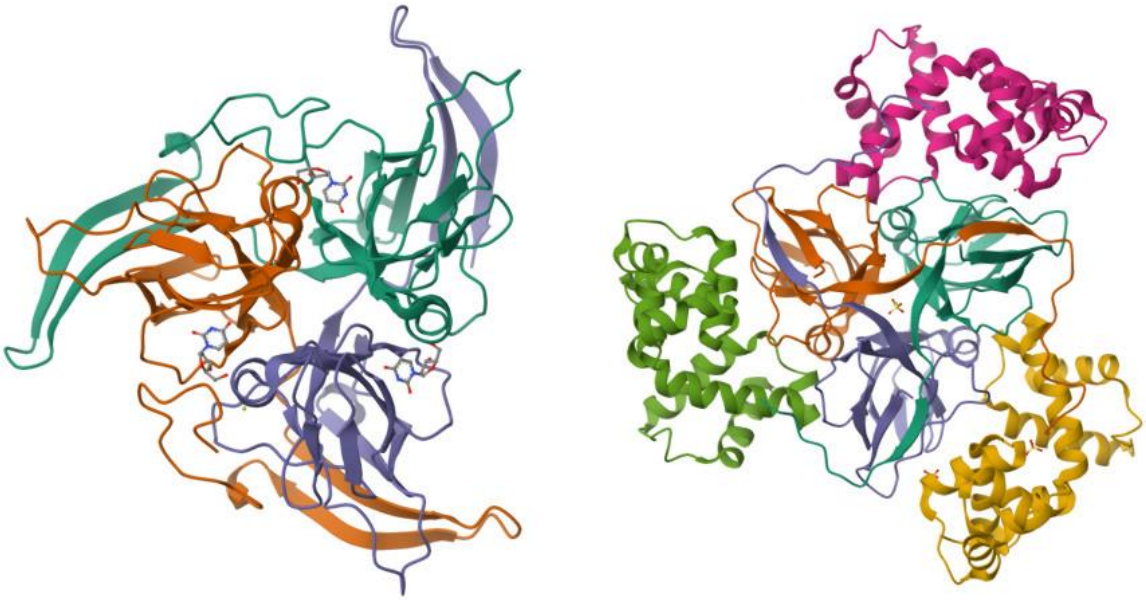
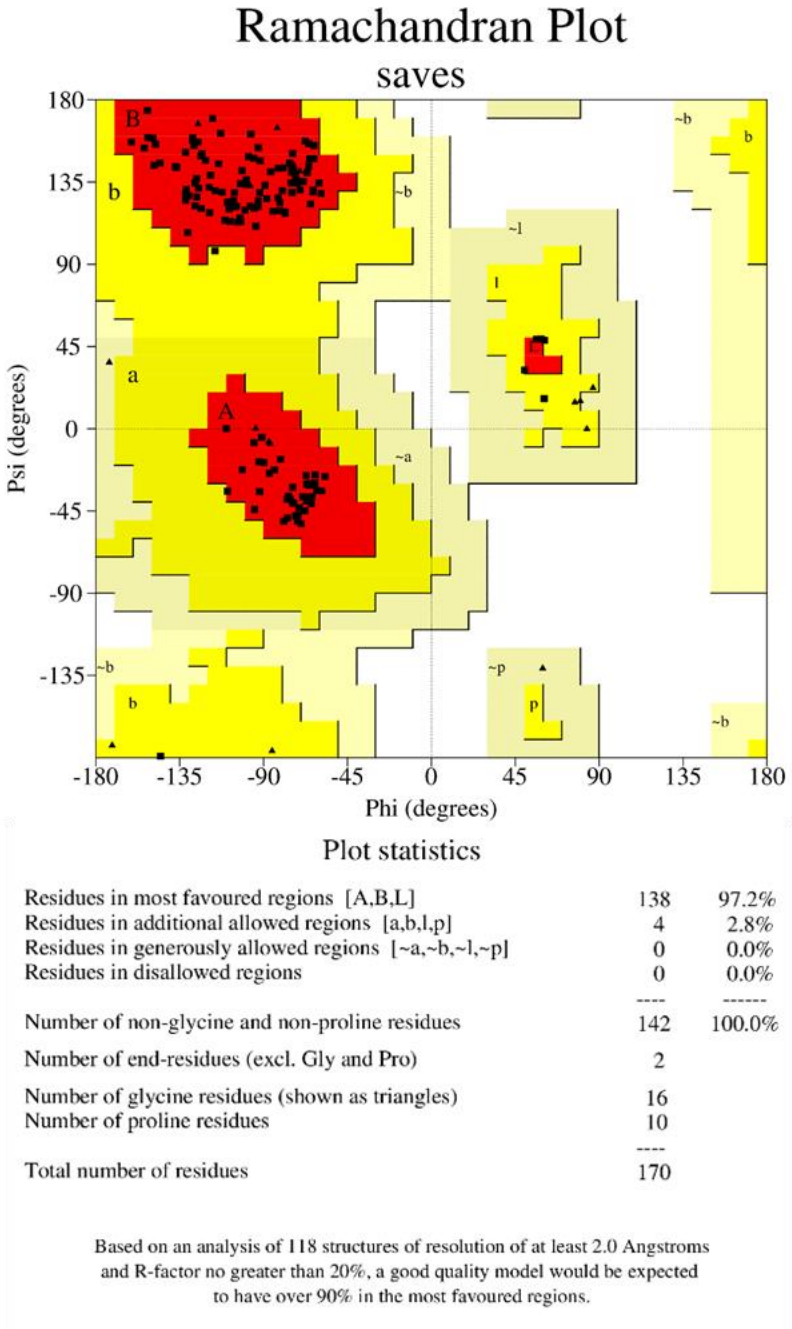
Biological Assembly of dUTPase



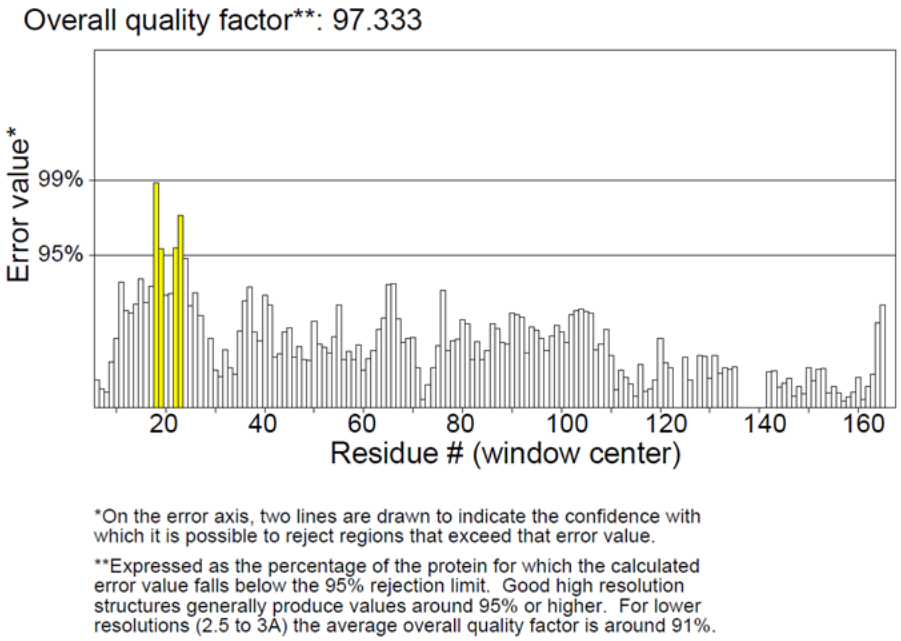
5Y5P WSSV dUTPase **7DLV Shrimp dUTPase**

Appendix 1. The shrimp dUTPase (7dLV) and the White Spot Syndrome Virus (5Y5P) dUTPase are enzymes that have a vital roles in nucleotide metabolism by hydrolyzing dUTP to dUMP. In spite of their similar functions, these enzymes reveal differences in their structural and sequence characteristics. While both share a trimeric structure, there are differences in surface residues, amino acids length, theoretical weight and others characteristics. The distinctive β-hairpin structure in the domain-swapping region, which facilitates a unique orientation of adjacent C-terminal segment that re-position the catalytic motif V. Thus, the shrimp enzyme employs three subunit active sites whereas WSSV enzyme have two-subunit active sites.



Appendix 2. Refined dUTPase protein Ramachandran plot. 97.2% of amino acid residues are present in the core region (red color).

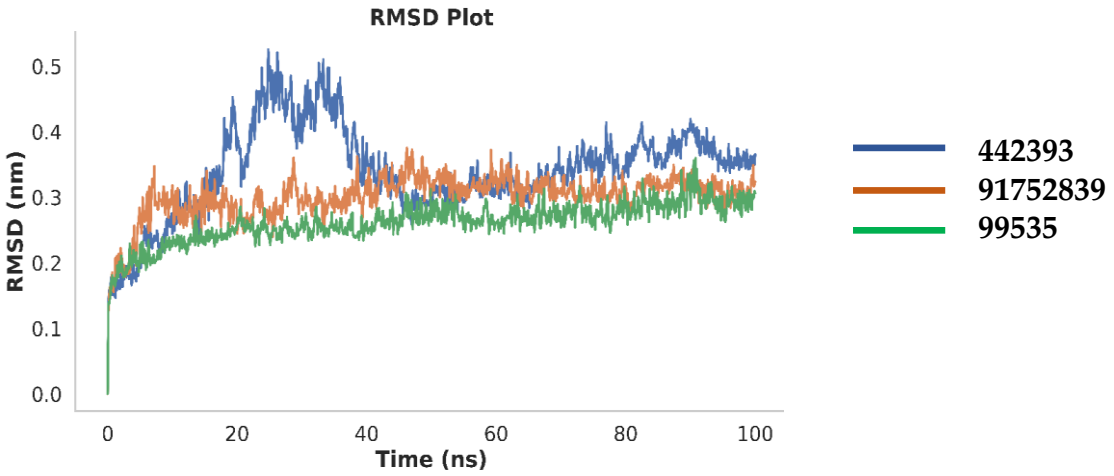
The Ramachandran plot is predominantly empirical with the backbone dihedral angles (φ and ψ) of a protein. Consequently, heteroatoms in side chains and water molecules are usually excluded from the plot. These elements have indirect influence on the plot by affecting the steric and energetic constraints of the polypeptide chain. When generating a Ramachandran plot, the focus is on the intrinsic conformational preferences of the backbone, therefore water molecules and heteroatoms are excluded from the plot.



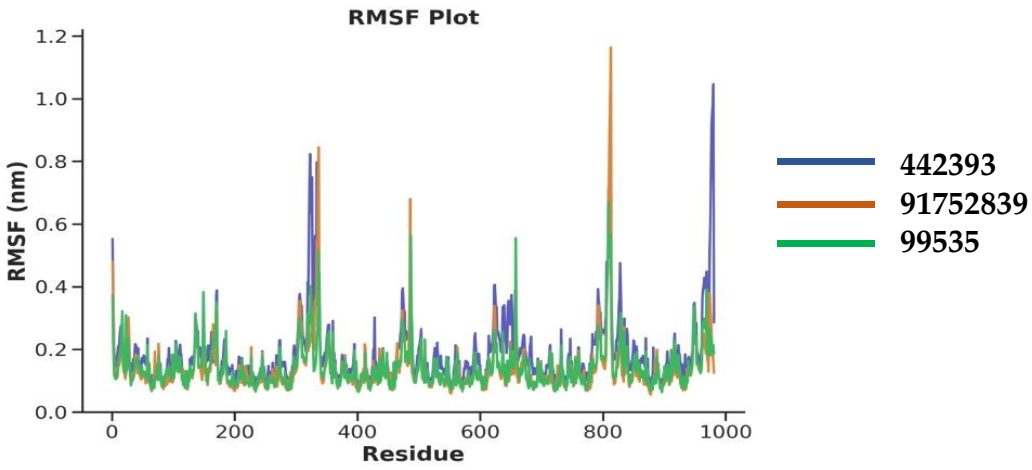
*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.

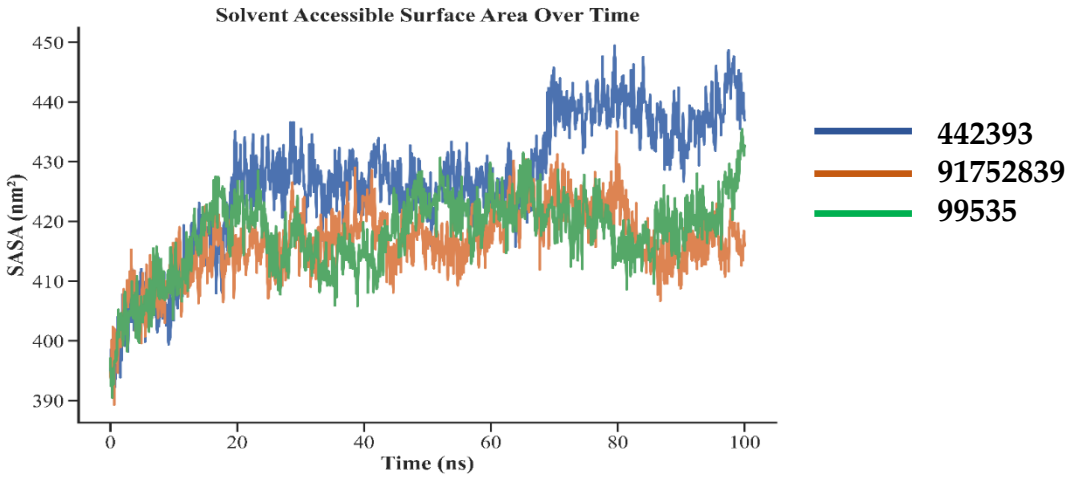
Appendix 3. Validation of refined protein using ERRAT score. The Y-axis represents the error value, and the X-axis represents the amino acid residues of the protein model.



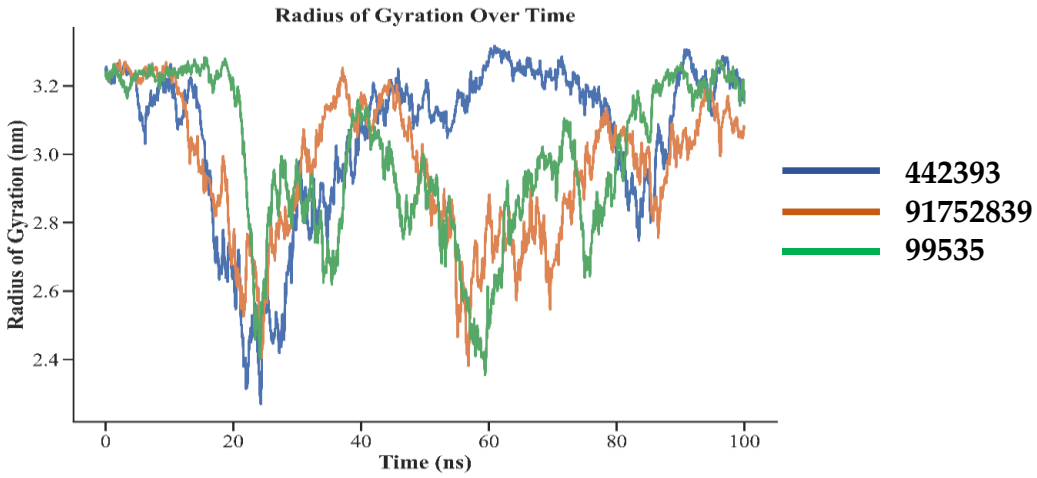
Appendix 4. The RMSD plot of WSSV dUTPase complexed with three different ligands, namely, 442393-blue line, 91752839-orange line and 99535-green line, during a 100-ns molecular dynamics simulation.



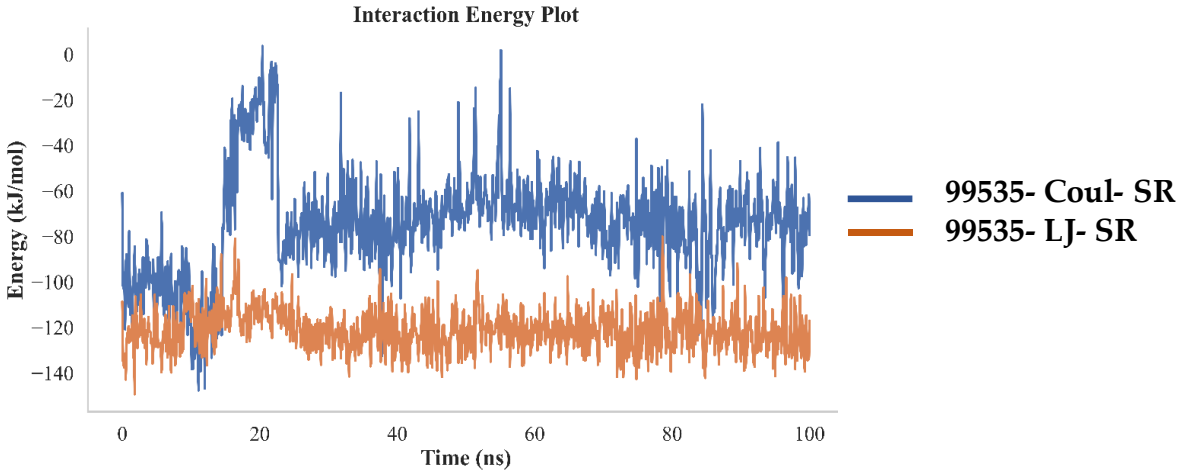
Appendix 5. The RMSF plot for per-residue flexibility of the dUTPase protein complexed with three ligands, namely, 442393, 5281600, and 99535represented respectively by blue, orange, and green lines at the end of the simulation.



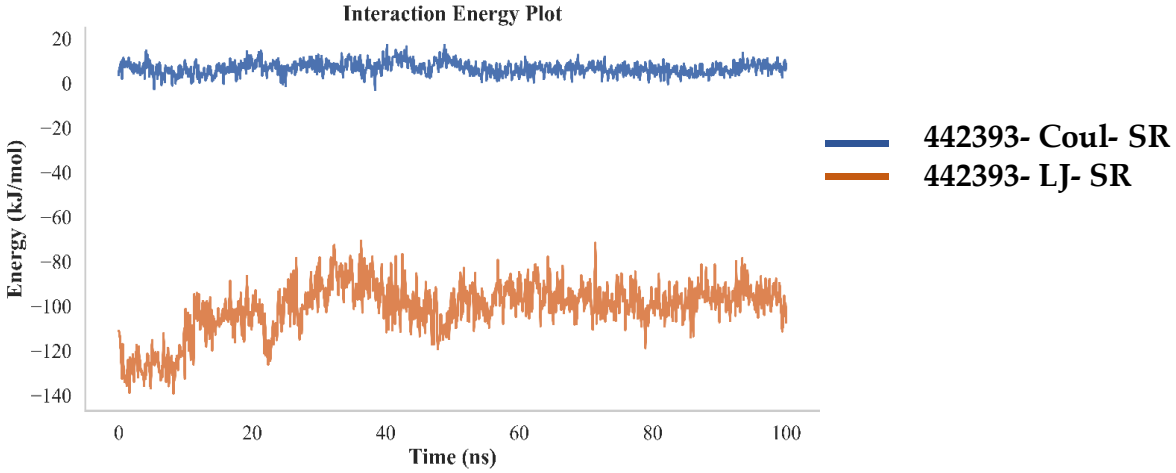
Appendix 6. Description of the SASA plot of the dUTPase protein exposed to the solvent molecules during interaction with the three ligands: 442393-blue line, 91752839-orange line and 99535-green line during a 100 ns.



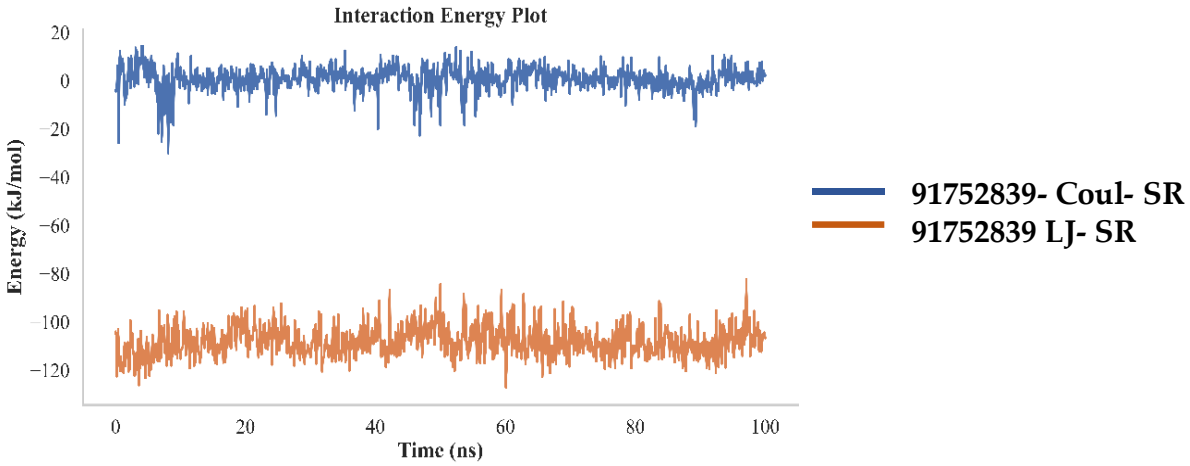
Appendix 7. Illustrat**ion of** the Radius of Gyration plot, representing the compactness of the dUTPase protein structure with respect to time upon binding with three ligands 442393 (blue line), 91752839 (orange line), and 99535 (green line), during a 100 ns.



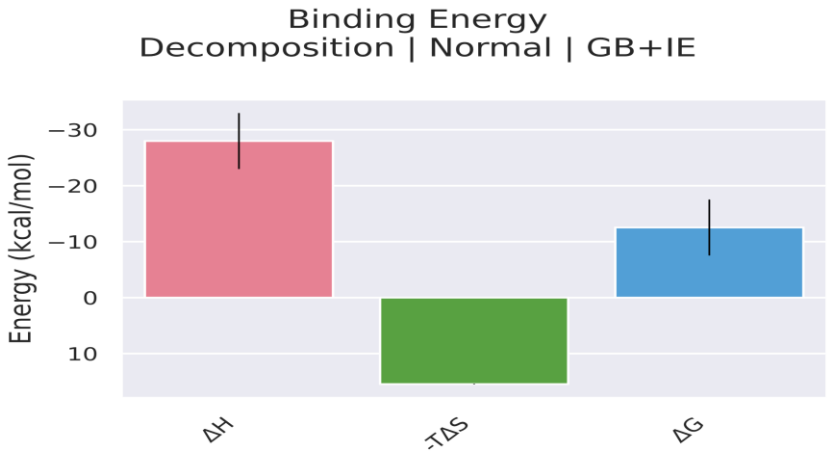
Appendix 8. The interaction energy plots of the Coulomb short-range and Lennard-Jones short-range interaction energies between the dUTPase protein and ligand 99535.



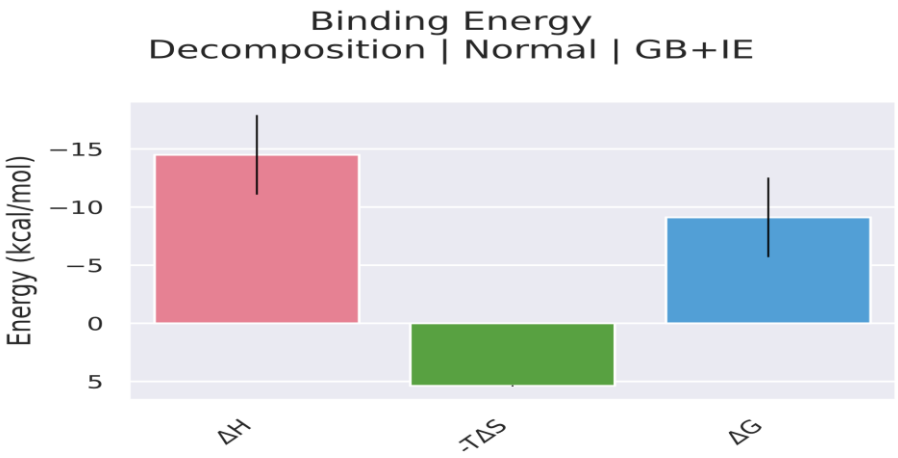
Appendix 9. The interaction energy plots of the Coulomb short-range and Lennard-Jones short-range interaction energies between the dUTPase protein and ligand 442393.



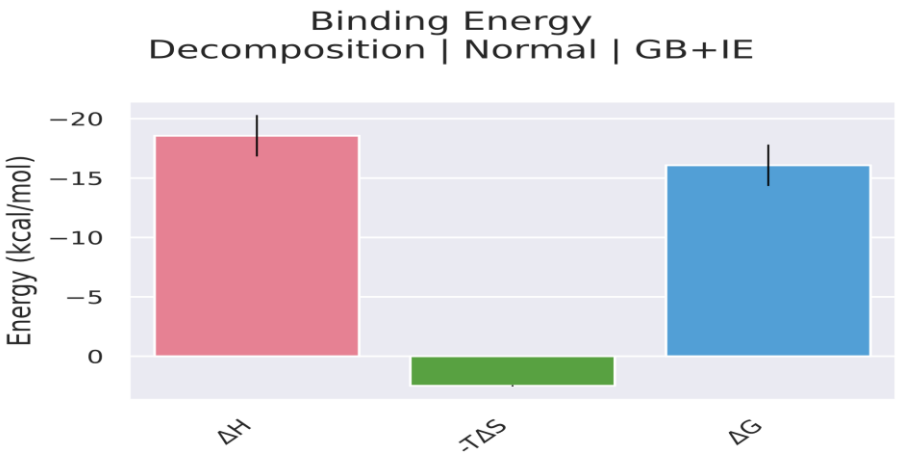
Appendix 10. The interaction energy plots of the Coulomb short-range and Lennard-Jones short-range interaction energies between the dUTPase protein and ligand 91752839



Appendix 11. The binding energy decomposition for dUTPase protein-ligand complexes and displays the contribution of enthalpy ΔH, entropy -TΔS, and Gibbs free energy ΔG to the binding affinity of ligand 99535.



Appendix 12. The binding energy decomposition for dUTPase protein-ligand complexes and displays the contribution of enthalpy ΔH, entropy -TΔS, and Gibbs free energy ΔG to the binding affinity of ligand 442393.

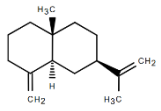
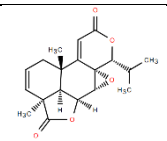
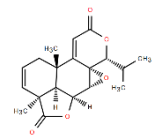


Appendix 13. The binding energy decomposition for dUTPase protein-ligand complexes and displays the contribution of enthalpy ΔH, entropy -TΔS, and Gibbs free energy ΔG to the binding affinity of ligand 91752839.

Appendix 14. Main differences between WSSV and shrimp dUTPase

Criteria	Shrimp dUTPase (7dlv)	WSSV dUTPase (5y5p)
Chains	A, B, C, G, H, I	A, B, C, D, E, F
Length	149 amino acids	174 amino acids
Theoretical weight	16.3 KDa	19.03 KDa
Accessible surface area	34507.71 Å2	19548.41 Å2
Buried surface area	19901.73 Å2	13793.96 Å2
Dissociation area	71.19 Å2	5,795.71 Å2
Dissociation energy (ΔGdiss)	11.85 kcal/mol	71.3 kcal/mol
Dissociation entropy (TΔSdiss)	0.05 kcal/mol	26.88 kcal/mol
Ligand binding sites	SO4, CA	MG, POP, DUR

Appendix 15. Top three phytochemicals and their binding affinities as well as other characteristics.

SI No.	PubChem ID	Chemical Name (Source)	Molecular Formula	Chemical structure	Molecular weight (g/mol)	Binding Affinity (Kcal/mol)
1	CID 442393	Selinene	$C_{15}H_{24}$		204.35	-9.3
2	CID 99535	Podolide	$C_{19}H_{22}O_5$		330.4	-8.8
3	CID 91752839	Zierone	$C_{15}H_{22}O$		218.33	-7.8

Appendix 16. In Silico ADME results of the top, three compounds.

Property	Model name (Unit)	alpha-Selinene (CID: 442393)	Podolide (CID: 99535)	Zierone (CID: 91752839)
Physicochemical Properties	Num. rotatable bonds	1	1	0
	Num. H-bond acceptors	0	5	1
	Num. H-bond donors	0	0	0
	Molar Refractivity	68.78	85.10	69.24
Lipophilicity	Log $P_{o/w}$ (iLOGP)	3.31	2.54	2.90
Absorption (Pharmacokinetics)	Water solubility (log mol/L)	-6.074	-3.871	-4.187
	Caco2 cell permeability (log Papp in 10^{-6} cm/s)	1.401	1.125	1.205
	Gastrointestinal (GI) absorption (human) (% Absorbed)	94.127	99.833	96.888
	Skin Permeability (log Kp)	-1.461	-3.363	-2.08
	P-glycoprotein substrate Categorical (Yes/No)	No	No	No
	P-glycoprotein I and II inhibitor (Yes/No)	No	No	No
Distribution (Pharmacokinetics)	BBB permeability (log BB)	0.776	-0.374	0.593
	CNS permeability (log PS)	-1.865	-2.89	-2.305
	Fraction unbound (human) (Fu)	0.186	0.304	0.301
	Volume of distribution (VDss) (log L/kg)	0.686	0.237	0.38
Metabolism (Pharmacokinetics)	CYP2D6 substrate (Yes/No)	No	No	No
	CYP1A2 inhibitor (Yes/No)	No	No	No
	CYP2C19 inhibitor (Yes/No)	No	No	Yes
	CYP2C9 inhibitor (Yes/No)	No	No	No
Excretion (Pharmacokinetics)	Total Clearance (log ml/min/kg)	1.172	0.788	1.199
	Organic cation transporter 2 (OCT2) (Categorical (Yes/No)	No	No	No
Drug likeness	Lipinski (Rule of five)	Yes; 1 violation: MLOGP>4.15	Yes; 0 violation	Yes; 0 violation
	Veber rule	Yes	Yes	Yes
	Bioavailability score	0.55	0.55	0.55
	Ghose filter	Yes	Yes	Yes
Medicinal Chemistry	Leadlikeness rule	No; 2 violations: MW<250, XLOGP3>3.5	Yes	No; 1 violation: MW<250,
	Synthetic accessibility	4.22	5.80	4.46

Appendix 17. The points of reference in Lipinski's rule of five.

No.	Rule	Criteria/ thresholds
1	Molecular Weight (MW)	**<500 g/mol**
2	Number of Hydrogen Donors	**<5**
	Number of Hydrogen Acceptors	**<10**
3	Log of 1-octanol/water partition co-efficient (neutral form) LogP	**<5**
4	Total Polar Surface Area (PSA)	**<140Å**
5	Number of Rotatable Bonds (RB)	**<10**

Appendix 18. The points of reference in Toxicity.

No.	Target	Criteria/ thresholds
1	Hepatotoxicity	Inactive to Slightly active
2	Carcinogenicity	Inactive
3	Mutagenicity	Inactive
4	Cytotoxicity	Inactive
5	LD50 (mg/kg)	**> 500**

Appendix 19. The Competitive Interactions between candidate phytochemicals and the dUTP.

SI No.	PubChem ID	Chemical Name (Source)	Binding Affinity (Kcal/mol)	Binding Site	Interactions
1	CID 65070	dUTP	-6.7	Active site	PHE A:166 ASP A:90 ARG A:161 THR A:169 ASP A:88
2	CID 442393	Selinene	-9.3	Active site	No overlap on the site, even though they competes only for PHE A:166
3	CID 99535	Podolide	-8.8	Active site	Strong overlap on the site, even though they competes for PHE A:166 and ARG A:161
4	CID 91752839	Zierone	-7.8	Active site	Side overlap on the site, even though they competes only for PHE A:166