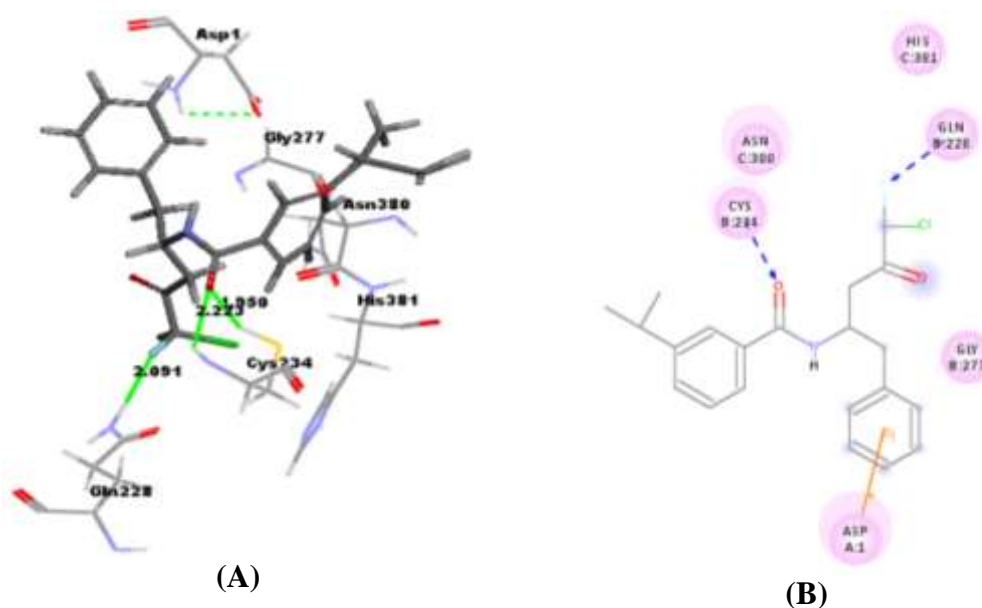


Fig. 4 Discovery Studio Client 3.1 result of compound **C-40**. (A) Computed structure of designed compound **C-40** showing hydrogen bond interactions with active site residues of Cathepsin C. (B) 2D diagram of the compound **C-40** with active site residues, showing with different types of interactions.

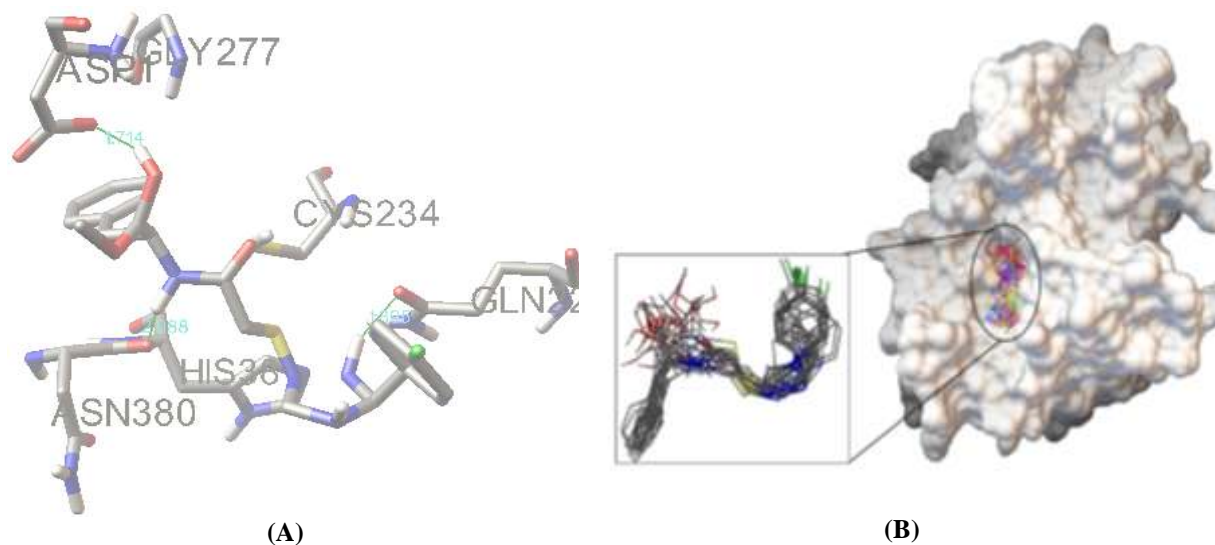


Fig. 5 AutoDock result of compound **C-48**. **(A)** Computed structure of compound **C-48** along with active site of Cathepsin C (1K3B) [12] showing interactions with active site residues. **(B)** Different conformations generated when ligand interacts with receptor, which explains the steady state of the ligand when it interacts with the enzyme.

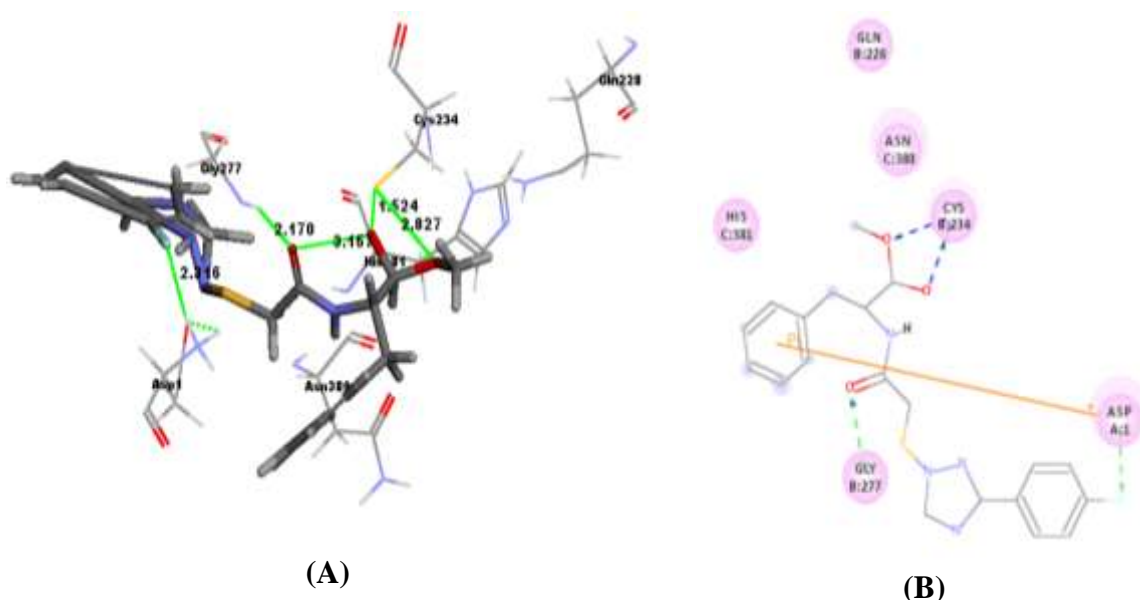


Fig. 6 Discovery Studio Client 3.1 result of compound **C-48**. **(A)** Optimized structure of compound **C-48** showing hydrogen bond interactions with active site residues of Cathepsin C. **(B)** 2D diagram of the compound **C-48**, with active site residues, showing with different type of interactions.

Along with, the Ni ions are bridged by a hydroxide ion (WB), which along with two terminal water molecules, W1 on Ni(1), W2 on Ni(2), and W3 located near the opening of the active site, which forms an H-



bonded water tetrahedral cluster filling the active site cavity. It is the cluster (the carboxylate group of the carbamylated lysine and the hydroxide molecule) that urea replaces when binding to the active site for the reaction [11]. As a consequence of above ligations, Ni(1) is pentacoordinated and Ni(2) hexacoordinated, and their coordination geometry is pseudo square pyramidal and pseudo octahedral, respectively. In another consideration, urease can severely decrease the efficiency of urea fertilizers to cause the release of large amounts of ammonia and further induce plant damage by ammonia toxicity and soil pH increase [11]. So, to control the rate of the enzymatic urea hydrolysis using urease inhibitors is an important goal. Large quantities of urea produced as a result of biological process. Each human being produces approximately 10 kg of urea per year. Spontaneous degradation of urea occurs with a half life of approximately 3.6 years [1], but in the presence of urease, the hydrolysis of urea is 104 times faster [7].

2. RESEARCH METHOD

For calculating interaction energy and different scoring functions of newly designed compounds two different programs were used: AutoDock and Discovery Studio Client 3.1, these software packages follows different algorithms for calculating receptor-ligand interactions.

2.1 Autodock package

AutoDock uses simulating annealing docking algorithms, a physical process for calculation of interaction energy between ligand and receptor. In AutoDock, ligand is treated as flexible and protein is treated as rigid, that states ligand will interact with receptor on its outer surface [54]. AutoDock docking procedure is divided into three steps related with tree programs: AutoTors, AutoGrid and AutoDock. The simulation begins at high temperature, accepting nearly every move and ligand explores large areas of conformation space. As the temperature is reduced, unfavorable moves are increasingly disallowed. The ligand ultimately finds an optimal position and conformation within the deepest energetic well it has sampled, by performing multiple separate simulations, starting from random initial states, consistent favorable binding conformations may be located. Intermolecular interaction energies are calculated at each time step, including dispersion/repulsion, hydrogen bonding and electrostatic terms, replacing the hard-sphere potential used in the original release [55] [56].

2.2 Accelrys Discovery Studio Client 3.1 package

Initially the structures of compounds were drawn using Accelrys Discovery Studio Client 3.1, force field CHARMM, Partial Charge-Momany Rone were applied [57][58] and it was also used to optimize receptor-ligand complex and to calculate the docking score of each compound. Missing bond orders, charges and angles were assigned to each ligand.

LigScore1, LigScore2, PLP1, PLP2, Jain, PMF, PMF04, Ludi_1, Ludi_2 and Ludi_3 are different scoring functions, which were calculated for newly designed compounds by Discovery Studio Client 3.1. Among these scoring functions, PMF and PMF04 are knowledge based functions and rest all empirical scoring functions [59]. Different scoring functions are grouped according to the methodologies used to derive them, these three main approaches are generally termed force field based, knowledge based, and empirical [59].

1. Force field scoring functions mainly account for the binding energy between ligand and protein by evaluation of non-bonded interactions, electrostatic and van der Waals potentials between protein and ligand atoms [59] [60] [61] [62].

2. Knowledge-based potentials are derived by collecting statistics on interatomic distances from a database of protein–ligand structures in order to determine potentials of mean force based on the probability of observing specific types of protein and ligand atoms at a given separation from each other [59] [63] [64] [65].

3. Empirical scoring Functions are a set that defines various aspects of receptor-ligand interaction energies like hydrogen bonding, steric interactions, lipophilic interactions, solvation and entropic effects. This states that empirical scoring function is related to force field function [59] [66] [67] [68] [69] [70].

3. RESULTS AND ANALYSIS

Newly designed compounds contains 1-phenylpropane-2-amine (Gly-Phe) fragment with: Non cyclic derivatives, Benzene, Benzene with modifications, saturated 6-member ring, saturated 5-member ring, unsaturated 5-member ring. Methodologies which were going to be applied in all the newly designed compounds were tested with one of already known inhibitor. As Gly-Phe-CHN2 is stated as one of the potential inhibitor [15] [33], it was selected for testing. The results which were obtained from both methodologies were compared with the data available for Gly-Phe-CHN2 inhibitor. AutoDock result of Gly-Phe-CHN2; lowest binding energy is -7.06 Kcal/mol, mean binding energy is -6.47 Kcal/mol. (Table 1)

Table 1 Discovery Studio Client 3.1 results of scoring functions of Gly-Phe-CHN₂

LigScore ₁	LigScore ₂	-PLP1	-PLP2	Jain	-PMF	-PMF 04	Ludi_1	Ludi_2	Ludi_3
-3.87	-7.9	34.13	24.4	2.58	27.3	7.99	274	295	279

Comparing interaction results of Gly-Phe-CHN₂ with enzyme obtained by AutoDock and Discovery Studio Client 3.1, it has been found that inhibitor shows interaction with Asp 1 residues of enzyme, which indicates that both different methods, supports each other. From these two methods, it has been obtained that inhibitor interacts with Asp 1, Asn 380 and Gly 277 residues of enzyme. By comparing results obtained from two methods with already available data [33], it is observed that both results correlate each other. These states that methodologies which are going to be applied to newly designed compounds are correct and they give reliable results of receptor-ligand interactions.

AutoDock results are described in terms of lowest binding energy, mean binding energy and total conformations in cluster. During evaluation of the results these three terms were taken in consideration. AutoDock software automatic ranks cluster according to their lowest binding energy, but values given in tables for each compounds may contain one of the best clusters, mean binding energy and lowest binding values. Cluster difference of less than 10 conformations was neglected, while selecting lowest energy. In some cases, selection of best cluster, mean binding and lowest energy was difficult, so two or three different conformations were selected for a single compound.

Discovery Studio Client 3.1 was used for calculation of different scoring functions: knowledge based scoring function like PMF and empirical scoring functions like LigScore, PLP, Jain and Ludi. All different scoring functions result was analyzed separately: LigScore, PLP and PMF score with lowest value were considered well, Jain and Ludi score with highest values were better. Results obtained states that scoring functions like LigScore1 and LigScore2 do not have much difference in their values and Ludi 1, 2 and 3 score values are almost similar with each other.

From all newly designed two compounds were considered as the appropriate potential targets for the Cathepsin C, they are C-40 and C-48. Their results are described below, describing two different methodologies. (Table 2)

Table 2 Scoring functions and AutoDock results of designed inhibitor

Com No.	Lowest Binding Energy	Mean Binding Energy	Total in Cluster	LigScore ₁	LigScore ₂	-PLP1	-PLP2	Jain	-PMF	-PMF 04	Ludi_1	Ludi_2	Ludi_3
C-1	-6.72	-5.62	23	-2.15	-5.05	35.07	29.68	0.83	7.61	-8.99	178	231	208
C-2	-7.11	-6.11	18	4.06	2.98	64.88	50.84	3.38	32.44	7.74	324	324	291

	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10	C-11	C-12	C-13
	-5.66	-5.34	-5.83	-5.71	-5.73	-6.24	-7.12	-6.61	-5.96	-6.43	-6.07
	-4.83	-4.81	-5.28	-5.09	-5.39	-5.67	-6.1	-6.04	-4.92	-6.04	-5.41
	79	64	54	70	88	77	57	64	25	15	31
	-1.46	-7.27	0.61	0.65	-1.26	-2.12	-11.2	-2.84	0.02	-6.8	-15.28
	-6.04	-16.45	-1.62	-1.95	-4.08	-5.47	-21.89	-3.25	-2.71	-12.53	-29.78
	21.4	-0.96	28.32	22.89	33.94	3.47	-36.38	42.99	30.37	-53.65	-118.78
	23.76	-11.79	31.37	19.11	37.45	13.58	-19.59	29.9	25.98	-22.28	-49.78
	4.95	4.84	2.05	4.59	5.44	5.16	5.61	-1.26	1.75	2.98	4.64
	11.92	8.23	-2.46	34.29	-16.29	-8.92	-46.46	15.35	15.78	10.22	-10.45
	-12.85	-14.49	-14.31	7.01	-27.03	-31.39	-59.89	5.62	-15.73	-30.51	-54.76
	483	653	154	409	245	302	410	183	213	339	342
	448	584	209	374	283	350	398	217	214	323	383
	437	563	188	355	267	325	366	192	257	410	419

	C-14	C-15	C-16	C-17	C-18	C-19	C-20	C-21	C-22	C-23	C-24
	6.18	-6.64	-6.65	-6.27	-7.4	-6.63	-6.35	-7.48	-7.99	-8.81	-8.83
	-5.71	-5.81	-5.82	-5.7	-6.5	-5.98	-5.93	-6.58	-6.8	-7.53	-7.65
	15	85	77	58	61	95	117	78	52	63	39
	-13.98	-2.08	-9.4	3.44	-5.24	1.76	-2.07	-10.14	-12.92	-12.09	-18.86
	-25.29	-2.37	-14.5	3.66	-10.97	1.5	-2.67	-16.38	-25.35	-22.67	-33.76
	-49.49	20.38	-39.18	34.23	-55.79	34.53	15.24	-52.22	-119.24	-68.95	-143.05
	-43.16	23.42	-19.57	53.77	-23.52	59.97	22.44	-23.32	-60.73	-24.97	-78.53
	2.6	1.26	1.76	1.67	4.29	3.5	2.25	4.17	8.24	8.8	5.78
	14.48	18.79	-9.71	3.33	6.77	-40.74	20.94	-7.85	-15.2	-40.62	-10.61
	-66.77	-6.94	-42.09	7.51	-35.54	-26.6	-11.36	-22.73	-69.22	-78.42	-83.13
	548	131	276	279	332	267	338	297	577	463	457
	466	224	314	336	370	329	343	342	515	512	443
	498	305	385	411	451	361	361	419	613	586	644

	C-35	C-34	C-33	C-32	C-31	C-30	C-29	C-28	C-27	C-26	C-25
	-6.62	-5.92	-7.93	-7.42	-8.19	-7.43	-7.12	-6.88	-6.51	-6.39	-5.15
	-6.2	-5.64	-7.22	-6.68	-6.92	-6.25	-5.53	-6.22	-5.98	-5.44	-4.34
	69	96	26	38	31	31	30	28	30	19	23
	-8.68	-7.59	-6.88	-5.37	-25.73	-19.34	-2.44	0.56	-19.63	-2.78	3.65
	-17.65	-15.04	-13.79	-12.01	-47.25	-35.03	-3.59	-0.31	-35.49	-3.3	4.32
	-127.54	-108.41	-85.8	0.19	-110.02	-197.97	7.79	40.14	-141.78	11.09	59.63
	-51.78	-60.43	-69.57	3.72	-57.84	-139.67	7	40.05	-35.28	11.45	68.91
	NA	5.54	6.63	2.75	5.78	6.34	-0.17	1.88	5.83	-0.96	1.09
	2.25	9.5	22.15	8.98	-33.13	-1.38	5.41	5.25	-101.28	8.73	28.44
	-24.37	-37.05	-37.74	-44.15	-97.16	-100.3	-15.09	-2.25	-81.53	-6.93	23.15
	399	390	389	452	625	476	76	331	527	70	377
	422	420	386	400	529	461	193	345	541	147	336
	513	497	427	430	610	531	168	377	614	111	303



C-36	-7.69	-6.81	55	-6.2	-11.92	-80.46	-32.45	6.99	3.07	-36.8	494	492	567
C-37	-7.44	-6.82	73	-22.64	-40.07	-216.1	-90.02	NA	-60.24	-67.55	390	411	553
C-38	-7.5	-6.98	62	-0.71	-2.49	31.23	29.14	1.59	-28.06	-25.52	102	200	171
C-39	-7.98	-7.34	45	-1.5	-3.43	17.44	18.85	2.61	-5.87	-25.26	221	267	240
C-40	-7.53	-7.31	104	-9.78	-18.13	-23.19	8.24	6.55	6.17	-25.9	440	448	509
C-41	-6.4	-5.71	24	-23.16	-39.5	-97.24	-47.09	8.13	-10.12	-63.74	619	566	541
C-42	-7.14	-6.24	20	-43	-73.21	-200.36	-135.52	8.53	-51.7	-92.47	357	417	386
C-43	-6.27	-5.68	92	-1.94	-6.13	-19.14	8.07	6.86	-1	-6.67	453	433	414
C-44	-7.44	-6.34	52	0.34	-1.46	43.15	19.28	1.08	6.07	-18.15	219	257	240
C-45	-7.62	-6.48	82	-7.17	-13.38	-33.41	-7.27	6.22	-4.07	-24.13	433	435	397
C-46	-5.8	-5.14	34	-4.45	-6.13	33.19	27.79	1.81	0.32	-5.1	167	257	222



	C-47	C-48	C-49	C-50	C-51	C-52	C-53	C-54	C-55	C-56	C-57
	-6.2	-7.37	-7.28	-6.78	-11.03	-7.44	-7.53	-6.97	-7.53	-6.5	-7.6
	-5.58	-6.55	-6.4	-6.21	-9.93	-6.15	-6.88	-6.08	-6.85	-5.57	-6.35
	43	18	34	63	19	15	12	28	23	38	29
	-1.4	-38.46	-3.09	4.92	-0.06	-3.3	-0.94	-10.62	3.38	-0.82	-9.08
	-3.03	-66.37	-7.12	2.78	-4.49	-3.63	-1.82	-18.35	3.33	-2.45	-13.57
	30.34	-260.25	0.48	26.27	6.77	2.2	16.34	-88.66	40.93	10.29	-67.26
	18.84	-163.18	6.5	48	17.83	2.77	19.93	-52.01	42.52	3.52	-49.81
	5.14	NA	5.55	7.9	6.62	0.19	0.99	2.56	3.71	1.86	4.33
	7.02	-70.17	-5.94	-22.66	1.85	2.1	2.1	-34.29	15.9	3.2	-17.46
	-39.9	-124.36	-34.54	-5.93	-17.76	-5.37	-11.28	-58.92	-3.6	-20.55	-48.6
	539	563	227	321	457	123	158	467	204	182	189
	492	552	290	336	419	173	240	432	253	209	297
	463	662	255	321	396	260	194	392	234	178	313



C-58	-7.26	-6.59	17	-3.52	-5.89	6.22	13.35	3.51	-0.27	-23.25	259	326	356
C-59	-7.22	-6.62	143	-0.96	-2.8	24.61	25.6	4.2	2.73	-19.47	461	434	470
C-60	-8.02	-7.07	71	3.49	2.73	52.46	53.08	5.68	29.29	-12.54	417	389	421
C-61	-8.74	-7.64	52	-8.2	-15.45	-75.71	-56.04	4.24	9.77	-32.42	411	378	457
C-62	-9.09	-8.6	56	-6.75	-14.71	-68.07	-42.16	6.5	8.42	-34.94	389	398	479
C-63	-7.91	-6.57	59	-22.1	-38.17	-65.33	-24.36	2.49	-29.92	-83.62	308	365	497
C-64	-7.65	-6.88	25	-19.56	-37.23	-116.82	-57.99	7.97	0.67	-66.26	769	633	707
C-65	-8.28	-7.17	56	-12.09	-22.84	-69.3	-24.78	5.63	-6.8	-49.51	380	409	477
C-66	-9.02	-8	28	4.01	4.74	57.15	48.74	0.13	16.43	9	130	174	151

There are two hydrogen bond formed between receptor and ligand, Figure 3A; NH group of compound C-40 forms hydrogen bond of bond length 1.9 Å with oxygen of Asn 380 residue and oxygen of ligand forms hydrogen bond with SH of Cys 234 of receptor, with bond length 1.9 Å. During receptor-ligand interaction, lowest binding energy obtained is -7.5 kcal/mol, mean binding energy of the cluster is -7.31 kcal/mol and total number of conformation obtained in the cluster is 104, which is 52 % of total cluster with respect to 200 poses calculated, that states ligand is quite steady when it binds to receptor.

Figure 4A, Oxygen of long end of compound C-40 forms hydrogen bond with NH group and SH groups of Cys 234 residue, with bond length 1.9 Å and 2.2 Å respectively. Fluorine from the short end of compound forms hydrogen bond with NH group of Gln 228 residue of heavy chain with bond length 2.0 Å.

Figure 4B, is a 2D diagram of computed structure of compound a with enzyme residues, which shows different types of interactions. All residues are colored in pink circles, which indicate they take part in hydrogen-bond, charge or polar interactions between ligand and receptor. Some atoms of compound are surrounded by a blue halo, it indicates about solvent accessible surface of an atom and diameter of circle is

proportional to the solvent accessible surface. Compound C-40 forms three hydrogen bonds with heavy chain of the enzyme, represented with blue dashed lines, arrows are directed towards electron donor. There is a Pi interaction formed between compound C-40 and Asp 1 residue of enzyme represented with an orange line.

Compound C-48 form hydrogen bonds with exclusion domain, heavy chain and light chain. Hydroxyl group of short end of compound forms hydrogen bond with oxygen of Asp 1 residue with bond length 1.7 Å. NH group of long end form hydrogen bond with oxygen of Asn 380 residue of light chain with bond length 2.0 Å and NH group present at the tail of long end of compound forms hydrogen bond with oxygen of Gln 228 residue with bond length 1.8 Å. Lowest binding energy of the compound obtained is -7.37 kcal/mol, and mean binding of 18 conformations is -6.55 kcal/mol.

Figure 6A, two oxygen of short end of compound C-48 forms hydrogen bonds with SH group of Cys 234 residue of heavy chain with bond lengths 1.5 Å and 2.8 Å respectively. Oxygen of long end of the compound forms hydrogen bond with NH group of Gly 277 residue with bond length 2.1 Å and fluorine of long end forms hydrogen bond with NH group of Asp 1 residue with bond length of 2.3 Å.

Figure 6B, is a 2D diagram of computed structure of compound belong with residues of enzyme and it shows different types of interactions formed between ligand and receptor. All residues are colored with a pink circle around them, which indicates that they take part in hydrogen-bond, charge or polar interactions between the ligand and receptor. Some atoms of compound are surrounded by a blue halo, it indicates solvent accessible surface of an atom and diameter of circle is proportional to the solvent accessible surface.

There is a Pi interaction formed between compound C-48 and Asp 1 residue of receptor represented with an orange line. Hydrogen bonds are represented in green and blue dotted lines, which indicate ligand binds with main chain and side chain of residues respectively; arrows are directed towards electron donor.

BPU is an heteropolymeric molecule $(\alpha\beta\gamma)^3$ with exact threefold symmetry and contains flexible subunit composition which depends on organism. The structure of the active site of urease is highly conserved which contains two nickel ions and has a comparably small volume. Urease inhibitors which are distinct in structure have been effectively identified. The most efficient urease inhibitors are Diamidophosphate (DAP) (Scheme 2) and its derivatives, which hydrolyze to the active molecule (DAP) in the active site [43].

DAP is a transition state analogue. It loses stability in aqueous environments because it contains hydrolyzable (especially at low pH) P-N bonds in its structure [44]. So the effort was done to modify the structure of this transition state analogue by using highly stable P-C or C-P-C bonds to improve its activity and stability against *Bacillus pasteurii* urease.

By using the knowledge-based design approach, 44 different compounds were designed to evaluate their potency against BPU. After designing the novel compounds using the knowledge of already synthesized urease inhibitors, they were energy minimized to the closest local minimum using the molecular mechanics CHARMM force field implemented in Discovery Studio. To study the interaction between the ligand and enzyme active site, all 44 compounds were docked in to the enzyme active site and were evaluated with 10 different scoring functions (Ligscore 1, Ligscore 2, PLP1, PLP2, Jain, PMF, PMF04, Ludi-1, Ludi-2, Ludi-3) of Discovery Studio package.

Automated docking was used to locate the appropriate binding orientations and conformations of different inhibitors in the BPU. To perform the task, genetic algorithm routine implemented in the program AutoDock was employed. Kollman charge, atomic solvation parameters and fragmental volumes were assigned to the protein using MGL Tools package. The program AutoGrid was used to generate the grid maps. Lamarckian genetic algorithm was applied for minimization using default parameters. The standard docking protocol was then applied by using AutoDock software package and binding free energies (ΔG_b , Kcal/mol) were obtained.

Inhibitor C-44, C-43, C-42, C-40, C-39, C-38, C-36 and C-35 (Table 1) (Scheme 3) are top ranked compounds according to evaluation of 10 scoring functions and binding free energy of AutoDock and these 8 compounds have also obtained highest number of conformations which were quite well overlaid. The inhibitors constructed by using phenyl/aromatic/aliphatic rings with fluorine/chlorine substitutions were obtained top ranked compounds in most of the scoring functions among 44 different structures.

4. CONCLUSION

To conclude the work, it is an approach to design new compounds that can be potential inhibitors for DPPI. For designing new potential inhibitors of DPPI several known inhibitors and their activities with the enzyme were considered. Compounds containing non cyclic derivatives, benzene, benzene with modifications, saturated 6-member ring, saturated 5-member ring and unsaturated 5-member ring were designed.

Binding free energy and binding affinity of all compounds with enzyme were calculated by AutoDock and Discovery Studio Client 3.1 and their interactions with receptor were studied. It has been found that, in

AutoDock ligand is treated as flexible and protein is treated as rigid, which states that ligand interacts with receptor on its outer surface and water molecules actively participates in the receptor-ligand interaction.

Using Discovery Studio Client 3.1 package, PDB coordinates of enzyme are kept rigid, whereas structure of enzyme fluctuates and the same is for inhibitor. It indicates that ligand has possibilities to interact with receptor at their inner and outer surface. Possible active role of water molecules in the receptor-ligand interaction is neglected. As the active site of the DPPI is located on its outer surface, there are not much difference seen in interactions between enzyme and all newly designed compounds in both methodologies.

Binding energy of all generated conformations for each inhibitor has been calculated by Autodock, and the best conformations showing lowest energy in cluster were used for ranking compounds. From newly designed compounds C-40 and C-48 (Scheme 2) can be considered as top ranked compounds, as they show strong interaction with enzyme. It has been observed that lowest binding energy, mean binding energy and total conformations in clusters, and scores of compound C-40 and C-48 are much better than standard inhibitor and their results correlates in both methodologies. In Autodock lowest binding energy from all generated conformations were considered for ranking the compounds. Compound C-40 shows lowest binding energy -7.53 kcal/mol, 104 conformations in cluster and Ludi_3 score 509, all these score describes that compound 40 form strong interactions with enzyme in a steady state. Compound C-48 shows lowest binding energy -7.37 kcal/mol, -PLP1 score -260.25, -PMF04 score -124.36 and Ludi_3 score 662, which states this compound could be a potential structure of DPPI.

Compounds containing non cyclic derivatives have shown weak interaction with enzyme than other compounds and standard inhibitor. Gly-Phe fragment containing benzene, benzene with modifications, saturated 6-member ring, saturated 5-member ring or unsaturated 5-member ring at one end and functional group at other end, can be potential structure for investigated protease. Further modification in these newly designed compounds could increase their inhibitory activity.

ACKNOWLEDGEMENTS

Authors are thankful to Prof. W. Andrzej Sokalski for sharing his knowledge, and The Supercomputing and Networking Centre in Wroclaw, for providing their hardware and software resources on which molecular modeling was performed.

REFERENCES

- [1] M. J. McGuire, P. E. Lipsky and D. L. Thiele, "Cloning and characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C)", *Biochim. Biophys. Acta*, 273:1351-1267, (1997).
- [2] L. Hola-Jamriska, J. F. Tort, J. P. Dalton, S. R. Day, J. Fan, J. Aaskov and P. J. Brindley, "Cathepsin C from *Schistosoma japonicum*: cDNA encoding the proenzyme and its phylogenetic relationships", *Eur. J. Biochem*, 255:527-534, (1998).
- [3] D. W. Hutchinson and A. Tunnicliffe, "The preparation and properties of immobilised dipeptidyl-aminopeptidase I (cathepsin C)", *Biochim. Biophys. Acta*, 916:1-4, (1987).
- [4] M. Klemba, I. Gluzman and D. E. Goldberg, "A *Plasmodium falciparum* dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation", *J. Biol. Chem*, 279:43000-43007, (2004).
- [5] J. M. Clair, C. T. Pham, S. A. Villalta, G. H. Caughey and P. J. Wolters, "Mast cell dipeptidyl peptidase I mediates survival from sepsis", *J. Clin. Invest*, 113:628-634, (2004).
- [6] A. M. Adkison, S. Z. Raptis, D. G. Kelley and C. T. Pham, "Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis", *J. Clin. Invest*, 109:363-371, (2002).
- [7] J. Bondebjerg, H. Fugsang and K. R. Valeur, "Novel Smicarbazine-derived inhibitors of human Dipeptidyl peptidase I (hDPPI)", *Bioorg. Med. Chem*, 13:4408-4424, (2002).
- [8] H. R. Gutmann and J. S. Frutin, "On the Proteolytic Enzymes of Animal Tissues: VIII. An Intracellular Enzyme Related to Chymotrypsin", *Biol. Chem*, 174:851-858, (1948).
- [9] D. Turk, V. Janjic, I. Stern, M. Podobnik, D. Lamba, S. W. Dahl, C. Lauritzen, J. Pedersen, V. Turk and B. Turk, "Structure of human Dipeptidyl peptidase I (Cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases", *EMBO. J*, 20:6570-6582, (2001).
- [10] M. Sajida, J. H. McKerrow, E. Hansella, M. A. Mathieua, K. D. Lucasa, I. Hsieha, D. Greenbauma, M. Bogyob, J. P. Saltera, K. C. Lima, C. Franklina, J. H. Kima and C. R. Caffreya, "Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase", *Mol. Biochem. Parasitol*, 131:65-75, (2003).
- [11] E. Kominami, K. Ishido, D. Muno and N. Sato, "The primary structure and tissue distribution of cathepsin C", *Bio. Chem. Hoppe-Seyler*, 373:367-373, (1992).
- [12] R. Shapiro and B. L. Vallee, "Interaction of human placental ribonuclease with placental ribonuclease inhibitor", *Biochemistry*, 30:2246-2255, (1991).
- [13] Laine, Dramane, Palovich, Michael, McClelland, Brent, Petitjean, Emilie, Delhom, Isabelle, Xie, Haibo, Deng, Jianghe, Lin, Guoliang, Davis, Roderick, Jolit, Anais, Nevins, Neysa, Zhao, Baoguang, Villa, Jim, Schneck, Jessica, McDevitt, Patrick, Midgett, Robert, Kmett, Casey, Umbrecht, Sandra, Peck, Brian, Davis, A. Bacon, Bettoun and David, "Discovery of Novel Cyanamide-Based Inhibitors of Cathepsin C", *ACS Med. Chem. Lett*, 2:2, (2011).

- [14] Guay, Daniel, Beaulieu, Christian, Truchon, Jean-Francois, Reddy, T. Jagadeeswar, Bayly, Christopher I., (Zamboni, Robert, Methot, Nathalie, Rubin, Joel, Ethier, Diane, Percival and M. David, "Design and synthesis of dipeptidyl nitriles as potent, selective, and reversible inhibitors of cathepsin C", *Bioorg. Med. Chem. Lett.*, 20:1463, (2010).
- [15] C. M. Kam, M. G. Gotz, G. Koot, M. McGuire, D. Thiele, D. Hudig and C. J. Powers, "Design and evaluation of inhibitors for dipeptidyl peptidase I (Cathepsin C)", *Arch. Biochem. Biophys.*, 427:123-134, (2004).
- [16] Suban, Dejan, Zajc, Tajana, Renko, Miha, Turk, Boris, Turk, Vito, Dolenc and Iztok, "Cathepsin C and plasma glutamate carboxypeptidase secreted from Fischer rat thyroid cells liberate thyroxin from the N-terminus of thyroglobulin", *Biochimie*, 94:719-726, (2012).
- [17] H. Lindley, "The Specificity of Dipeptidyl Aminopeptidase I (Cathepsin C) and its Use in Peptide Sequence Studies", *Biochem. J.*, 126:683-688, (1972).
- [18] R. Gianetto and C. D. Duve, "Tissue fractionation studies. 4. Comparative study of the binding of acid phosphatase, β -glucuronidase and cathepsin by rat-liver particles", *Biochem. J.*, 59:433-438, (1955).
- [19] M. J. Guire, P. E. Lipsky and D. L. Thiele, "Cloning and characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C)", *Biochim. Biophys. Acta*, 1351:267-273, (1997).
- [20] L. H. Jamriska, J. F. Tort, J. P. Dalton, S. R. Day, J. Fan, J. Aaskov and P. J. Brindley, "Cathepsin C from *Schistosoma japonicum*, cDNA encoding the proenzyme and its phylogenetic relationships", *Eur. J. Biochem.*, 255:527-534, (1998).
- [21] D. W. Hutchinson and A. Tunncliffe, "The preparation and properties of immobilised dipeptidyl-aminopeptidase I (cathepsin C)", *Biochim. Biophys. Acta*, 916:1-4, (1987).
- [22] M. Klemba, I. Gluzman and D. E. Goldberg, "A *Plasmodium falciparum* Dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation", *J. Biol. Chem.*, 279:43000-43007, (2004).
- [23] M. J. McGuire, P. E. Lipsky and D.L. Thiele, "Generation of active myeloid and lymphoid granule serine proteases requires processing by the granule thiol protease dipeptidyl peptidase I", *J. Biol. Chem.*, 268:2458-2467, (1993).
- [24] A. M. Adkison, S. Z. Raptis, D. G. Kelley and C. T. Pham, "Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis", *J. Clin. Invest.*, 109:363-71, (2002).
- [25] C. T. Pham and T. J. Ley, "Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B in vivo", *Proc. Natl. Acad. Sci. USA*, 96:8627-32, (1999).
- [26] P. J. Wolters, C. T. Pham, D. J. Muilenburg, T. J. Ley and G. H. Caughey, "Dipeptidyl peptidase I is essential for activation of mast cell chymases, but not tryptases, in mice", *J. Biol. Chem.*, 276:18551-18556, (2001).
- [27] M. J. Guire, P. E. Lipsky and D. L. Thiele, "Purification and characterization of dipeptidyl peptidase I from human spleen", *Arch. Biochem. Biophys.*, 295:280-288, (1992).
- [28] N. M. Thot, D. Guay, J. Rubin, D. Ethier, K. Ortega, S. Wong, D. Normandin, C. Beaulieu, T.J. Reddy, D. Riendeau and M. D. Perciva, "In Vivo Inhibition of Serine Protease Processing Requires a High Fractional Inhibition of Cathepsin C", *Mol. Pharmacol.*, 73:1857-1865, (2008).
- [29] P. D. Sheth, J. Pedersen, A. F. Walls and A. R. McEuen, "Inhibition of dipeptidyl peptidase I in the human mast cell line HMC-1: blocked activation of tryptase, but not of the predominant chymotryptic activity", *Biochem. Pharmacol.*, 66:2251-2262, (2003).
- [30] P. J. Wolters, C. T. Pham, D. J. Muilenberg, T. J. Ley and G. H. Caughey, "Dipeptidyl Peptidase I is essential for activation of mast cell chymases, but not tryptases, in mice", *Biol. Chem.*, 276:18551-18556, (2001).
- [31] P. S. Hart, Y. Zhang, E. Firatli, C. Uygur, M. Lotfazar, M. D. Michale, J. J. Marks, X. Lu, B. J. Coates, W. K. Seow, R. Marshall, D. Williams, J. B. Reed, J. T. Wright, T. C. Hart, "Identification of cathepsin C mutations in ethnically diverse Papillon-Lefèvre syndrome patients", *Med. Genet.*, 37:927-932, (2000).
- [32] J. M. Clair, C. T. Pham, A. Villalta, G. H. Caughey and P. J. Wolter, "Mast cell dipeptidyl peptidase I mediates survival from sepsis", *J. Clin. Invest.*, 113:628-634, (2004).
- [33] A. Molgaard, J. Arnau, C. Lauritzen, S. Larsen, G. Petersen and J. Pedersen, "The crystal structure of human Dipeptidyl peptidase I (Cathepsin C) in complex with the inhibitor Gly-Phe-CHN₂", *Biochem. J.*, 401:645-650, (2007).
- [34] D. L. Thiele, P. E. lipsky and M. J. McGuire, "U.S. patent", 602:102, (1997).
- [35] G. D. Green and E. J. Shaw, "Peptidyl diazomethyl ketones are specific inactivators of thiol proteinases", *Biol. Chem.*, 256:1923-1928, (1981).
- [36] S. A. Thompson, P. R. Andrews and R. P. Hanzlik, "Carboxyl-modified Aminoacids and Peptides as Protease Inhibitors", *Med. Chem.*, 29:104-111, (1986).
- [37] J. Bondebjerg, H. Fuglsang, K. R. Valeur, J. Pedersen and L. Naerum, "Dipeptidyl nitriles as human dipeptidyl peptidase I inhibitors", *Bioorg. Med. Chem. Lett.*, 3614-3617, (2006).
- [38] C. M. Kam, M. G. Gotz, G. Koot, M. McGuire, D. Thiele, D. Hudig and J. C. Powers, "Design and Synthesis of Inhibitors for Dipeptidyl Peptidase", *Arch. Biochem. Biophys.*, 427:123-134, (2004).
- [39] H. U. Demuth, U. Heiser and A. Niestroj, WO 03/022871 A2, (2003).
- [40] M. Horn, M. Pavlik, L. Doleckova, M. Baudys and M. Mares, "Arginine-based structures are specific inhibitors of cathepsin C. Application of peptide combinatorial libraries", *Eur. J. Biochem.*, 267:3330-3336, (2000).
- [41] A. J. Barrett, A. A. Kembhavi, M. A. Brown, H. Kirschke, C. G. Knight, M. Tamai and K. Hanada, "L-trans-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L", *Biochem. J.*, 201:189-198, (1982).
- [42] A. Anastasi, M. A. Brown, A. A. Kembhavi, M. J. Nicklin, C. A. Sayers, D. C. Sunter and A. J. Barrett, "Cystatin, a protein inhibitor of cysteine proteinases: improved purification from egg white, characterization, and detection in chicken serum", *Biochem. J.*, 211:129-138, (1983).
- [43] F. H. Clarke, "How Modern Medicines are Discovered", *Futura*, (1973).

- [44] W. Sneader, "Drug Prototypes and Their Exploitation", Wiley, (1996).
- [45] M. A. Johnson and G. M. Maggiora, "Concepts and Applications of Molecular Similarity", John Wiley & Sons, (1990).
- [46] H. J. Bohm, A. Flohr and M. Stahl, "Scaffold hopping", *Drug Discov Today: Technologies*, 1:3, (2004).
- [47] H. Zhao, "Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective", *Drug Discov Today*, 12:3-4, (2007).
- [48] P. R. Andrews, "Functional group contributions to drug-receptor interactions", *J. Med. Chem*, 27:1648-1657, (1984).
- [49] A. C. Zapatero and J. T. Metz, "Ligand efficiency indices as guideposts for drug discovery", *Drug Discov. Today*, 10:464-469, (2005).
- [50] M. M. Hann and T. I. Oprea, "Pursuing the leadlikeness concept in pharmaceutical research", *Curr. Opin. Chem. Biol*, 8:255-263, (2004).
- [51] S. J. Teague, "The design of leadlike combinatorial libraries", *Angew. Chem. Int. Ed. Engl*, 38:3743-3748, (1999).
- [52] G. W. Bemis and M. A. Murcko, "The properties of known drugs. 1. Molecular frameworks", *J. Med. Chem*, 39:2887-2893, (1996).
- [53] A. Schuffenhauer, "Relationships between molecular complexity, biological Activity, and structural diversity", *J. Chem. Inf. Model*, 46:525-535, (2006).
- [54] O. V. Buzko, A. C. Bishop and K. M. Shokat, "AutoDock for accurate docking of protein kinase inhibitors", *J. Comput. Aided Mol. Des*, 16:113-127, (2002).
- [55] D. S. Goodsell, G. M. Morris and A. J. Olson, "Automated Docking of Flexible Ligands: Applications of AutoDock", *J. Mol. Recognit*, 9:1-5, (1996).
- [56] R. Huey, G. M. Morris, A. J. Olson and D. S. Goodsell, "A semiempirical free energy force field with charge-based desolvation", *J. Comput. Chem*, 28:1145-1152, (2007).
- [57] B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, "The Biomolecular simulation Program", *J. Comp. Chem*, 30:1545-1615, (2009).
- [58] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *J. Comp. Chem*, 4:187-217, (1983).
- [59] A. Krammer, P. D. Kirchhoff, X. Jiang, C. M. Venkatachalam and M. Waldman, "LigScore: a novel scoring function for predicting binding affinities", *J. Mol Graph Model*, 23:395-407, (2005).
- [60] P. A. Kollma, "Free energy calculations: applications to chemical and biochemical phenomena", *Chem. Rev*, 93:2395-2417, (1993).
- [61] J. Aqvist, V. B. Luzhkov and B. O. Brandsdal, "Ligand binding affinities from MD simulation", *Acc. Chem. Res*, 35:358-365, (2002).
- [62] H. A. Carlson and W. L. Jorgensen, "An extended linear response method for determining free energies of hydration", *J. Phys. Chem*, 99:10667-10673, (1995).
- [63] I. Muegge and Y. C. Martin, "A general fast scoring function for protein-ligand interactions: a simplified potential approach", *J. Med. Chem*, 42:791-804, (1999).
- [64] J. B. Mitchell, R. A. Laskowski, A. Alex and J. M. Thornton, "BLEEP potential of mean force describing, protein-ligand interactions. Part I. Generating potential", *J. Comp. Chem*, 20:1165-1176, (1999).
- [65] H. Gohlke, M. Hendlich and G. Klebe, "Knowledge-based scoring function to predict protein-ligand interactions", *J. Mol. Biol*, 295:337-356, (2000).
- [66] H. J. Bohm, "Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs", *J. Comput. Aided Mol. Des*, 12:309-323, (1998).
- [67] D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel and S. T. Freer, "Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming", *Chem. Biol*, 2:317-324, (1995).
- [68] G. M. Verkhivker, D. Bouzida, D. K. Gehlhaar, P. A. Rejto, S. Arthurs, A. B. Colson, S. T. Freer, V. Larson, B. A. Luty, T. Marrone and P. W. Rose, "Deciphering common failures in molecular docking of ligand-protein complexes", *J. Comput. Aided Mol. Des*, 14:731-751, (2000).
- [69] A. N. Jain, "Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities", *J. Comput. Aided Mol. Des*, 10:427-440, (1996).
- [70] A. E. Muryshev, D. N. Tarasov, A. V. Butygin, O. Y. Butygina, A. B. Aleksandrov and S. M. Nikitin, "A novel scoring function for molecular docking", *J. Comput. Aided Mol. Des*, 17:597-605, (2003).