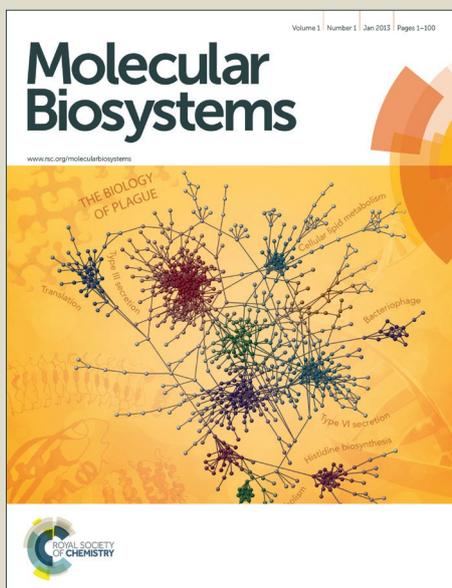


Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

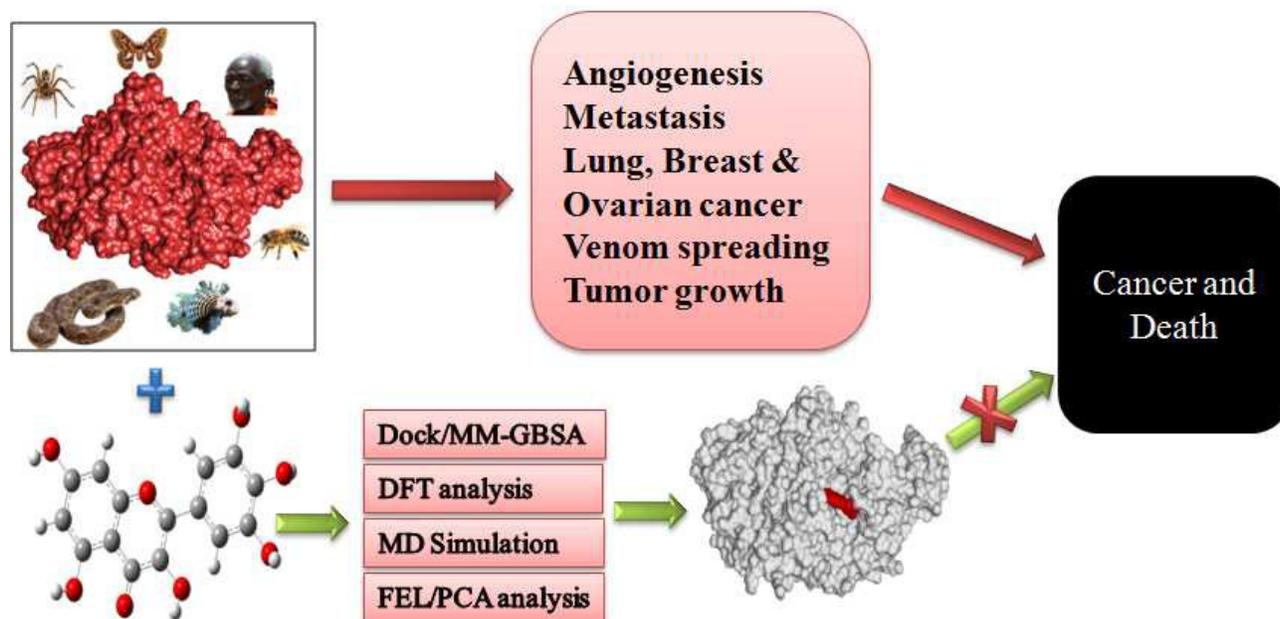


www.rsc.org/molecularbiosystems

Molecular BioSystems

Graphical Abstract

The study focused on structural, ligand induced conformational change of viper Hyaluronidase and also it provides insight to the structure based drug design for eukaryotic Hyaluronidases which could be the future drug target in cancer and spreading factor.



Sequence diversity and Ligand induced structural rearrangements of viper Hyaluronidase

A. Christian Bharathi^a, Pradeep Kumar Yadav^a and B. Syed Ibrahim^{*a}

Hyaluronidases (Hyal) are class of carbohydrate-active enzyme involved in angiogenesis, cancer proliferation, tumour growth and venom spreading. Functionally significant Hyals are responsible for fast spreading of venom to the target site of action. The absence of molecular diversity, structural and functional behaviour of snake venom Hyals were paved notion to the objective of the study. *Echis pyrimedium Leakyi*'s Hyaluronidase (EHY) based phylogenetic analysis was ensured the existence of two function group of Hyals which were diverged from coral snake (the ancestor). The structure was modelled and it was founded that E-Loop region (211-224 AA) was only present in EHY compare to templates which may account for significant function of viper Hyal. Best interacting ligands were screened from the selected plant derivatives and MYR was consisted of better pharmacophore features (AADDRRR) in comparison with other ligands. Further HOMO, LUMO, MO energies and energy gaps of CGA, MIM and MYR were calculated by DFT analysis. EHY-ligand complex stability and interactions were investigated through MD simulation and FEL analysis. These revealed that MIM and MYR or their derivative compounds could be prominent lead molecule for both EHY and other eukaryotic Hyal. PCA analysis of both unligated and ligated forms was confirmed that Loop-III (86-96 AA) and E-Loop regions structural rearrangements were very essential for association and dissociation process of substrate. Particularly, ARG92 and LYS219 are determined as important key residues from the conformational changes. These regions dynamic behaviour can be associated with HA binding and catalytic function of EHY. This result can extend our

knowledge of viper Hyal functional behaviour and provides structural insight to target eukaryotic Hyals as forthcoming drug target in cancer and venom spreading.

a: Centre for Bioinformatics, Pondicherry University, Pondicherry, India.

Corresponding author Email: ibrahim.bic@pondiuni.edu.in

Supplementary information available: find the ESI document file

Introduction

Hyaluronidases (EC:3.2.1.35) are belongs to Glycoside Hydrolases (GH), a subfamily in carbohydrate-active enzymes which is involved in the "carbohydrate metabolism". Hyals biological functions and mechanism of actions have been obscured for many decades so that they were placed in the group of neglected enzymes^{1,2}. Hyals are hydrolyses Hyaluronic acid (HA) at β (1-4) bond formed between NAG (N-acetyl Glucosamine) and GUA (Glucuronic acid). The catalysis is activated by two acidic residues, in which one can act as a proton donor (a general acid) while another can act as nucleophile/base³. HA polymers are an integrated component of ECM (Extra cellular matrix), cartilage, connective tissues, synovial fluid, tissues and skins. HA polymer and its subsequent degradation products participates in the cellular process like cell differentiation, tumour progression, inflammatory process, cell-cell adhesion, angiogenesis, signalling, tissue injury and repair mechanism⁴⁻⁷.

Hyaluronate Lyases (Polysacchiride Lyases, EC:4.2.2.1), an another class of enzymes, which is also cleaving the same substrate (HA) but with different structural motif (α and β domain) and mechanism of action (β -elimination) was well studied from *Streptococcus pneumonia* species⁸. This is one of the main virulent factor which is responsible for the bacterial invasion by degrading the HA. On the other hand, eukaryotic Hyal structures are containing (β/α)_x TIM barrel motif, thus the common structural fold observed in GH family. Most of the mammalian species were consist of catalytic domain for cleaving HA and EGF-like domain whose functional role yet to be understand⁹.

The crystal structure of Human and Bee Hyal are already reported but their mechanism of action is still unclear^{9, 10}. These structures were accorded with the proposed acid-base or double displacement mechanism of action of Hyal. From the analysis of both the structure and sequence, it revealed that they had similar shape and size of substrate binding cavity and shared 32.13 % sequence identity. The substrate binding is favoured by electrostatic (ARG and LYS at active site region to carboxylic group of HA) and hydrophobic interactions (TYR at cavity to hydrophobic regions of HA). Functional mutation study on human sperm PH-20 was identified D146 and E148 (possible proton donor) which is equivalent to D111 and E113 of 1FCV as most important catalytic residues Hyals¹¹.

Eukaryotic Hyals were regulating many biological processes like fertilization, extra cellular invasive, metastasis, angiogenesis, cell differentiation and proliferation, cancer and venom spreading⁹⁻¹². Venom contains a typical and functionally diverged Hyal which is non-toxic enzyme but ubiquitously presence in all secreted type venom. Biochemical studies are reported with the characterization of Hyal in the venom of scorpion, bee, hornet, spider, stone fish and snake species^{10, 13-17}. Venom Hyals are an essential and active key enzyme which enhance the fast spreading and diffusion of the venom to blood circulation, hence they termed as "spreading factor"¹⁷. Snake bite is a common socio-economic problem and especially snake bite affects more than 4.5 million people annually around the world. Snake Hyal is the main factor which influenced more for the local tissue damage and enhancing the rapid diffusion of venom. If this Hyal is neutralized earlier, the toxic proteins spreading and effect can be reduced dramatically. A leading drug against these Hyals will assist and improve the antivenom therapy when it administrated earlier or along with available treatments.

Plant compounds are known with their effective neutralizing capacity against viper and other snake Hyal's enzyme activity, spreading and delaying lethality^{18, 19}. There is only limited biochemical and structural studies of Hyal were reported but their mechanism of inhibition, structural and functional behaviour is still blurred. Hence, the present study spotlighted on phylogenetic analysis, inhibitor binding mechanism and conformational behaviour study of *Echis pyrimedium* Leakyi Hyal (EHY).

Materials and methods

Sequence analysis and Phylogeny of Viper Hyal:

A search for EHY similar sequences was performed using protein BLAST at NCBI on non-redundant protein sequence databases with the default parameters. The sequences were filtered based on the identity more than 65 percentage and less E-value. Partial and non-poisonous snake's Hyal sequence entries were excluded. Only thirteen matured and annotated sequences were selected and they were collected from the Uniprot Database (<http://www.uniprot.org>). First, the sequences were aligned with the help of ClustalX²⁰ by using the default parameters. Then MEGA5²¹ were used to construct the phylogeny and diversity relationship of these sequences based on Maximum likelihood method in particular JTT matrix-based model²². All positions with gaps and missing data of sequences were eliminated. There were a total of 424 positions were found in the final dataset. The global tree was obtained from the calculation of a maximum-likelihood distance matrix normalized to the amino acid composition of the data set. The resulting tree topology was used as a seed to search for a better topology with highest likelihood value by the same amino acid substitution model. Further, the MSA of bee (PDB ID: 1FCV), human (PDB ID: 2PE4), human sperm PH-20 (Uniprot:P38567) and *Echis* (Uniprot:A3QVN6) Hyal were performed to understand the eukaryotic Hyal sequence similarity and conserved regions.

Structural model of EHY

In order to get the three dimensional structure, EHY complete model was predicted using comparative homology modeling²³. The sequence of EHY²⁴ (445aa, UniProt ID: A3QVN6) was obtained from the UniProt database; and BLAST search was performed Protein database to deduce appropriate homologue templates. Human Hyal (PDB ID:2PE4|chainA) and bee Hyal (PDB ID:2PE4|chainA) showing 41.3% and 33.4% identity and 92% and 70% query coverage respectively were selected as templates for comparative homology. The selected templates were used to generate 10 models using Modeller9.14 (22). Here, the N-terminal region (1-34) was eliminated from model structure due to devoid of homologues template. The best model was selected from least discrete optimized protein energy (DOPE) and modeler objective function (MOF) scores and carried for further refinement with loop refinement script. The qualities of the refined model were validated using SAVES Server which includes PROCHECK²⁵, ERRAT²⁶ and VERYFY3D²⁷. The loop refined model was subjected to MD run of 10ns in order to obtain near native and average ensemble structure.

MD simulation of EHY

The EHY average ensemble structure was obtained by MD run using GROMACS 5.0.4 package²⁸. First, the protein was altered by adding the hydrogen atom and the topology was generated using the GROMOS96 54a7 force field. The cubic box type used to place the protein where the edge of the box from the protein molecules set as 1.0 nm in all the directions. The SPC216 water model was used to solvate the system and total charges were neutralized by adding 12 Cl⁻ ions. Then, the solvated system was allowed to minimize their energy by steepest descent algorithm with maximum of 50,000 steps in order to obtain the convergence tolerance of 1000 kJ mol⁻¹ nm⁻¹ which subsequently allowed for the conjugate gradient algorithm with same convergence tolerance. The system equilibrium was obtained by using both NVT and NPT methodology where the PME²⁹ electrostatic were used to treat long range electrostatic with PME order of 4 and a Fourier spacing of 0.16nm. In both, the bonded parameters were used LINCS holonomic constraints³⁰ and SETTLE algorithm³¹ was used to constrain the geometry of water. In addition, both protein and non-protein coupling groups were allowed to retain the temperature at 300 K using the V-rescale temperature coupling method³². But in NPT equilibration, the Parrinello-Rahman pressure coupling³³ was introduced to maintain a uniform scaling of box vectors with 1.0 bar as the reference pressure. Finally, the MD runs of 10ns time scale were carried out, and the simulated cluster average EHY structure was preceded for docking interaction with selected plant compounds.

Optimization and HOMO/LUMO of ligands using DFT method

All plant compounds (Ref. ESI, Table.1) geometries were optimized by DFT method using Gaussian09 package³⁴. The frequency at DFT level was performed to achieve the gradient geometry, bonding feature and charge optimization. DFT based optimization of compounds were carried out using Lee-Yang-Parr correlation functional (B3LYP) theory³⁵ with 6-311G* basis set. All the compounds geometry parameters were minimized without any constraint at a standard DFT level in potential energy surface. The optimized compounds were preceded to find better binding natural compounds with EHY. In addition to that, the HOMO and LUMO properties of screened compounds were calculated using the same method in gas phase. HOMO (Highest Occupied Molecular Orbital) energy determines the electron donor regions of ligand molecule while LUMO (Lowest Occupied Molecular Orbital) determines the electron acceptor regions.

Molecular docking and binding energy calculation

The optimized compounds were prepared with ligprep module of Maestro9.2 suite. EHY structure was prepared using Protein preparation wizard. The receptor grids of EHY was prepared at the substrate binding site with grid box size of 12*12*12

Å. The docking was performed using the SP method; based on interactions and docking energy the best five ligands were redocked with the XP method. The outputs were subjected to free energy calculation on MM-GBSA module of Maestro suite. Based on the docking scores and MM-GBSA free energy, best three plant compound were carried out to the EHY-ligand complex MD simulation to examine their interaction stability and conformational behavior of EHY. Topologies and gromacs inputs of three ligands were collected from the PRODRG server. MD simulations of EHY-ligand complexes was performed with the same MD parameter setup as mentioned previously except the box size, which was kept as 12 Å.

Conformational behavior and FEL analysis:

Principal component analysis (PCA) represents the concerted collective motion of molecule based on their eigenvectors from the mass-weighted covariance matrix. The functional mode changes of protein is determined by collective coordinate displacement of domains and they are termed as essential dynamics^{36, 37}. The conformational changes of EHY in unligated and ligated forms for the 415 C α atoms were subjected to the PCA calculation using the VMD NMZWiz plugin³⁸ (36). The covariance matrix was generated to define their collective coordinate motion from the mass-weighted matrix of principal components,

$$C_{ij} = (x_i - \langle x_i \rangle) (x_j - \langle x_j \rangle)$$

$\langle \rangle$ is the average over the instantaneous structures sampled during the period of simulation and x_i is a mass-weighted atomic coordinate. PCA calculation was performed after removal of overall translation and rotation of each configuration of EHY. The top most eigenvector represents the large amplitude fluctuations due to random diffusion; hence trivial modes were discarded in order to define the functional mode changes of EHY. Further, an absolute and sensitive quantification of trajectory convergence is an efficient method to bring the free energy landscape (FEL) of EHY-ligand complexes³⁹. From the calculated principal component (PCs), the cosine content which are lesser than 0.2 to 0.5 were selected and subjected to produce FEL smoother and single basin⁴⁰. The FEL of EHY-ligand complexes were mapped to get minimum energy configurations using the g_sham script from the two PCs chosen based on above mentioned cosine contents.

Results and discussion

Phylogenetic analysis of viper Hyal

To understand the molecular evolutionary and diversity relationship among the viper Hyal, the seed tree with the highest log likelihood score (-2277.6254) is derived from the dataset mentioned in the methods. All nodes are having high bootstrap score admits that the phylogenetic tree was most reliable. The tree consists of major two groups (shaded in two oval circles) and the target sequence (EHY: A3QVN6) was grouped in the group-I which is highlighted by red triangle. These two monophyletic groups were diverged from a common hypothetical node which was derived from the out group taxa (Eastern coral snake: U3FYQ4). Horned desert viper isoforms (A3QVN3, A3QVN4 & A3QVN5) and *Echis carinatus leakeyi* (A3QVN6) were closely related Hyal than *Echis ocellatus* (A3QVN2) and African puff adder isoforms (A3QVP0, A3QVN9) within the group-I. In group-II, Diamond black rattlesnake (J3S820) and Timber rattlesnake (T1D6Q3) were placed in subgroup I whereas Ryukyu island pit viper iso-forms (U3TDI3, U3TBU1) and Habu pit viper (U3TDE8) were in subgroup II as they are most sequence similarity. EHY (A3QVN6) was diverged simultaneously from pit viper and rattlesnakes and it derived from the common ancestor (coral snake). The branch length (number of change or mutation rate per amino acid) was denoted with a scale bar in the Fig.1-A.

From the MSA dataset, there are two highly conserved consensus patterns i) between 126-146 (FHGLGVIDWENWRPQWDRNWG) and ii) 148-159 (FHGLGVIDWENWRPQWDRNWG) were observed. These regions can play significant role in functional and structural activity of eukaryotic Hyals. Herein, all the sequences were found to have the common catalytic domain and EGF-like domain. In general, mammalian species have six Hyal gene variants either active or inactive forms with same structural fold⁴¹. These viper Hyals fall into the group of functionally active Hyal due to the presence of catalytic important consensus pattern (i). In the phylogenetic group-II, N-terminal regions of U3TDI3 and U3TDE8 were observed with deletion and addition of 23 and 76 residues respectively. This Insertion and deletion regions made the sister group within the group-II. The indel regions would be resultant of post-translation modification region or exon scission region of Hyal genes for the production of diverse Hyal variants. From this analysis, it ensured that the viper species have two major functional group of Hyal but their specific functional difference has to be confirmed through experiment.

Structure of EHY:

The three dimensional structure of EHY was obtained by comparative homology modeling as mentioned in the methodology. First, the model was secured 69.28 and 81 of ERRAT and VERIFY3D score with less RMSD variation with templates (1FCV: 1.176 Å and 2PE4: 0.316 Å). Followed by the refinement, the model was subjected to get the MD cluster average structure. The quality of resulted model was validated based on PROCHECK, ERRAT (92.83) and VERIFY3D (92.29) values. By using CATH, the built model was found to have (β/α)₈ TIM barrel domain between 34-355; and EGF-like domain between 357-440 residues. The (β/α)_x TIM barrel domain was a common structural fold of GH families which degrades GAG, but they are having different kind of mechanism of actions due to their functional diversity³. EGF-like domains were so far only observed in mammalian species in comparison with human Hyal-1 structure (PDBID:2PE4) whose distinct role in EHY and other Hyals yet to be determined⁹.

The tertiary structure of EHY consists of 11 α -helices, 10 β -strands (for more details Ref. ESI; text 1) and 5 disulfide bridges contribute for the compactness and structural stability (Fig.1-B). EHY contains five glycosylation sites (ASN67, ASN103, ASN111, ASN357 and ASN401) which were lying outside of EHY catalytic site. Additionally, extended loop (E-loop, between 211-224 AA) was found to be present only in EHY compared to template structures. So that it may account for significant functional difference in viper Hyal but peculiar precise role yet to be validated by experimental study. The catalytic region of eukaryotic Hyal have the consensus pattern of [G/A]-V-I-D-[W/F]-E-[X]-W-R-P which was observed from the MSA of bee, human, viper and human sperm (PH-20) species. This result shows that two catalytic residues ASP133 and GLU135 and neighboring tyrosine and arginine residues are conserved in most eukaryotic species and they are playing major role in the substrate binding and catalytic process¹⁰. Highly conserved cystine and functional important residues were highlighted in ESI, Fig.1.

EHY-ligand molecular docking interaction:

In order to define the binding mode of selected compounds, the docking was performed with EHY and the outputs were analyzed using glide module. Out of 24 plant compounds only five, namely Catechin, Chlorogenic acid (CGA), Mimosine (MIM), Myricetin (MYR) and Kaemferol were filtered based on glide score (cut off was set as -4.00) and glide energy in comparison with other compounds. The filtered compounds were residing in EHY cavity by interacting with catalytic (ASP133 or GLU135), basic (LYS148, ARG (271,295) and hydrophobic TYR (206,214) residues which are important for the substrate binding¹⁰ (Fig.2). The details of glide score, MM-GBSA energy, residues involved in hydrogen bonds were given in Table.1 (Ref. also ESI, Table.2 for details of interaction). Based on free energy and interactions, best three compounds (CGA: 37.44 kcal/mol, MIM: 34.22 kcal/mol and MYR: 35.78 kcal/mol) were chosen to the analysis of their drug-like properties using DFT method.

DFT analysis of compounds:

The selected three compounds were subjected to the HOMO and LUMO calculations for the determination of the reactivity and kinetic stability through DFT study. HOMO energy is directly proportional to the electron rich regions of molecule; in contrast the LUMO is directly proportional electron deficiency region. Both energies will define the reactive regions of a ligand molecule to form stable interactions with receptors. The respective HOMO and LUMO regions of CGA, MIM and MYR were depicted in 3-D mesh plots (Fig.3). Here blue and red colour regions represent the positive and negative phase of the 3-D orbital. Each molecules HOMO and LUMO energy and energy gap are also given in the same figure. It shows that CGA were secured reactive regions (both electron donating and accepting) in caffeic acid moiety but not in quinic acid moiety. MYR has the wide spread electron rich and deficient regions to form stable interactions with the receptor protein compared to other compounds. Further, the energy gap (between HOMO and LUMO energy) is directly proportional to the stability of drug molecule. Also it deciphers the intramolecular charge transfer between protein and ligand which indicating the biological importance of ligand⁴². The highest energy gap 4.8445 eV was procured by MIM and the lowest energy gap 3.872 eV was accomplished by MYR. From this, we can conclude that MIM and MYR were having the low chemical and high kinetic stability orderly.

The optimized geometries of these compounds corresponding to their minimum potential energy surface have been solicited by self-consistent field energy. The SCF (self-consistent field) energy of CGA, MIM and MYR (1297.5156a.u., 721.2631a.u and 1179.366 a.u. respectively) were calculated using the same method. HOMO, LUMO, energy gap and SCF energy of three compounds were listed in the Table.2.

Ligand-based Pharmacophore model

The ligand-based pharmacophore model was developed for these compounds (CGA, MIM and MYR) to identify the best pharmacophore featured compound using e-pharmacophore module of Maestro suite. The pharmacophore hypothesis was developed for single ligand based method and receptor van der Waals cut off value was kept as 0.5. MYR was gained better pharmacophore features (AADDRRR) than the others. The model hypothesis was consist of seven leading pharmacophore features (two acceptor; A, two donor; D and three aromatics rings; R15, R16 & R17) as shown in Fig.4-A. The inter-feature distances (in Å) are also mentioned in the same figure. Thus, the features and drug-likeness suggests that Myricetin or its derivative compounds could be a prominent base compound for structure-based drug design of eukaryotic Hyals. Consequently, the EHY-ligand complexes were subjected to MD simulation up to 10ns period towards investigating their interaction stability and conformational behaviour.

MD simulation Results

All trajectories were consistent with potential energy, temperature and kinetics during the simulation period. The RMSD graph of EHY (apo form) was deviated maximum to 3.3 Å than the ligated form; and this observation confirmed that the ligated form controlling the conformational flexibility of EHY (Fig.4-B). RMS fluctuations in N-terminal region (50-58 AA) and near the functionally important residues (TYR79, ASP133, GLU135, TYR206 and TYR284) were high in apo form due to devoid of any ligand (Ref. Fig.4-D). The regions loop (53-62 AA); β -8 strand & α -9 helix (372-405) and Loop-2 (87-101 AA); α -4 and α -5 helix fluctuations were restricted by CGA and MIM binding. Whereas the active site residues flexibility was largely arrested by the MIM and average less fluctuation was achieved by MYR.

The compactness in term of radius of gyration of EHY reduces gradually over time period and converged at 2.27 nm due to absence of any inhibitors in catalytic site. But EHY-ligands gyration was not converged and the globular nature of EHY was not revamp (Ref. ESI, Fig.2). The total solvent accessible surface area (SASA) of apo and ligand complexes was maintained within the range of 225-230 nm (Fig.4-C). This suggests that the ligand binding does not alter EHY surface charge, hydrophobic and hydrophilic property of EHY. The occupancy of hydrogen bonds between EHY-ligands were analysed and depicted in the figure (Ref. ESI, Fig.3-A). MIM and MYR were retained maximum of 5 and 4 hydrogen bonds and two of them were occupied more than 75% and 60% life span. The DFT results of these ligands were highly correlated with hydrogen bond occupancy with EHY.

The converged trajectories were subjected to LigPlus⁴³ and HBplus⁴⁴ analysis to find the stable non-bonded interactions within the cut-off distance of 4 Å. The list of residues and atoms involved in non-bonded interactions are tabulated (please refer in ESI, Table.3 for interaction details). Caffeic acid moiety of CGA was actively interacting with EHY whereas Quinic acid moiety was lacking of any interaction due to devoid of electron donating or accepting regions (Ref. Fig.3). MIM was secured strong electrostatic interaction within catalytic (ASP133 and GLU135) and hydrophobic interactions with TYR79 and TYR206 residues. Moreover MYR was preserved by TYR214, TYR253, ARG295 and TYR300 residue interactions and stably residing near β -5 strand and α -9 helix region. Functionally important residues (TYR79, ASP133, GLU135, ARG138 and TYR206) Φ and ψ angles were

distributed for the apo and ligated form of EHY and shown in figure (Refer in the ESI, Figure.3: B to F). The dihedral angle distributions were followed the similar observation as aforementioned MD results.

Energy Landscape of EHY-ligand complex

The global minimum energy conformation of protein-ligand can be derived from FEL analysis. If protein-ligand interaction was very weak or unstable, it can attain multiple minimum energy clusters; whereas the strong and stable interaction can bring almost a single conformation cluster in potential energy distribution. Selected ligands FEL contour maps were derived from PC1 to PC2 (as mentioned in methodology) and depicted in the Fig.5. The structural transition was occurred from the docking to the Gibbs minimum energy state of EHY. The corresponding initial to FEL conformational change of helices and backbone of EHY in ligated forms were compared and depicted in the figure (refer ESI, Fig.5).

The minimal energy conformation of CGA, MIM and MYR binding was attained with the backbone RMSD deviation of 2.753 Å, 1.12 Å and 1.2 Å from initial conformation. Herein, the large displacement regions after CGA (β -7; β -8 strand; helices α -1, α -10, α -11 & α -12), MIM (E-Loop; loop between β -8 and β -9) and MYR (E-Loop and β -7 and β -8 strands) binding was observed and highlighted in Fig.5. FEL interaction of CGA was stabilized by a hydrogen bond formed by ARG92 with 3.2 Å distance and further non-bonded interactions were established by TYR79, ASN67, ASN81, TRP328 and MET331 residues as shown in Fig.5-A. EHY-CGA FEL contour map was shown more than one cluster region due its moderate binding with EHY.

MIM was strongly held by the core catalytic region so that the whole catalytic region flexibility was controlled which influenced over the suppression of EGF-like domain motion. This supports that the positive correlated motion was observed between the catalytic and EGF-like domains and thus EGF-like domain could control the catalytic activity of EHY. FEL conformation of MIM was sustained by two hydrogen bond with ASN136 and HIS88 and strong hydrophobic interactions with TYR79, TYR206 and TYR293 residues. Additionally, a cation-pi interaction was acquired between N1 atom of MIM to the ring centre of TYR206. FEL of MIM were brought almost single cluster as shown in the FEL counter (Fig.5-B).

MYR interaction was retained by three hydrogen bonds with ASN67, ARG295 and MET331 and hydrophobic interactions with TYR79, ALA299 & TYR300 residues. Further, MYR was procured by cation-pi interaction between ARG295-NH2 to ring centre of R17 and pi-pi interaction formed between centres of TYR79 to R17 ring of MYR (Ref. Fig.5-C). FEL conformation of MYR was attained a smooth single basin in free energy counter map. From the FEL analysis, we can conclude that MIM and MYR were having durable interaction with EHY and they could be a preferable inhibitor for eukaryotic Hyals also because they have same kinetics of inhibition⁴⁵.

PCA analysis of EHY:

The motions responsible for the functional behaviours of protein can be determined by PCA analysis. The two dimensional projection of PC1 to PC2 (between largest two eigen values) in all forms were plotted and given in ESI, Fig.4. Here each dot represents the conformation of EHY in apo or ligated forms. More the conformational change, the spreading of distribution will be more in the conformational space. Here, the apo form distributions were widely scattered due to absence of ligand whereas the conformational space of EHY in ligated forms were more rigid.

The structural and conformational rearrangements of EHY were investigated from the PCA analysis as mentioned earlier. The mode vector cut-off for the functional mode changes of EHY was considered as 3.5 Å distance. Due to over fluctuation of C-terminal region they were not included for the mode vector analysis. The average to maximum transformation of EHY backbone in apo, CGA, MIM and MYR bound forms were founded with 1.7 Å, 1.48 Å, 1.54 Å and 2 Å RMS deviations as the result of ligand induced functional mode change. The mode vectors of unligated and ligated forms were depicted in Fig.6 with magnitude and direction of vector [A-EHY (apo form), B-EHY1, C-EHY2 and D-EHY3].

Close-Open conformation of EHY

The large structural translocation of Loop-III, E-Loop region and EGF-like domain were noticed in both unligated and ligated forms. Functionally important residues maximum displacements by means of C α atom were given in ESI, Fig.6 with distances (in Å). The closed conformation of EHY was achieved by the close proximity of loop regions surrounded to catalytic domain chiefly by Loop-III and E-Loop. The distances between ALA299 to ARG92 (14.37 Å), ASN67 to ARG92 (15.56 Å), MET331 to R92 (15.25 Å) and SER260 to LYS219 (9.82) residues were approached as much as nearer towards the closed conformation. Interestingly, in open-I conformation, it was formed by the notable dislocation of above mentioned residues (ALA299 to ARG92: 22.12 Å, ASN67 to ARG92: 19.48 Å, MET331 to R92: 22.15 Å and SER260 to LYS219: 19.56 Å) as the result of ligand binding. In this, SER260 to LYS219 dislocated nearly 10 Å (as shown in Fig.7) which may block the entry of large HA fragment to the EHY cavity. Whereas, the open-II conformation was most similar like open-I conformation (ALA299 to ARG92: 23.54 Å, ASN67 to

ARG92: 19.55 Å, MET331 to R92: 20.93 Å) but the SER260 to LYS219 distance was retained as in the closed conformation. These observations leads to the postulation that open-I and open-II conformations can be attained by entry of small and large HA fragment binding respectively. The overall structural rearrangement has brought with a closed and two open conformations of EHY. In conclusion, Loop-III (notably ARG92 AA) and E-loop (notably LYS219 AA) region can play essential role in catalytic dynamic behaviour of EHY.

Conclusions

Hyalys are a crucial carbohydrate-active enzyme and they have wide biological role in tumour growth, metastasis, angiogenesis and cell proliferation. Specifically Human Hyal isoforms are identified as the diagnostic marker as they expressed in many cancer cells. Likewise, Venom Hyalys are also an important weapon for the venom spreading which primitively targets for the fast action of the toxic proteins.

The phylogenetic analysis of viper Hyal revealed the existence of two different exons or post-translation scission regions and two functional group of Hyal. Significant difference was observed in the E-loop region of the EHY model built in comparison with template. We propose that this region may show unique functional property. Drug-like molecular property and stability of CGA, MIM and MYR were calculated in terms of HOMO and LUMO using DFT method. Interaction stability and free energy landscape of each ligand revealed that MIM and MYR can be prominent inhibitors for eukaryotic Hyal since they follow similar kinetics of inhibition. High flexibility and structural changes in the Loop-III and E-loop regions results due to the ligand binding; and we propose that this flip-flop may be due to the inhibitor inducement. From the structural and conformational rearrangements, HA large and small fragment binding conformers were predicted and possible closed and open conformation of EHY was speculated from PCA analysis.

The results strongly supports the inhibitory activity of ligands and further it provides spectacle about the ligand induced dynamic behaviour of viper Hyaluronidase. These finding will roadmap to target eukaryotic Hyal as forthcoming important drug target in cancer and venom spreading.

Abbreviations

Hyal - Hyaluronidase, EHY - *Ehics pyrimedium Leakyi's* Hyaluronidase, CGA - Chlorogenic acid, MIM - Mimosine, MYR - Myricetin, ECM - Extra cellular matrix, NAG - N-acetyl Glucosamine, GUA - Glucuronic acid, GAG - Glycosaminoglycan, EGF - Epiderman growth factor, DOPE - discrete optimized protein energy, SAVES - structural analysis and verification server, E-Loop - Extend loop, MIM-GBSA -Molecular mechanics based Generalized Born/surface area, BLY3P - Lee-Yang-Parr correlation functional theory, HOMO - Highest occupied molecular orbital, LUMO - Lowest unoccupied molecular orbital, EHY1 - EHY-Chlorogenic acid complex, EHY2 - EHY-Mimosin complex, EHY3 - Myricetin complex, SASA -Solvent accessible surface area, SPC - Simple point charge, FEL - Free energy landscape, PCA - Principal component analysis.

Acknowledgements

We thank Centre for Bioinformatics, Pondicherry University for the computational facilities and UGC, Government of India for their financial support for the research work. BSI thank SERB for funding the project. We also thank Dr. Kannan, M., Centre for Bioinformatics and Dr. Baskaran, S., Department of Chemistry, Pondicherry University for their valuable suggestion and support throughout this work.

Notes

The complete reference of 34 is available in ESI.

References

1. G. Kreil, *Protein Sci.*, 1995, 4, 1666.
2. R. Stern and M. J. Jędrzejak, *Chem Rev.*, 2006, 106, 818-839.
3. G. Davies and B. Henrissat, *Structure*, 1995, 3, 853-859.
4. C. Termeer, J. P. Sleeman and J. C. Simon, *Trends Immunol.*, 2003, 24, 112-114.

5. B. P. Toole, *Nat Rev Cancer.*, 2004, 4, 528-539.
6. D. C. West, I. N. Hampson, F. Arnold and S. Kumar, *Science*, 1985, 228, 1324-1326.
7. P. W. Noble, *Matrix Biol.*, 2002, 21, 25-29.
8. M. J. Jedrzejak, L. V. Mello, B. L. de Groot and S. Li, *J Biol Chem.*, 2002, 277, 28287-28297.
9. K. L. Chao, L. Muthukumar and O. Herzberg, *Biochemistry*, 2007, 46, 6911-6920.
10. Z. Marković-Housley, G. Miglierini, L. Soldatova, P. J. Rizkallah, U. Müller and T. Schirmer, *Structure*, 2000, 8, 1025-1035.
11. S. Arming, B. Strobl, C. Wechselberger and G. Kreil, *Eur J Biochem.*, 1997, 247, 810-814.
12. J.-X. Tan, X.-Y. Wang, X.-L. Su, H.-Y. Li, Y. Shi, L. Wang and G.-S. Ren, *PLoS One*, 2011, 6, e22836.
13. M. Ramanaiah, P. Parthasarathy and B. Venkaiah, *Biochem Int.*, 1989, 20, 301-310.
14. G. Lu and L. Kochoumian, *J Biol Chem.*, 1995, 270, 4457-4465.
15. R. Wright, K. Elgert, B. Campbell and J. Barrett, *Arch Biochem Biophys.*, 1973, 159, 415-426.
16. C. Poh, R. Yuen, M. Chung and H. Khoo, *Comp. Biochem. Physiol., B*, 1992, 101, 159-163.
17. K. Girish, R. Shashidharamurthy, S. Nagaraju, T. V. Gowda and K. Kemparaju, *Biochimie*, 2004, 86, 193-202.
18. K. Girish and K. Kemparaju, *Biochemistry (Mosc)*. 2005, 70, 948-952.
19. M. Sebastin Santhosh, M. Hemshekhar, K. Sunitha, R. M Thushara, S. Jnaneshwari, K. Kemparaju and K. S Girish, *Mini Rev Med Chem.*, 2013, 13, 106-123.
20. M. A. Larkin, G. Blackshields, N. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm and R. Lopez, *Bioinformatics*, 2007, 23, 2947-2948.
21. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, *Mol Biol Evol.*, 2011, 28, 2731-2739.
22. D. T. Jones, W. R. Taylor and J. M. Thornton, *Comput Appl Biosci.*, 1992, 8, 275-282.
23. A. Šali and T. L. Blundell, *J Mol Biol.*, 1993, 234, 779-815.
24. R. A. Harrison, F. Ibison, D. Wilbraham and S. C. Wagstaff, *Gene*, 2007, 392, 22-33.
25. R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Cryst.*, 1993, 26, 283-291.
26. C. Colovos and T. O. Yeates, *Protein Sci.*, 1993, 2, 1511.
27. R. Liithy, J. Bowie and D. Eisenberg, *Nature*, 1992, 356, 83-85.
28. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J Comput Chem.*, 2005, 26, 1701-1718.
29. M. Kawata and U. Nagashima, *Chem. Phys. Lett*, 2001, 340, 165-172.
30. B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comp. Chem.*, 1997, 18, 1463-1472.
31. S. Miyamoto and P. A. Kollman, *J. Comp. Chem.*, 1992, 13, 952-962.
32. H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, *J. Chem. Phys.*, 1984, 81, 3684-3690.
33. M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, 52, 7182-7190.
34. M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb and J. Cheeseman, *Gaussian 09, Revision D. 01; Gaussian: Wallingford, CT, USA*, 2009.
35. P. M. Gill, B. G. Johnson, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett*, 1992, 197, 499-505.
36. H. J. Berendsen and S. Hayward, *Curr Opin Struct Biol.*, 2000, 10, 165-169.
37. A. Amadei, A. Linssen and H. Berendsen, *PROTEINS:Structure, Function, and Genetics*, 1993, 412-425.
38. W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics* 1996, 14, 33-38.
39. J. Wang and G. M. Verkhivker, *Phys Rev Lett.*, 2003, 90, 188101.
40. K. Muthu, M. Panneerselvam, N. S. Topno, M. Jayaraman and K. Ramadas, *Cell Signal.*, 2015, 27, 739-755.
41. V. B. Lokeshwar, G. L. Schroeder, R. I. Carey, M. S. Soloway and N. Iida, *J Biol Chem.*, 2002, 277, 33654-33663.
42. L. Padmaja, C. Ravikumar, D. Sajan, I. Hubert Joe, V. Jayakumar, G. Pettit and O. Fauriskov Nielsen, *J. Raman Spectrosc.*, 2009, 40, 419-428.
43. R. A. Laskowski and M. B. Swindells, *J Chem Inf Model.*, 2011, 51, 2778-2786.
44. I. K. McDonald and J. M. Thornton, *J. Mol. Biol.*, 1994, 238, 777-793.
45. K. Mio and R. Stern, *Matrix Biol.*, 2002, 21, 31-37.
46. M. W. Kramer, R. Golshani, A. S. Merseburger, J. Knapp, A. Garcia, J. Hennenlotter, R. C. Duncan, M. S. Soloway, M. Jorda and M. A. Kuczyk, *Eur Urol*, 2010, 57, 86-94.

Figure Captions

Fig.1: A) The tree with highest likelihood value of -2277.6254 is depicted. Phylogentic relationship among selected sequences (Uniprot ID and short name of each species were mentioned) was inferred by Maximum-likelihood method based on JTT-matrix model using MEGA5 and MSA was performed using ClustalX. Bootstrap score values are given at each node and branch length (substitution/mutation rate per amino acid) is given below as bar scale (0.02). B) Three dimension structure of EHY is shown in the figure. Catalytic domain region loops are in highlighted in green colour and EGF-like domain loops are highlighted in blue colour and helix and β -sheet are in red and cyan colour respectively. The functionally important Loop III and E-loop region are highlighted in magenta colour.

Fig.2: EHY-ligand dock pose and interactions are depicted in this figure. a) The ligand (yellow sticks) occupied site of EHY (in surface) is highlighted in circle. The hydrogen bond interactions with b) CGA, c) MIM and d) MYR are shown with the bond length distances (in Å). Here Helix, β -strand and loop is in red, blue and plum colour respectively.

Fig.3: Frontier molecular orbital HOMO (H) and LUMO (L) of CGA (I), MIM (II) and MYR (III) are calculated using DFT B3LYP/ 6-311G* method and their corresponding electron densities are shown in mesh form. Each forms orbital energy and energy gap is denoted adjacent to the corresponding figures. Here green and red represent the positive and negative phase of three dimensional atomic orbital.

Fig.4: Pharmacophore model hypothesis of MYR with AADDRRR feature is depicted in figure (A) along with inter-feature distances in Å. (A-Hydrogen acceptor, D-Hydrogen donor and R-Aromatic ring). MD simulation analysis of EHY flexibility and compactness in term of B: RMSD, C: total SASA and D: RMSF are shown. The color representation for each form is denoted inside the legend box. EHY-*apo*, EHY1-CGA, EHY2-MIM and EHY3-MYR bound form of EHY.

Fig.5: Two dimensional FEL contour map of CGA (A), MIM (B) and MYR (C) bound form of EHY are plotted against PC1 to PC2 using Mathematica. Each ligands corresponding minimum energy conformer are shown with the interacting residue and bond distances in Å. Here color code represents magenta-charge, green-main and blue-side chain hydrogen bonds.

Fig.6: Functional mode changes of EHY backbone in EHY-*apo* (A), EHY1 (B), EHY2 (C) and EHY3 (D) forms are depicted along with mode vector direction and amplitude strength (length of vector).

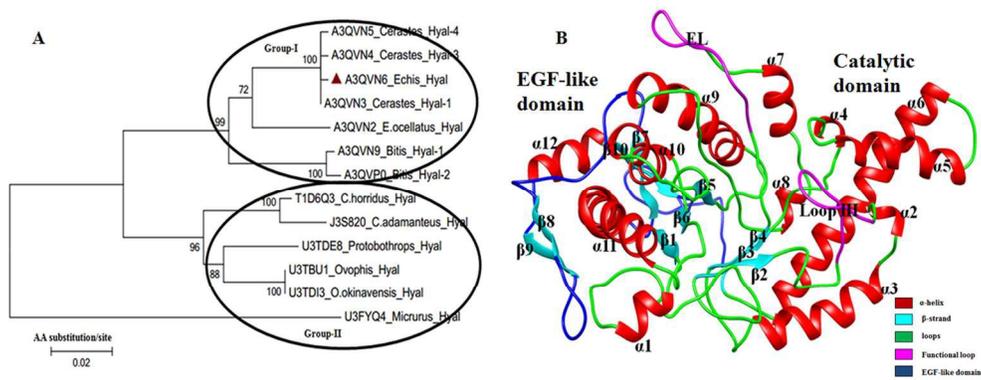
Fig.7: The postulated close to open conformation of EHY is derived from the PCA analysis. Close and Open-I conformations are represented in red colour while Open-II is shown in dark brown colour. The loop residues around catalytic site are labeled and distances between them are given in Å.

Table 1

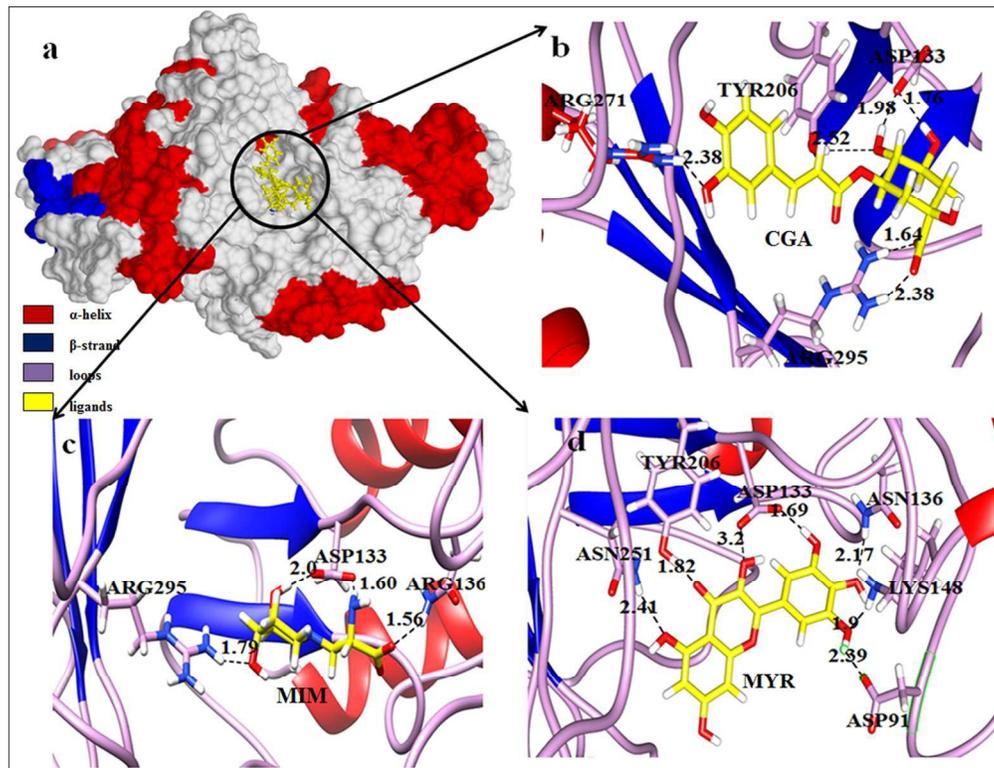
Ligand	Interacting Residues	Glide energy	Glide emodel	MM-GBSA (ΔG) kcal/mol
Chlorogenicacid	Asp133,Tyr214, Arg271 and Arg295	-5.7	-53.56	-37.44
Mimosine	Asp133,Asn136 and Arg295	-5.6	-50.64	-34.22
Myricetin	Asp91,Asp133,Asn136 ,Lys148,Tyr206 and Asn251	-5.5	-53.50	-35.78
Catechin	Asp91, Asp133 and Lys148	-4.5	-42.37	-21.53
Kaempferol	Asp133, Arg271 and Arg295	-4.3	-38.25	-22.91

Table 2

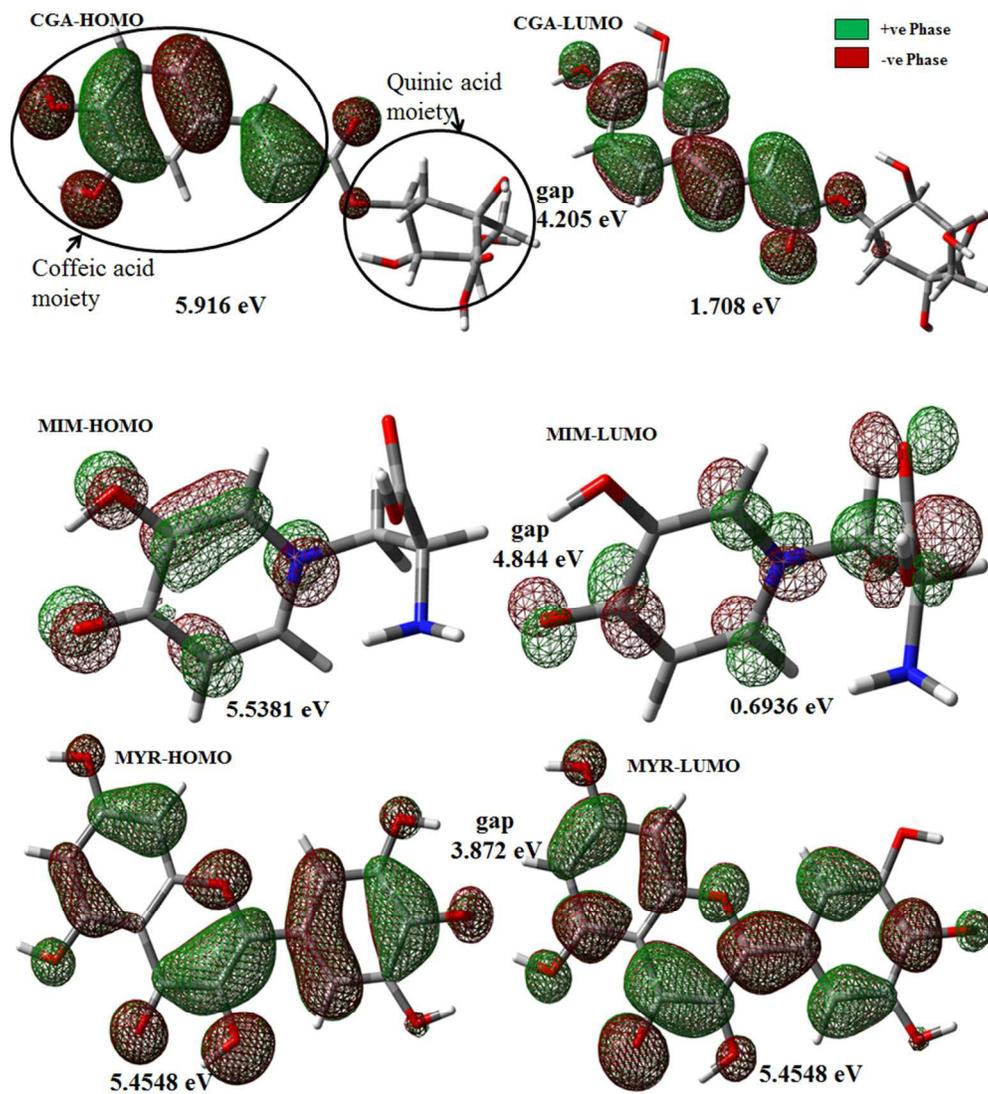
Name of the compound	HOMO energy (eV)	LUMO energy (eV)	Energy gap (eV)	SCF energy (a.u)
Chlorogenicacid (CGA)	5.913	1.708	4.205	1297.5156
Mimosine (MIM)	5.5381	0.6936	4.845	0721.2631
Myricetin (MYR)	5.4548	1.5828	3.872	1179.366

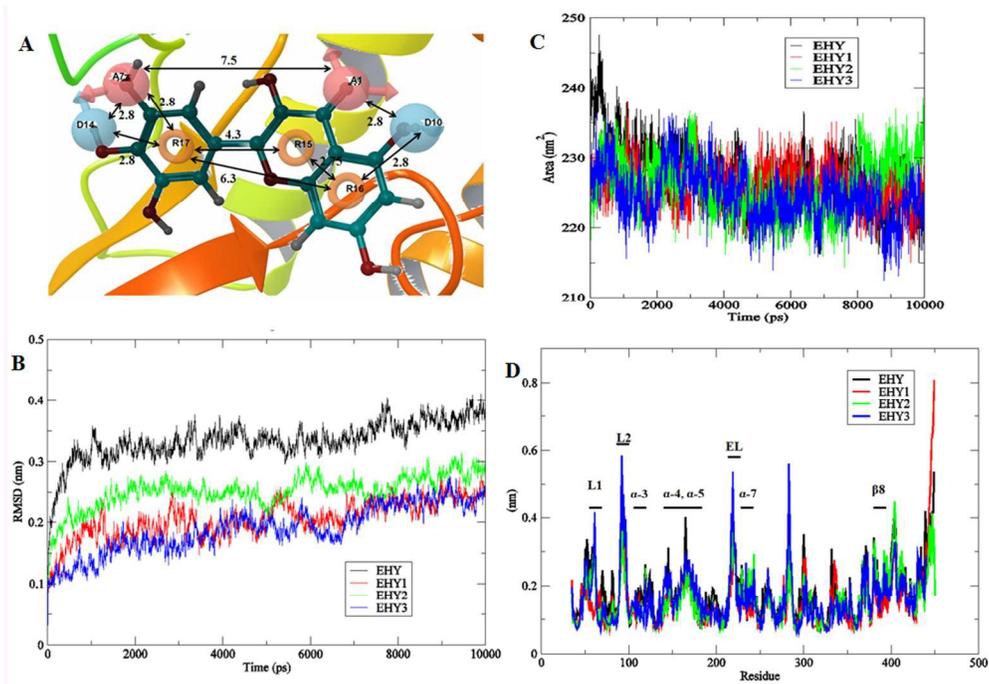


346x134mm (300 x 300 DPI)

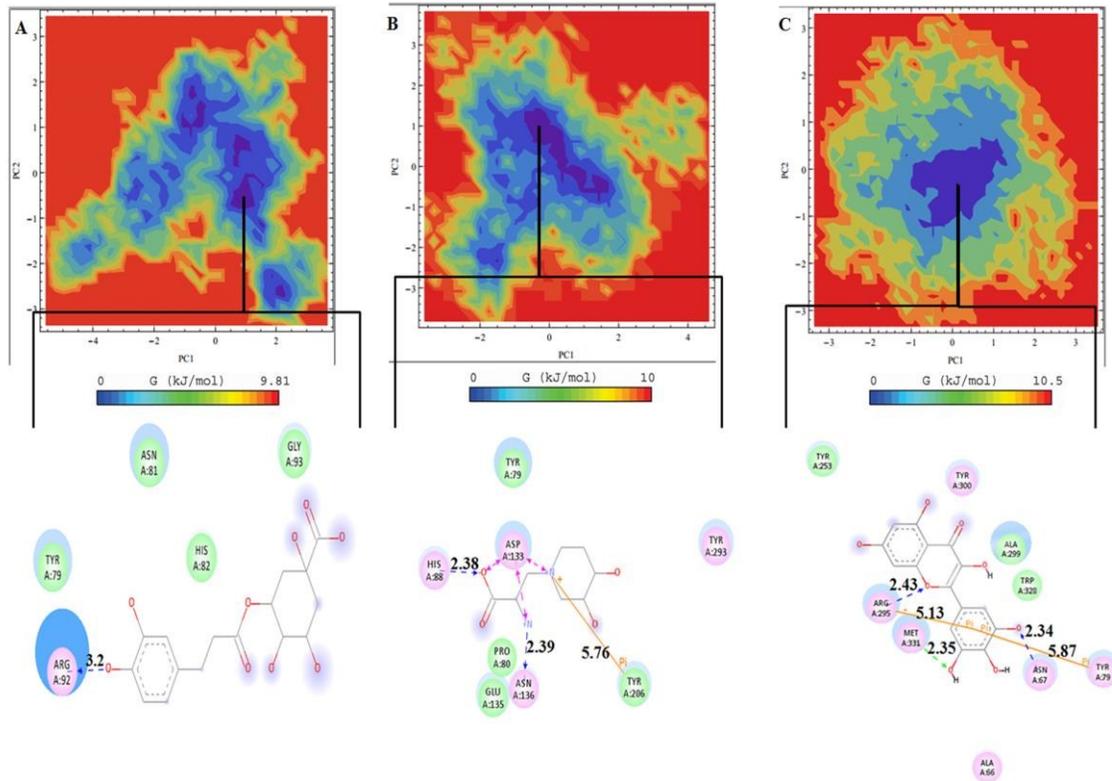


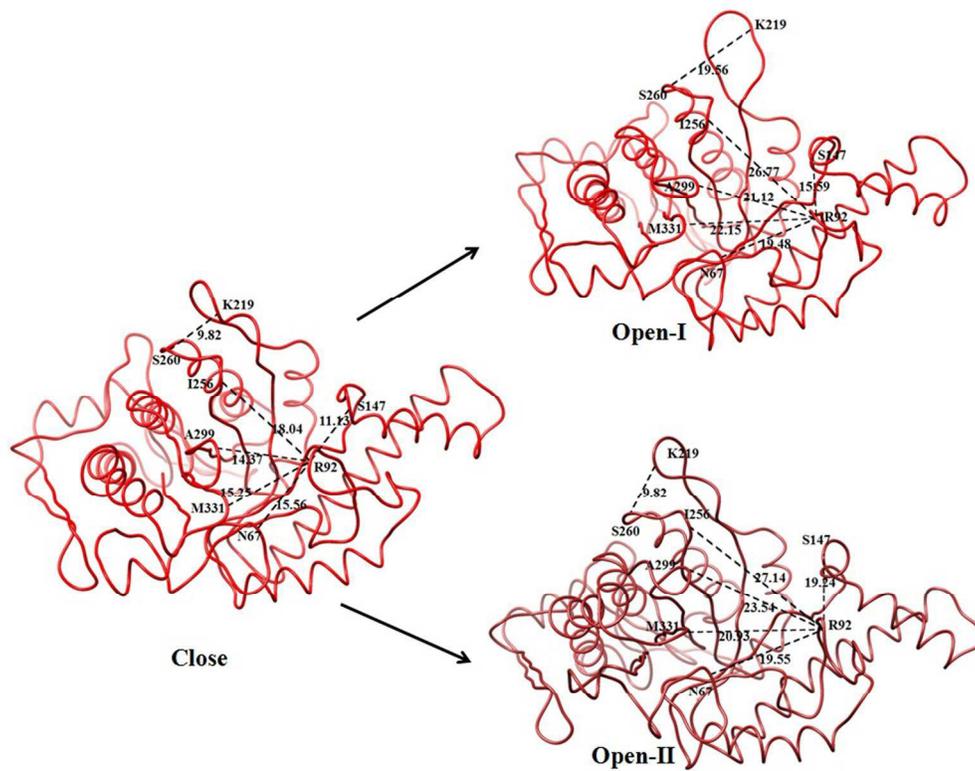
131x101mm (300 x 300 DPI)



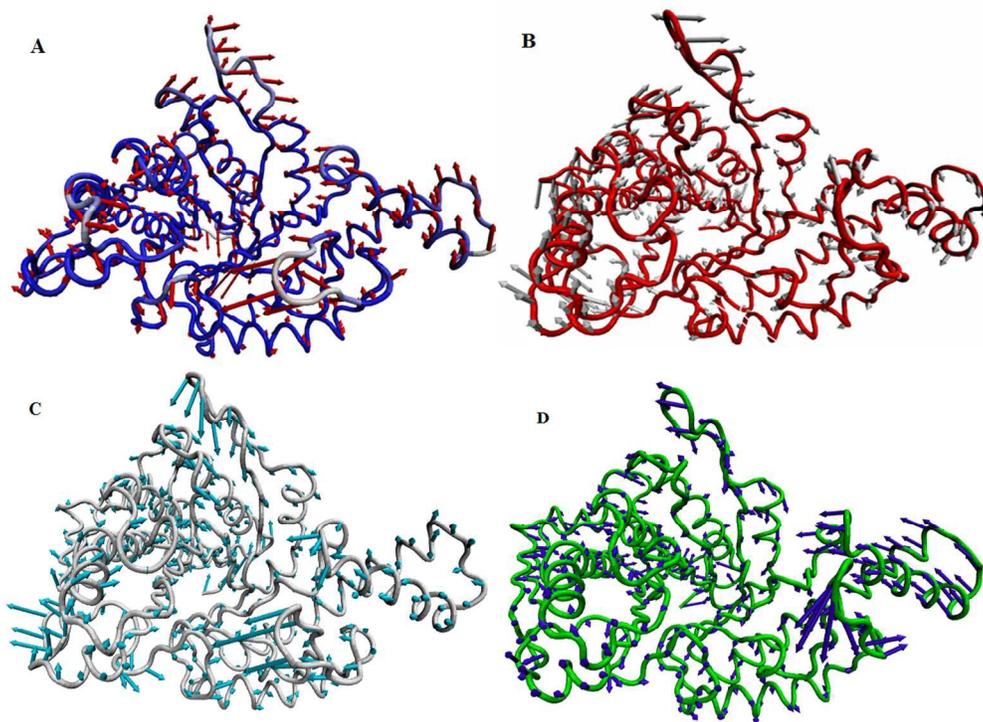


290x201mm (300 x 300 DPI)





205x163mm (300 x 300 DPI)



186x136mm (300 x 300 DPI)