Available Online through
www.ijpbs.com

IJPBS |Volume 2| Issue 1 |JAN-MARCH |2012|37-45

In-silico generation of three-dimensional structure and analyzing binding sites of Cynops pyrrhogaster thrombin protein which has not yet been resolved by X-ray Crystallography or NMR

Naincy Chandan[1*] *and Mukeshchand Thakur*[2]

[1*,2] Padmashree Dr.D.Y.Patil University, Department of Biotechnology and Bioinformatics, Navi Mumbai-400 614, Maharashtra, India

*Corresponding Author Email: naincychandan@gmail.com

BIOLOGICAL SCIENCES | Research Article

RECEIVED ON 06-01-2011 | ACCEPTED ON 17-01-2012

ABSTRACT

Thrombin protein three-dimensional structure of Cynops pyrrhogaster has not been resolved and confirmed by X-ray crystallography and NMR experiments. In current work, in silico homology modeling of the thrombin protein sequence of the organism is performed to generate three-dimensional structure using Modeler9v7 software. The energy was minimized using Swiss-pdbViewer 4.0.1 software. Structural validation was performed using PROCHECK from Swiss Model Workspace. The validated best end-model, thus generated had its RMSD value from its template as 0.18 Å and 100% residues within the favorable region of Ramachandran plot. It was hence predicted to be the most suitable candidate similar to the actual three dimensional structure of Cynops pyrrhogaster thrombin molecule in vivo. Its active site was found out using Q-site finder tool.

KEYWORDS: *Homology Modeling, Modeler9v7, PROCHECK, Q-site finder, SwissPDBViewer 4.0.1, Thrombin.*

INTRODUCTION:

Thrombin is an active serine-protease enzyme derived by Vitamin-K dependent proteolysis of its precursor zymogen- Prothrombin. The mechanism of coagulation needs to be clearly understood for determining the preoperative bleeding risk to patients undergoing surgery and managing haemostatic therapy preoperatively [1]. The coagulatory pathway of blood usually occurs by two pathways- the contact activation pathway (intrinsic pathway), and the tissue factor pathway (extrinsic pathway), which lead to fibrin formation. Previously it was inferred that the coagulation cascade consisted of two pathways of equal importance which were amalgamated together to a single common pathway. However, it is now known that the primary pathway for the initiation of blood coagulation is the tissue factor pathway [2]. Prothrombin is a zymogen, i.e an inactive precursor serine protease which is proteolytically cleaved to give thrombin. The reactions are the series of pathways also known as cascades which ultimately leads to an insoluble cross-linked fibrin. The active coagulation factors are denoted by Roman numerals along with small letter 'a' to signify that they are active. In the blood coagulation pathway, thrombin acts to convert factor XI to XIa, VIII to VIIIa, V to Va, and fibrinogen to fibrin. The contact activation pathway commences with formation of the primary complex of collagen by high-molecular weight kininogen (HMWK), prekallikrein, and factor XII (FXII). Prekallikrein is converted to kallikrein and FXII becomes activated factor XII (FXIIa). FXIIain turn converts factor XI (FXI) into activated factor XI (FXIa). Activated factor XI (XIa) activates factor IX (FIX) and with its co-factor activated factor FVIII (FVIIIa) form the tenase complex, which activates factor X (FX) to activated factor X (Fxa). Thrombin has a large array of functions. Its primary role is the conversion of fibrinogen to fibrin, the building block of a hemostatic plug. In addition, it activates Factors VIII and V and their inhibitor protein C (in the presence of thrombomodulin), and it activates Factor XIII, which forms covalent bonds that crosslink the fibrin polymers that form from activated monomers[3].Thrombin receptor activation mechanisms are done by two forms-

Direct activation and transactivation. In direct activation, soluble thrombin directly cleaves the extracellular domain of the receptor to unmask receptor activation whereas in transactivation, soluble thrombin, in low concentrations, interacts with the first receptor and the thrombin-receptor complex itself serves as an enzyme to cleave the second receptor [4,5]. No three dimensional structure of *Cynops pyrrhogaster* has been resolved by X-ray or NMR spectroscopy. Understanding the structure is important since for the development of anti-thrombin molecules. Thrombin is an endolytic serine protease that selectively cleaves the Arg-Gly bond of fibrinogen to form fibrin and release fibrinopeptides A and B. To understand the blood clotting in lower animals, we have selected the animal model of *Cynops pyrrhogaster for* studying its thrombin protein and its active sites.

MATERIALS AND METHODS

C. pyrrhogaster (GI:213388) sequence was taken as target sequence from NCBI proteins and homology modeling performed using Modeler9v7 Windows Platform[6] taking *H. sapiens* thrombin protein (PDB ID:1GJ5) as a template. Ten models were generated and their energy minimized using SPDBV 4.0.1. These models were structurally assessed using PROCHECK from Swiss Model Workspace[7] and the best model were selected on the basis of their Ramachandran plot. After validation, model 04 was found to be best and its RMSD value was calculated using SPDBV 4.0.1. Further its active site analysis was performed using Q-site finder [8].

RESULTS

For thrombin protein sequences of *C. pyrrhogaster,* the BLASTp[9] resultsshowed that Homo sapiens thrombin protein PDB ID: 1GJ5 is 71% similar to the target having e-value: 6.41e-132, bit-score: 372 for an aligned length of 234[10].Ten models were generated using Modeler9v7 (Windows Platform) out of which model 04 gave a better results as compared to other 9 models as shown in **Table I**. The RMSD value between model 04 and 1GJ5 was calculated using SPDBV 4.0.1 and was found to be 0.18 Å. The Ramachandran plot for the model 04 showed residues falling in: Core regions, favorable allowed regions, generously allowed regions and disallowed regions as 88.1%, 11.4%, 0.5% and 0% respectively **Figure I**. Overall plot gave 100% of residues within favorable region hence model 04 can be used for further analysis. The final model is shown in **Figure III** represented using RasMol[11].The active sites were identified using Q-site finder and the best five sites are shown in **Figure II**. Their binding region details are shown in **Table II**.

Table I showing the geometric and PROCHECK statistical data for ten models generated using Modeler9v7.

Model no.	Core region (%)	Allowed region (%)	Generously allowed region (%)	Disallowed region	Bad contacts	G-factor	M/c Bond lengths	M/c bond angles	Planar region
1	85.1	13.9	1	0	2	-0.05	99.5	96.6	90.5
2	85.6	12.9	1	0.5	1	-0.11	99.9	94.9	89.5
3	86.1	12.9	1	0	0	0.01	100	96.5	87.4
4	88.1	11.4	0.5	0	1	-0.01	100	96	88.4
5	86.6	12.4	1	0	0	0.55	99.8	95.8	84.2
6	87.6	11.4	1	0	2	-0.09	100	95.6	86.3
7	85.6	13.4	1	0	2	-0.12	100	95	88.4
8	85.6	12.9	1.5	0	1	0.5	99.6	95.6	89.5
9	84.7	14.4	0.5	0.5	2	-0.1	99.8	96.1	85.3
10	87.1	11.9	1	0	1	-0.09	100	94.9	88.4

Table II (a) & (b) showing the details of best five probable binding sites for model 04 generated using Q-site finder webserver

Binding Site	Site Volume	Binding Box Around Selected Sites
Site 1	563 Cubic Angstroms	Min Coords: (-4, -28, 10) Max Coords: (11, 0, 29)
Site 2	432 Cubic Angstroms	Min Coords: (-4, -28, 10) Max Coords: (11, 0, 29)
Site 3	395 Cubic Angstroms	Min Coords: (-4, -28, 10) Max Coords: (11, 0, 29)
Site 4	263 Cubic Angstroms	Min Coords: (-10, -25, 20) Max Coords: (11, -7, 34)
Site 5	135 Cubic Angstroms	Min Coords: (-7, -6, 21) Max Coords: (7, 9, 36)

a) **Showing the volume in cubic angstroms and the co-ordinates of the respective binding sites.**

b) **Showing the amino acids present at the respective binding sites generated using Q-site finder. The atom number, atom type, residue name, chain identifier and residue number are shown.**

Site 1	Site 2	Site 3	Site 4	Site 5
7 CD GLN 1	27 C LEU 3	1252 CA VAL 115	1509 CD1 LEU 141	1209 CG HIS 111
9 NE2 GLN 1	28 O LEU 3	1253 C VAL 115	1725 C ALA 164	1210 ND1 HIS 111
11 1HE2 GLN 1	34 N ILE 4	1254 O VAL 115	1726 O ALA 164	1211 CD2 HIS 111
12 2HE2 GLN 1	35 CA ILE 4	1255 CB VAL 115	1729 N GLY 165	1212 CE1 HIS 111
17 C GLU 2	36 C ILE 4	1256 CG1 VAL 115	1730 CA GLY 165	1213 NE2 HIS 111
18 O GLU 2	37 O ILE 4	1257 CG2 VAL 115	1731 C GLY 165	1216 N LYS 112
25 N LEU 3	38 CB ILE 4	1258 H VAL 115	1732 O GLY 165	1217 CA LYS 112
26 CA LEU 3	39 CG1 ILE 4	1259 N SER 116	1734 N TYR 166	1218 C LYS 112
27 C LEU 3	40 CG2 ILE 4	1260 CA SER 116	1735 CA TYR 166	1219 O LYS 112
29 CB LEU 3	41 CD ILE 4	1261 C SER 116	1736 C TYR 166	1220 CB LYS 112
30 CG LEU 3	43 N CYS 5	1262 O SER 116	1737 O TYR 166	1221 CG LYS 112
31 CD1 LEU 3	48 SG CYS 5	1265 H SER 116	1738 CB TYR 166	1222 CD LYS 112
32 CD2 LEU 3	182 C HIS 20	1267 N GLY 117	1739 CG TYR 166	1225 H LYS 112
34 N ILE 4	183 O HIS 20	1268 CA GLY 117	1741 CD2 TYR 166	1229 N GLY 113
35 CA ILE 4	184 CB HIS 20	1269 C GLY 117	1746 H TYR 166	1230 CA GLY 113
36 C ILE 4	185 CG HIS 20	1270 O GLY 117	1748 HD2 TYR 166	1231 C GLY 113
38 CB ILE 4	186 ND1 HIS 20	1271 H GLY 117	1752 N LYS 167	1232 O GLY 113
39 CG1 ILE 4	187 CD2 HIS 20	1272 N TRP 118	1753 CA LYS 167	1233 H GLY 113
40 CG2 ILE 4	188 CE1 HIS 20	1273 CA TRP 118	1754 C LYS 167	1234 N ARG 114
41 CD ILE 4	189 NE2 HIS 20	1274 C TRP 118	1755 O LYS 167	1235 CA ARG 114
42 H ILE 4	192 N CYS 21	1293 N GLY 119	1765 N PRO 168	1238 CB ARG 114
43 N CYS 5	193 CA CYS 21	1294 CA GLY 119	1766 CA PRO 168	1239 CG ARG 114
44 CA CYS 5	194 C CYS 21	1295 C GLY 119	1767 C PRO 168	1240 CD ARG 114
45 C CYS 5	195 O CYS 21	1296 O GLY 119	1768 O PRO 168	1241 NE ARG 114
46 O CYS 5	196 CB CYS 21	1297 H GLY 119	1769 CB PRO 168	1242 CZ ARG 114
47 CB CYS 5	197 SG CYS 21	1298 N ASN 120	1770 CG PRO 168	1243 NH1 ARG
48 SG CYS 5	198 H CYS 21	1299 CA ASN 120	1771 CD PRO 168	114

49	**H**	**CYS**	**5**	199	N	ILE	22	1300	C	ASN	120
50	**N**	**GLY**	**6**	200	CA	ILE	22	1301	O	ASN	120
51	**CA**	**GLY**	**6**	201	C	ILE	22	1309	N	LEU	121
52	**C**	**GLY**	**6**	202	O	ILE	22	1310	CA	LEU	121
53	**O**	**GLY**	**6**	208	N	PHE	23	1311	C	LEU	121
54	**H**	**GLY**	**6**	210	C	PHE	23	1312	O	LEU	121
55	**N**	**ALA**	**7**	225	N	TYR	24	1313	CB	LEU	121
56	**CA**	**ALA**	**7**	226	CA	TYR	24	1314	CG	LEU	121
57	**C**	**ALA**	**7**	228	O	TYR	24	1315	CD1	LEU	121
58	**O**	**ALA**	**7**	229	CB	TYR	24	1316	CD2	LEU	121
59	**CB**	**ALA**	**7**	230	CG	TYR	24	1318	N	HIS	122
60	**H**	**ALA**	**7**	237	H	TYR	24	1319	CA	HIS	122
61	**N**	**SER**	**8**	239	HD2	TYR	24	1320	C	HIS	122
62	**CA**	**SER**	**8**	261	CB	TRP	27	1321	O	HIS	122
63	**C**	**SER**	**8**	264	CD2	TRP	27	1322	CB	HIS	122
64	**O**	**SER**	**8**	267	CE3	TRP	27	1323	CG	HIS	122
65	**CB**	**SER**	**8**	268	CZ2	TRP	27	1324	ND1	HIS	122
66	**OG**	**SER**	**8**	269	CZ3	TRP	27	1325	CD2	HIS	122
67	**H**	**SER**	**8**	270	CH2	TRP	27	1326	CE1	HIS	122
69	**N**	**ILE**	**9**	274	HE3	TRP	27	1327	NE2	HIS	122
70	**CA**	**ILE**	**9**	275	HZ2	TRP	27	1328	H	HIS	122
73	**CB**	**ILE**	**9**	276	HZ3	TRP	27	1329	HD1	HIS	122
74	**CG1**	**ILE**	**9**	277	HH2	TRP	27	1330	N	GLU	123
75	**CG2**	**ILE**	**9**	288	CA	LYS	29	1331	CA	GLU	123
76	**CD**	**ILE**	**9**	289	C	LYS	29	1334	CB	GLU	123
77	**H**	**ILE**	**9**	290	O	LYS	29	1335	CG	GLU	123
146	**CB**	**VAL**	**15**	291	CB	LYS	29	1336	CD	GLU	123
147	**CG1**	**VAL**	**15**	292	CG	LYS	29	1337	OE1	GLU	
148	**CG2**	**VAL**	**15**	293	CD	LYS	29	123			
160	**CA**	**THR**	**17**	294	CE	LYS	29	1338	OE2	GLU	
163	**CB**	**THR**	**17**	295	NZ	LYS	29	123			
164	**OG1**	**THR**	**17**	296	H	LYS	29	1339	H	GLU	123
165	**CG2**	**THR**	**17**	297	HZ1	LYS	29	1460	N	GLN	137
167	**HG1**	**THR**	**17**	298	HZ2	LYS	29	1461	CA	GLN	137
173	**H**	**ALA**	**18**	299	HZ3	LYS	29	1462	C	GLN	137
193	**CA**	**CYS**	**21**	318	CD2	TYR	31	1463	O	GLN	137
194	**C**	**CYS**	**21**	320	CE2	TYR	31	1464	CB	GLN	137
195	**O**	**CYS**	**21**	321	CZ	TYR	31	1465	CG	GLN	137
196	**CB**	**CYS**	**21**	322	OH	TYR	31	1466	CD	GLN	137
197	**SG**	**CYS**	**21**	325	HD2	TYR	31	1467	OE1	GLN	
199	**N**	**ILE**	**22**	327	HE2	TYR	31	137			
200	**CA**	**ILE**	**22**	328	HH	TYR	31	1468	NE2	GLN	
203	**CB**	**ILE**	**22**	1835	CB	ASP	175	137			
204	**CG1**	**ILE**	**22**	1836	CG	ASP	175	1469	H	GLN	137
205	**CG2**	**ILE**	**22**	1837	OD1	ASP		1470	1HE2	GLN	
206	**CD**	**ILE**	**22**	175				137			
207	**H**	**ILE**	**22**	1838	OD2	ASP		1471	2HE2	GLN	
312	**CA**	**TYR**	**31**	175				137			
313	**C**	**TYR**	**31**	1840	N	ALA	176	1472	N	GLN	138
314	**O**	**TYR**	**31**	1841	CA	ALA	176	1473	CA	GLN	138
315	**CB**	**TYR**	**31**	1842	C	ALA	176	1474	C	GLN	138
316	**CG**	**TYR**	**31**	1843	O	ALA	176	1475	O	GLN	138
318	**CD2**	**TYR**	**31**	1844	CB	ALA	176	1481	H	GLN	138
320	**CE2**	**TYR**	**31**	1845	H	ALA	176	1484	N	VAL	139

1772	N	ASP	169	1244	NH2	ARG					
1781	N	GLU	170	114							
1783	C	GLU	170	1245	H	ARG	114				
1790	H	GLU	170	1246	HE	ARG	114				
1791	N	PRO	171	1247	1HH1	ARG					
1792	CA	PRO	171	114							
1793	C	PRO	171	1248	2HH1	ARG					
1794	O	PRO	171	114							
1795	CB	PRO	171	1249	1HH2	ARG					
1796	CG	PRO	171	114							
1797	CD	PRO	171	1250	2HH2	ARG					
1798	N	ASN	172	114							
1799	CA	ASN	172	1493	CA	ASN	140				
1800	C	ASN	172	1496	CB	ASN	140				
1801	O	ASN	172	1497	CG	ASN	140				
1806	H	ASN	172	1498	OD1	ASN					
1809	N	ARG	173	140							
1810	CA	ARG	173	1499	ND2	ASN					
1811	C	ARG	173	140							
1812	O	ARG	173	1501	1HD2	ASN					
1813	CB	ARG	173	140							
1814	CG	ARG	173	1502	2HD2	ASN					
1815	CD	ARG	173	140							
1816	NE	ARG	173	1511	H	LEU	141				
1817	CZ	ARG	173	1920	CA	VAL	186				
1819	NH2	ARG		1921	C	VAL	186				
173				1922	O	VAL	186				
1821	HE	ARG	173	1923	CB	VAL	186				
1824	1HH2	ARG		1925	CG2	VAL	186				
173				1927	N	MET	187				
1825	2HH2	ARG		1928	CA	MET	187				
173				1929	C	MET	187				
1826	N	GLY	174	1930	O	MET	187				
1827	CA	GLY	174	1931	CB	MET	187				
1828	C	GLY	174	1932	CG	MET	187				
1829	O	GLY	174	1935	H	MET	187				
1830	H	GLY	174	1936	N	LYS	188				
1831	N	ASP	175	1937	CA	LYS	188				
1832	CA	ASP	175	1938	C	LYS	188				
1835	CB	ASP	175	1939	O	LYS	188				
1836	CG	ASP	175	1940	CB	LYS	188				
1837	OD1	ASP		1941	CG	LYS	188				
175				1945	H	LYS	188				
1838	OD2	ASP		2017	CD2	TRP	195				
175				2020	CE3	TRP	195				
2147	N	ASP	208	2021	CZ2	TRP	195				
2148	CA	ASP	208	2022	CZ3	TRP	195				
2149	C	ASP	208	2023	CH2	TRP	195				
2150	O	ASP	208	2027	HE3	TRP	195				
2151	CB	ASP	208	2029	HZ3	TRP	195				
2152	CG	ASP	208	2030	HH2	TRP					
2153	OD1	ASP		195							
208											
2154	OD2	ASP									

323	H	TYR	31	1846	N	CYS	177	1485	CA VAL 139 208
325	HD2	TYR	31	1847	CA	CYS	177	1488	CB VAL 139 2155 H ASP 208
327	HE2	TYR	31	1848	C	CYS	177	1489	CG1 VAL 139 2156 N ARG 209
329	N	THR	32	1849	O	CYS	177	1490	CG2 VAL 139 2157 CA ARG 209
330	CA	THR	32	1850	CB	CYS	177	1491	H VAL 139 2158 C ARG 209
331	C	THR	32	1851	SG	CYS	177	1811	C ARG 173 2159 O ARG 209
332	O	THR	32	1852	H	CYS	177	1817	CZ ARG 173 2167 H ARG 209
336	H	THR	32	1853	N	GLU	178	1818	NH1 ARG 2173 N ASP 210
338	N	THR	33	1854	CA	GLU	178	173	2174 CA ASP 210
339	CA	THR	33	1855	C	GLU	178	1819	NH2 ARG 2175 C ASP 210
340	C	THR	33	1857	CB	GLU	178	173	2176 O ASP 210
341	O	THR	33	1858	CG	GLU	178	1822	1HH1 ARG 2177 CB ASP 210
345	H	THR	33	1859	CD	GLU	178	173	2178 CG ASP 210
365	H	ASP	35	1861	OE2	GLU		1823	2HH1 ARG 2179 OD1 ASP
366	N	ILE	36	178				173	210
367	CA	ILE	36	1862	H	GLU	178	1824	1HH2 ARG 2180 OD2 ASP
370	CB	ILE	36	1863	N	GLY	179	173	210
371	CG1	ILE	36	1864	CA	GLY	179	1825	2HH2 ARG 2181 H ASP 210
372	CG2	ILE	36	1865	C	GLY	179	173	2182 N GLY 211
373	CD	ILE	36	1867	H	GLY	179	1826	N GLY 174 2183 CA GLY 211
374	H	ILE	36	1868	N	ASP	180	1827	CA GLY 174 2184 C GLY 211
375	N	LEU	37	1869	CA	ASP	180	1828	C GLY 174 2186 H GLY 211
377	C	LEU	37	1870	C	ASP	180	1829	O GLY 174 2187 N LYS 212
378	O	LEU	37	1872	CB	ASP	180	1830	H GLY 174 2188 CA LYS 212
383	H	LEU	37	1876	H	ASP	180	1831	N ASP 175 2189 C LYS 212
384	N	VAL	38	1877	N	SER	181	1832	CA ASP 175 2190 O LYS 212
385	CA	VAL	38	1878	CA	SER	181	1833	C ASP 175 2191 CB LYS 212
386	C	VAL	38	1881	CB	SER	181	1834	O ASP 175 2196 H LYS 212
388	CB	VAL	38	1882	OG	SER	181	1835	CB ASP 175 2200 N TYR 213
389	CG1	VAL	38	1883	H	SER	181	1836	CG ASP 175 2201 CA TYR 213
390	CG2	VAL	38	1884	HG	SER	181	1838	OD2 ASP 2202 C TYR 213
392	N	ARG	39	2085	C	VAL	201	175	2203 O TYR 213
393	CA	ARG	39	2086	O	VAL	201	1839	H ASP 175 2204 CB TYR 213
394	C	ARG	39	2087	CB	VAL	201	1840	N ALA 176 2205 CG TYR 213
395	O	ARG	39	2088	CG1	VAL	201	1841	CA ALA 176 2206 CD1 TYR 213
396	CB	ARG	39	2091	N	SER	202	1842	C ALA 176 2208 CE1 TYR 213
397	CG	ARG	39	2092	CA	SER	202	1843	O ALA 176 2210 CZ TYR 213
398	CD	ARG	39	2093	C	SER	202	1844	CB ALA 176 2211 OH TYR 213
399	NE	ARG	39	2094	O	SER	202	1845	H ALA 176 2213 HD1 TYR 213
401	NH1	ARG	39	2097	H	SER	202	1846	N CYS 177 2215 HE1 TYR 213
403	H	ARG	39	2099	N	TRP	203	1847	CA CYS 177 2218 N GLY 214
404	HE	ARG	39	2100	CA	TRP	203	1850	CB CYS 177 2219 CA GLY 214
406	2HH1	ARG		2101	C	TRP	203	1851	SG CYS 177 2222 H GLY 214
39				2102	O	TRP	203	1852	H CYS 177
409	N	ILE	40	2103	CB	TRP	203	1872	CB ASP 180
410	CA	ILE	40	2113	H	TRP	203	1873	CG ASP 180
411	C	ILE	40	2120	N	GLY	204	1874	OD1 ASP
413	CB	ILE	40	2121	CA	GLY	204	180	
414	CG1	ILE	40	2122	C	GLY	204	1875	OD2 ASP
415	CG2	ILE	40	2123	O	GLY	204	180	
416	CD	ILE	40	2124	H	GLY	204	2140	N CYS 207
418	N	GLY	41	2125	N	GLU	205	2141	CA CYS 207
422	H	GLY	41	2126	CA	GLU	205	2142	C CYS 207
427	CB	LYS	42	2127	C	GLU	205	2143	O CYS 207

428 CG LYS 42	2134 H GLU 205	2144 CB CYS 207
429 CD LYS 42	2135 N GLY 206	2145 SG CYS 207
430 CE LYS 42	2136 CA GLY 206	2146 H CYS 207
432 H LYS 42	2137 C GLY 206	2147 N ASP 208
637 CG LEU 58	2138 O GLY 206	2148 CA ASP 208
638 CD1 LEU 58	2139 H GLY 206	2149 C ASP 208
639 CD2 LEU 58	2140 N CYS 207	2150 O ASP 208
672 CB ILE 61	2141 CA CYS 207	2151 CB ASP 208
673 CG1 ILE 61	2142 C CYS 207	2152 CG ASP 208
674 CG2 ILE 61	2144 CB CYS 207	2153 OD1 ASP
675 CD ILE 61	2145 SG CYS 207	208
887 CB ILE 80	2147 N ASP 208	2154 OD2 ASP
888 CG1 ILE 80	2155 H ASP 208	208
889 CG2 ILE 80	2203 O TYR 213	2155 H ASP 208
890 CD ILE 80	2218 N GLY 214	2156 N ARG 209
1279 CD2 TRP 118	2219 CA GLY 214	2157 CA ARG 209
1281 CE2 TRP 118	2220 C GLY 214	2158 C ARG 209
1282 CE3 TRP 118	2223 N PHE 215	2160 CB ARG 209
1283 CZ2 TRP 118	2234 H PHE 215	2161 CG ARG 209
1284 CZ3 TRP 118	2251 OH TYR 216	2162 CD ARG 209
1285 CH2 TRP 118		2163 NE ARG 209
1289 HE3 TRP 118		2164 CZ ARG 209
1290 HZ2 TRP 118		2165 NH1 ARG
1291 HZ3 TRP 118		209
1292 HH2 TRP		2166 NH2 ARG
118		209
		2167 H ARG 209
		2169 1HH1 ARG
		209
		2170 2HH1 ARG
		209
		2171 1HH2 ARG
		209
		2173 N ASP 210
		2174 CA ASP 210
		2177 CB ASP 210
		2178 CG ASP 210
		2179 OD1 ASP
		210
		2180 OD2 ASP
		210
		2181 H ASP 210

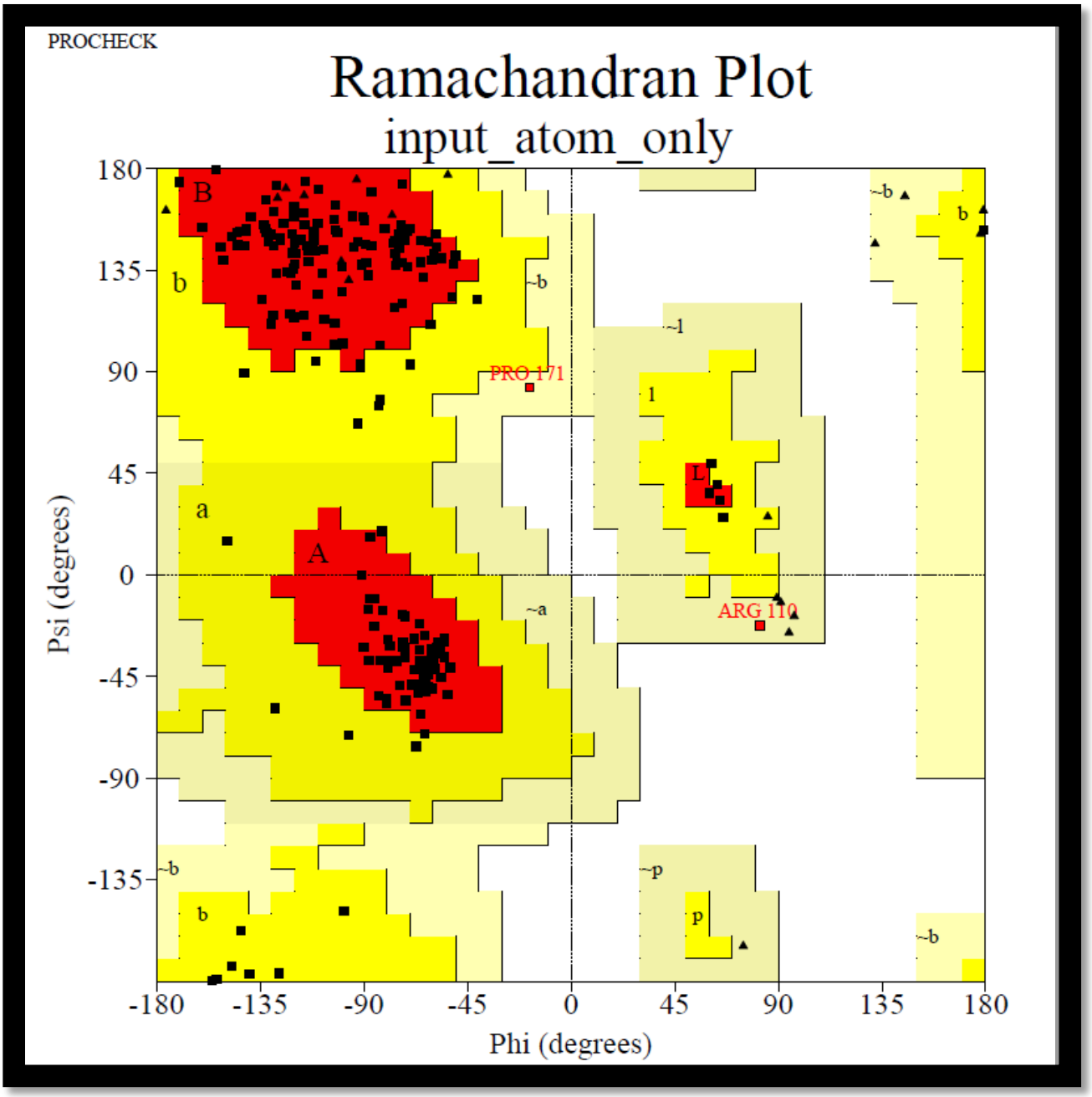


Figure I showing Ramachandran plot of phi/psi values of the modeled *C. pyrrhogaster* thrombin protein homology model 04. The colour codes are: Red color- most favorable regions, yellow color- allowed region, and pale yellow- generously allowed region and white color- disallowed regions.

Figure II showing model 04- Color codes are Yellow colour – Beta-sheets, Pink colour helices, red colour turns and white colour loops. The model 04 contains in all 7 helices, 22 strands and 26 turns.

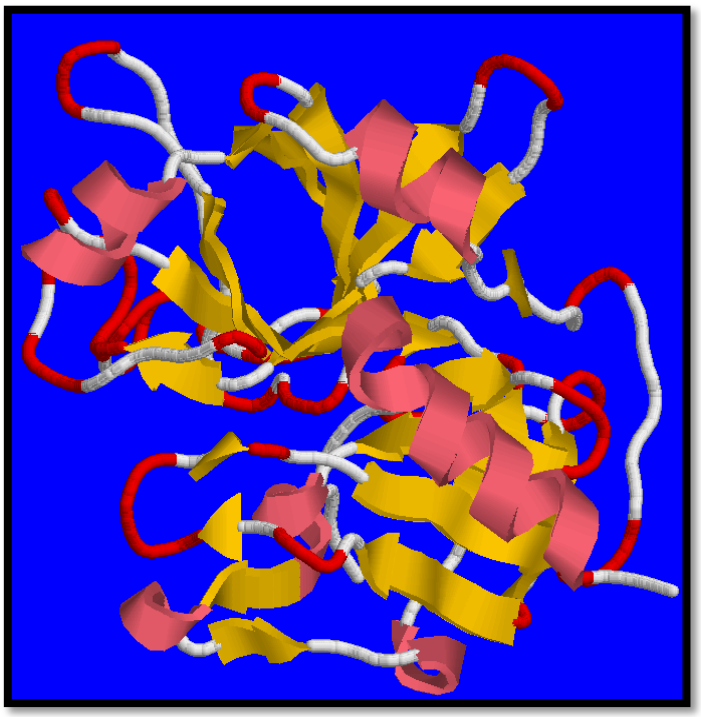
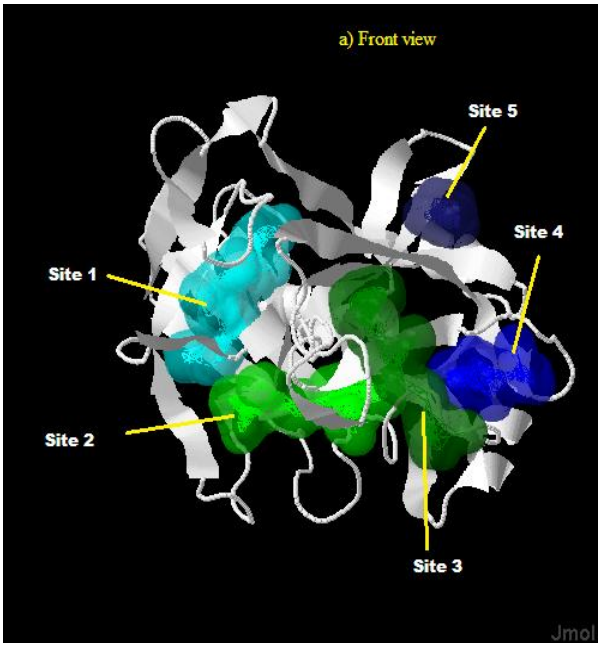
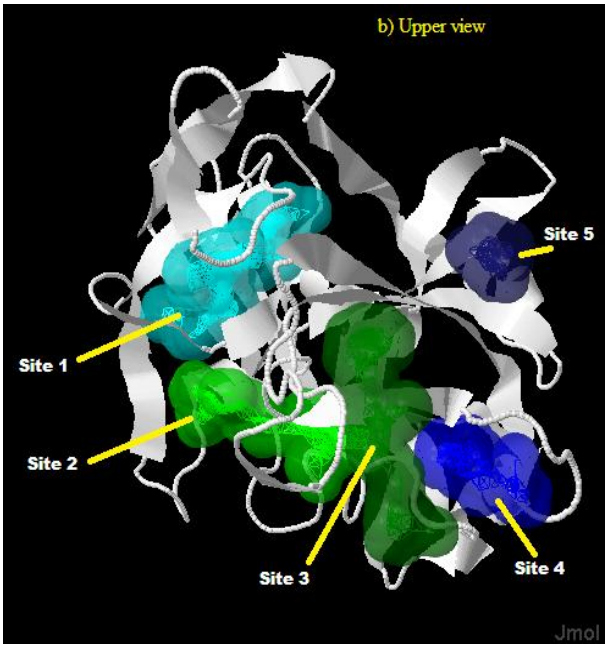


Figure III showing the toggled view of binding region in model 04: (a) Showing the front view of the structure and (b) showing the upper view of the structure with respective binding sites. Colour codes representing the sites are: Cyan color for site 1, Light green colour for site 2, dark green colour for site 3, blue colour for site 4 and dark blue colour for site 5.

(a) **(b)**





DISCUSSION

Bleeding complications may arise out of anticoagulant therapy. Hence it is essential to develop specific antidotes against the anticoagulants being used. Heparin and Warfarin are being used as a anti-coagulant have specific effects. Due to high cost required in computer hardware to cover-up the graphics required for molecular modeling, programs like Modeler have been developed. Its error rate is less but it is a bit inconvenient since human made errors may arise at any step. The homology model we have developed may provide an insight for further studies. The model can further be used for docking studies which can help in drug development procedures to develop new anti-thrombin molecules.

REFERENCES

1. Johri P, Nagare S, Swamy KV and Kumar CS(2011). An insight into blood clotting disorders in humans. Journal of Computational Biology and Bioinformatics Research Vol. 3(1), pp. 8-14.
2. Tanaka KA, Key NS, Levy JH (2009). Blood coagulation: hemostasis and thrombin regulation. Anesth. Anal., 108: 1433-1446.
3. Vine AK (2009). Recent advances in haemostasis and thrombosis. Retina, 29: 1-7.
4. Esmon CT (1989). The roles of protein C and thrombomodulin in the regulation of blood coagulation. J. Biol. Chem., 264: 4743-4746.
5. Esmon CT (2008). Crosstalk between inflammation and thrombosis.Maturitas, 61: 122-131.
6. Fiser A, Sali A (2003). Modeller: generation and refinement of homology-based protein structure models. Methods in Enzymology, 374: 461-491.
7. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 18: 2714-2723.
8. Laurie AT, Jackson RM (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics,* **21:** 1908-1916.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol., 215: 403-410
10. Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673 -4680.
11. Roger Sayle and E. James Milner-White. "RasMol: Biomolecular graphics for all", *Trends in Biochemical Sciences (TIBS)*, September 1995, Vol. 20, No. 9, p. 374.

*** Corresponding Author**
1.Naincy RamniklalChandan
Email : naincychandan@gmail.com
2.Mukeshchand Thakur
Email: mukeshchandthakur@yahoo.com
PadmashreeDr.D.Y.Patil University, Department of Biotechnology and Bioinformatics, Navi Mumbai-400 614, Maharashtra. India