Supporting Information

AADS- An automated active site identification, docking and scoring protocol for protein targets based on physico-chemical descriptors

Tanya Singh, D. Biswas and B. Jayaram*

Table of Contents

Supplementary Table 1	Some docking programs reported in the literature
Supplementary Table 2	Comparative Evaluation of *Bappl* scoring function
Supplementary Table 3	PDB IDs of the 620 protein data set used to identify active sites
Supplementary Table 4.	A comparison of different active site identifier programs on 48 protein-ligand complexes
Supplementary Figure 1	A schematic representation of active site identification protocol
Supplementary Figure 2	A schematic representation of parallel implementation of docking and scoring protocol

Supplementary Table 1. Some docking programs reported in the literature

Program	Search Algorithm	Program	Search Algorithm
Affinity	MC	ICM[123,124]	MC
ADAM[77-79]	RBD	LibDock[125]	MA
Autodock 4.0[80,81]	GA	LIDAEUS[126]	MA
Autodock Vina[82]	-	LIGIN[127]	RBD
CDOCKER[83]	MD-SA	MacDOCK[128]	IC
CHARMM[84]	GA and MC	MCDOCK[129]	MC
CLIX[85]	RBD	MolDock[130]	DE
DARWIN[86]	GA	MVP	IC
DIVALI[87]	GA	PatchDOCK[131]	Shape complemantarity
DOCK 6[88-91]	IC	PAS-Dock[132]	TS
Dock*Blaster*[92]	-	PhDOCK[133]	MA
DockIt	-	Ph4DOCK[134]	MA
DockVision 1.0.3[93]	MC	PIPER[135]	Fast fourier transform
DoMCoSAR[94]	SA	Plants[136]	Ant colony optimisation
DragHOME[95]	RBD	Prodock[137]	MC
EADock[96]	EA	PRO_LEADS[138]	TS
eHiTs[97]	RBD	ProPose[139]	IC
EUDOC[98]	RBD	PSI_DOCK[140]	GA and TS
FDS[99]	MC	Q-fit[141]	MA
DAIM-SEED-FFLD[100-102]	RB	Quantum	-
FITTED[103]	GA	QXP/Flo+[142]	MC
FlexX 2.2[104]	IC	RiboDock[143]	MC
FlexE[105,106]	-	ROSETTALIGAND[144]	MC
FlipDock[107]	GA	SANDOCK[145]	MC
FLOG[108]	IC	SDOCKER[146]	Random walk
FRED[109]	RBD	SLIDE[147,148]	IC
FTDock[110]	RBD	SKELGEN[149]	IC
GAMBLER[111]	GA	SO DOCK[150]	Swarm Optimisation
GasDock[112]	GA	SG-DOCK/SP-DOCK[151]	IC
GEMDOCK[113]	EA	Surflex 2.1[152,153]	IC with MA
GlamDock[114]	MC/SA	Yucca[154]	MC
Glide 4.0[115,116]	MC	ParDOCK[155]	MC
GOLD 3.1[117,118]	GA		
Hammerhead[119]	IC		
HADDOCK[120]	SA		
HierDOCK/HierVLS[121,122]	IC		

Abbreviations: N/A, not available. [a]DE: differential evolution, EA: evolutionary algorithm, GA: genetic algorithm, IC: incremental construction, MA: matching algorithm, MD: molecular dynamics, RBD: rigid-body docking, SA: simulated annealing, TS: tabu search.

Supplementary Table 2. Comparative Evaluation of *Bappl* scoring function

S. No.	Scoring Function	Method	Correlation Coefficient (r)
1	Bappl[167]	Force field / Empirical	r = 0.92
2	DOCK[176]	DOCK[167]	-
3	EUDOC[177]	Force field	-
4	CHARMm[178]	Force field	-
5	AutoDock[179]	Force field	-
6	DrugScore[180]	Knowledge	-
7	SMoG[181]	Knowledge	r=0.79
8	BLEEP[182]	Knowledge	r = 0.74
9	PMF[183]	Knowledge	r = 0.78
10	DFIRE[184]	Knowledge	r = 0.63
11	SCORE[185]	Empirical	r = 0.81
12	GOLD[186]	Empirical	-
13	LUDI[187]	Empirical	r = 0.83
14	FlexX[188]	Empirical	-
15	ChemScore[189]	Empirical	r = 0.84
16	VALIDATE[190]	Empirical	r = 0.90
17	Ligscore[191]	Empirical	r = 0.87
18	X-CSCORE[192]	Empirical (consensus)	r = 0.77
19	GLIDE[193]	Force field / Empirical	-

Supplementary Table 3. PDB ID's of the 620 proteins used to identify active sites

1	1A30	40	1BA8	79	1CX9	118	1ERE	157	1HSL
2	1A4K	41	1BAP	80	1D3D	119	1ETR	158	1HTE
3	1A4Q	42	1BB0	81	1D3H	120	1ETS	159	1HTF
4	1A4W	43	1BDR	82	1D3P	121	1ETT	160	1HTG
5	1A9M	44	1BID	83	1D3Q	122	1EZQ	161	1HVH
6	1AAQ	45	1BIL	84	1D3T	123	1F0R	162	1HVI
7	1ABE	46	1BIM	85	1D4L	124	1F0S	163	1HVJ
8	1ABF	47	1BJI	86	1D4P	125	1F0T	164	1HVK
9	1AC4	48	1BMA	87	1DBB	126	1F0U	165	1HVL
10	1ACJ	49	1BMM	88	1DBJ	127	1FAX	166	1HVR
11	1ACM	50	1BMN	89	1DBK	128	1FBC	167	1HWR
12	1ADB	51	1BRA	90	1DBM	129	1FKG	168	1HXB
13	1ADF	52	1BTO	91	1DD7	130	1FLR	169	1HXW
14	1AE8	53	1BV7	92	1DG5	131	1FRP	170	1IDA
15	1AEB	54	1BV9	93	1DHF	132	1G2K	171	1IMB
16	1AED	55	1BWA	94	1DID	133	1GHB	172	1IVB
17	1AFK	56	1BWB	95	1DIE	134	1GNO	173	1IVD
18	1AFL	57	1BXO	96	1DMP	135	1GPY	174	1IVF
19	1AI4	58	1BXQ	97	1DOG	136	1HBV	175	1K1L
20	1AI5	59	1BYB	98	1DR1	137	1HDC	176	1K1M
21	1AJP	60	1C29	99	1DRF	138	1HDT	177	1K1N
22	1AJV	61	1C2T	100	1DWB	139	1HEF	178	1K6C
23	1AJX	62	1C5C	101	1DWC	140	1HEW	179	1LCP
24	1ANF	63	1C84	102	1DWD	141	1HGE	180	1LDM
25	1APB	64	1C85	103	1DY9	142	1HGG	181	1LGR
26	1APT	65	1C86	104	1EAP	143	1HGH	182	1LYB
27	1APU	66	1C87	105	1EBG	144	1HGI	183	1MCB
28	1APV	67	1C88	106	1ECV	145	1HGJ	184	1MCF
29	1APW	68	1C8K	107	1EED	146	1HIH	185	1MCH
30	1B38	69	1C9D	108	1EKO	147	1HII	186	1MCJ
31	1B39	70	1CBS	109	1ELA	148	1HIV	187	1MCS
32	1B5G	71	1CF8	110	1ELB	149	1HOS	188	1MDR
33	1B6J	72	1COM	111	1ELC	150	1HPO	189	1MRK
34	1B6K	73	1CPI	112	1ELD	151	1HPS	190	1MTW
35	1B6L	74	1CQP	113	1ELE	152	1HPV	191	1NCO
36	1B6M	75	1CTR	114	1ENT	153	1HPX	192	1NNB
37	1B9S	76	1CTT	115	1ENU	154	1HRN	193	1NSC
38	1B9T	77	1CW2	116	1EPO	155	1HSG	194	1NSD
39	1B9V	78	1CX2	117	1EPP	156	1HSH	195	1OLA

196	1PDZ	235	2ER7	274	7TIM	313	1CHM	352	1LOS
197	1PGP	236	2GBP	275	8GPB	314	1DBT	353	1MMP
198	1PHH	237	2IFB	276	9ABP	315	1DI8	354	1MMQ
199	1PPC	238	2MCP	277	9HVP	316	1DQX	355	1MMR
200	1PPH	239	2MSB	278	1HRI	317	1DVJ	356	1MU6
201	1PPK	240	2PHH	279	1F3E	318	1DY4	357	1MU8
202	1PPL	241	2PK4	280	1CVU	319	1E1V	358	1MUE
203	1PPM	242	2R04	281	1CSC	320	1E1X	359	1NM5
204	1PSO	243	2SIM	282	1C83	321	1E9H	360	1NM6
205	1Q6N	244	2UPJ	283	1AJN	322	1EIX	361	1NT1
206	1QBR	245	2WEA	284	1AEE	323	1EYN	362	1OI9
207	1QBT	246	2WEB	285	1ADC	324	1G30	363	1OIQ
208	1QBU	247	2WEC	286	1LIC	325	1G32	364	1OIR
209	1RBP	248	2XIS	287	1MMQ	326	1H00	365	1OIU
210	1RGK	249	2YPI	288	1POC	327	1H01	366	1P2A
211	1RGL	250	3CLA	289	1SLT	328	1H07	367	1PPW
212	1RNE	251	3ER3	290	1SRJ	329	1H1Q	368	1PY5
213	1RPA	252	3ER5	291	1TKA	330	1H1S	369	1PYE
214	1SNC	253	3ERT	292	1TRK	331	1HZZ	370	1R78
215	1SRE	254	3PTB	293	1WAP	332	1IEI	371	1RE8
216	1STP	255	4DFR	294	1XIE	333	1IYK	372	1REJ
217	1TLC	256	4ER2	295	2ADA	334	1J01	373	1REK
218	1TMT	257	4EST	296	2CHT	335	1JKI	374	1RM0
219	1TNG	258	4HMG	297	2LGS	336	1JLR	375	1RV1
220	1TNH	259	4PHV	298	2MCP	337	1JRB	376	1RW8
221	1TNI	260	4TS1	299	2PLV	338	1JRC	377	1SL3
222	1TNJ	261	5ABP	300	2R07	339	1JWT	378	1SO2
223	1TNK	262	5CNA	301	4AAH	340	1KE5	379	1T1R
224	1TNL	263	5CPP	302	5P2P	341	1KE6	380	1T1S
225	1TPH	264	5ENL	303	6RSA	342	1KE7	381	1TA2
226	1TPP	265	5HVP	304	1IVE	343	1KE8	382	1UPF
227	1ULB	266	5TIM	305	1LNA	344	1KE9	383	1VJY
228	1UVS	267	6ABP	306	1LST	345	1KTS	384	1VYZ
229	2ACS	268	6ENL	307	1TMN	346	1KTT	385	1WCC
230	2AK3	269	6RNT	308	2AK3	347	1LEV	386	1X38
231	2CGR	270	6TIM	309	3CPA	348	1LO7	387	1X97
232	2CMD	271	7ABP	310	3TPI	349	1LO8	388	1X98
233	2CTC	272	7DFR	311	4CTS	350	1LOL	389	1XLX
234	2ER6	273	7GPB	312	8GCH	351	1LOR	390	1XLZ

391	1XM4	430	2BHE	469	2IPW	508	2RFY	547	1ACJ
392	1XM6	431	2BPM	470	2IQ0	509	2UZE	548	1BLH
393	1XMU	432	2BTR	471	2IS7	510	2UZN	549	1CDO
394	1XMY	433	2BTS	472	2ISF	511	2UZO	550	1FBP
395	1XN0	434	2C5Y	473	2IW6	512	2VNP	551	1FBP
396	1XON	435	2C68	474	2IW8	513	2VTZ	552	1GCA
397	1XOQ	436	2C69	475	2IW9	514	2VV9	553	1INC
398	1XOS	437	2C6I	476	2JF4	515	2W05	554	1OKM
399	1XOT	438	2C6K	477	2JG0	516	2W17	555	1QPE
400	1XUC	439	2C6L	478	2NVC	517	3B8Z	556	1SRF
401	1XUD	440	2C6M	479	2NVD	518	3BHT	557	2H4N
402	1XUR	441	2D1N	480	2NXE	519	3BHU	558	2TMN
403	1Y2B	442	2D41	481	2O5D	520	3BHV	559	7CPA
404	1Y2C	443	2DKN	482	2O8H	521	3BJC	560	1CSH
405	1Y2D	444	2DS1	483	2OVV	522	3BR9	561	1PDA
406	1Y2E	445	2DUV	484	2OVY	523	3C1K	562	1PII
407	1Y2H	446	2DXS	485	2OW9	524	3CJ3	563	2DKB
408	1Y2K	447	2EGH	486	2PD5	525	3CJ4	564	2POR
409	1Y8Y	448	2FM0	487	2PD9	526	3CM2	565	1ADS
410	1Y91	449	2FM5	488	2PDB	527	3ECN	566	1ARP
411	1YKR	450	2FQO	489	2PDC	528	3HAD	567	1CCA
412	1YOU	451	2FQT	490	2PDF	529	456C	568	1GOX
413	1YVF	452	2FSV	491	2PDG	530	830C	569	1OYB
414	1YYZ	453	2FVD	492	2PDI	531	1ADD	570	1TDE
415	1YZ0	454	2G9X	493	2PDN	532	1BYH	571	2NPX
416	1Z3N	455	2GC8	494	2PDU	533	1EPM	572	9ICD
417	1Z3T	456	2GIQ	495	2PGJ	534	1FMP	573	1ACL
418	1Z3V	457	2GIR	496	2QYN	535	1FUT	574	1AEC
419	1Z4N	458	2HAI	497	2R3F	536	1HNE	575	1AHA
420	1Z4O	459	2HWH	498	2R3G	537	1ONC	576	1AZM
421	1Z71	460	2HWI	499	2R3H	538	1PEK	577	1BAF
422	1ZGV	461	2I1R	500	2R3J	539	1PIP	578	1BLH
423	1ZKL	462	2IKG	501	2R3K	540	1RPA	579	1CBX
424	1ZRB	463	2IKH	502	2R3L	541	2ALP	580	1CIL
425	1ZTQ	464	2IKI	503	2R3M	542	2CUT	581	1CPS
426	2A4L	465	2IKJ	504	2R3N	543	40636	582	1FEN
427	2B52	466	2ILZ	505	2R3O	544	3SGA	583	1FKI
428	2B53	467	2INE	506	2R64	545	1MNS	584	1GLP
429	2B55	468	2INZ	507	2R8Q	546	3PGM	585	1GLQ

586	1HFC
587	1ICN
588	1IGJ
589	1LAH
590	1MRG
591	1MUP
592	1PBD
593	1PHA
594	1PHD
595	1PHG
596	1RDS
597	1ROB
598	1SRG
599	1TDB
600	1XID
601	2ACK
602	2ADA
603	2DBL
604	3HVT
605	1COM
606	1COY
607	1DID
608	1EPB
609	1HDY
610	1HYT
611	1GKY
612	1GPB
613	1HMY
614	1PHP
615	1STO
616	1ULA
617	2YHX
618	1A6W
619	3GCH
620	3MTH

Supplementary Table 4. A comparison of different active site identifier programs on 48 bound complexes[7,8] .

Sl. No.	PDBID	Protein	Ligand	Distance[a] (Å)	[c]Rank		
					ASF[b]	Pocket Picker[7]	Ligsite[csc6]
1	1bid	Thymidylate Synthatase	UMP	1.7	1	1	1
2	1cdo	Alcohol Dehydrogeanse	NAD	1.6	2	1	1
3	1dwd	Alpha thrombin + hirudin	MID	0.9	1	1	1
4	1fbp	Phosphohydrolase	AMP	1.4	2	1	1
5	1gca	Galactose-binding protein	GAL	0.9	1	1	1
6	1hew	Acetylchitotriose	NAG	2.8	1	1	1
7	1hyt	Thermolysin	BZS	3.4	1	1	1
8	1inc	Elastase	ICL	1.0	2	1	1
9	1rbp	Retinol Binding protein	RTL	0.9	2	1	1
10	1rob	Ribonuclease A	C2P	2.0	2	1	1
11	1stp	Strptavidin	BTN	1.5	1	1	1
12	1ulb	Purine nucleoside	GUN	2.6	1	1	-
13	2ifb	Fatty acid binding protein	PLM	0.7	1	1	1
14	3ptb	Beta trypsin	BEN	2.2	1	1	1
15	2ypi	Triose Phosphate Isomerase	PGA	2.3	5	5	-
16	4dfr	Dihydrofolate reductase	MTX	1.8	1	-	1
17	4phv	HIV 1 protease	VAC	1.4	1	2	1
18	5cna	Concanavalin A	MMA	3.1	3	-	1
19	7cpa	Carboxypeptidase	FVF	1.7	1	1	1
20	1a6w	B1-8 FV fragment	NIP	1.6	1	2	1

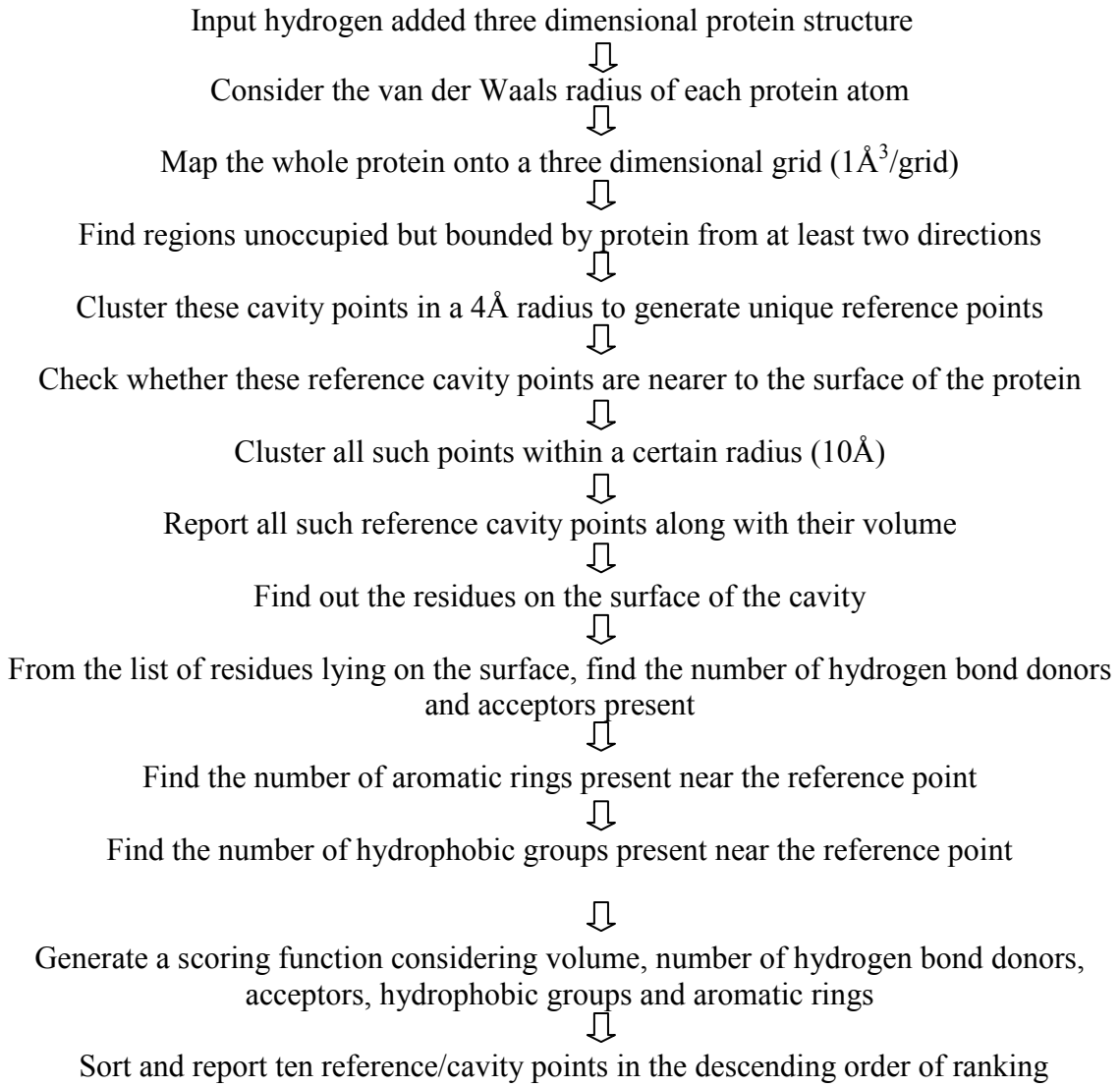
21	1apu	Penicillopepsin	IVA,VAL,VAL, STA,OET	1.3	1	1	-
22	1acj	Acetylcholinestarase	THA	1.2	1	1	-
23	1blh	Methyl Phosphonate	FOS	1.9	1	1	1
24	1byb	Beta amylase	GLC	0.4	1	1	1
25	1hfc	Fibroblast collagenase	PLH	1.8	1	1	1
26	1ida	HIV 2 protease	QND-Val-HPB-PPL-PY2	0.4	2	1	1
27	1ivd	Sialidase	ST1	1.3	2	2	1
28	1mrg	Alpha momorcharin	AND	3.3	3	-	1
29	1mtw	Trysin	DX9	0.7	1	2	1
30	1okm	Carbonic anhydrase 2	SAB	3.7	1	2	1
31	1pdz	Enolase	PGA	3.4	1	1	1
32	1phd	Camphor 5-monooxygenase	PIM	2.0	1	1	1
33	1pso	Pepsin 3a	IVA,VAL,VAL, TA,ALA,STA	0.6	1	1	1
34	1qpe	Lck kinase	PP2	1.6	4	1	2
35	1rne	Renin	C60	4.0	1	1	1
36	1snc	Staphylococcal nuclease	THP	2.1	1	1	1
37	1srf	Streptavidin	MTB	1.0	1	1	1
38	1igj	Immunoglobulin	DGX	1.2	3	4	-
39	2ctc	Carboxypeptidase	HFA	2.3	1	1	1
40	2h4n	Carbonic anhydrase 2	AZM	3.4	1	1	1
41	2pk4	Plasminogen kringle	ACA	-	-	2	1
42	2sim	Sialidase (neuraminidase)	DAN	1.5	1	2	1
43	2tmn	Thermolysin	LEP	2.0	2	1	-
44	3gch	Gamma chymotrypsin	OAC	2.8	1	1	10
45	3mth	Methylparaben insulin	MPB	4.0	5	2	9

46	5p2p	Phosphilipase	DHG	2.3	3	1	1
47	1imb	Inositol monophosphatase	LIP	1.0	1	-	1
48	6rsa	Ribonuclease	UVC	0.7	1	1	1

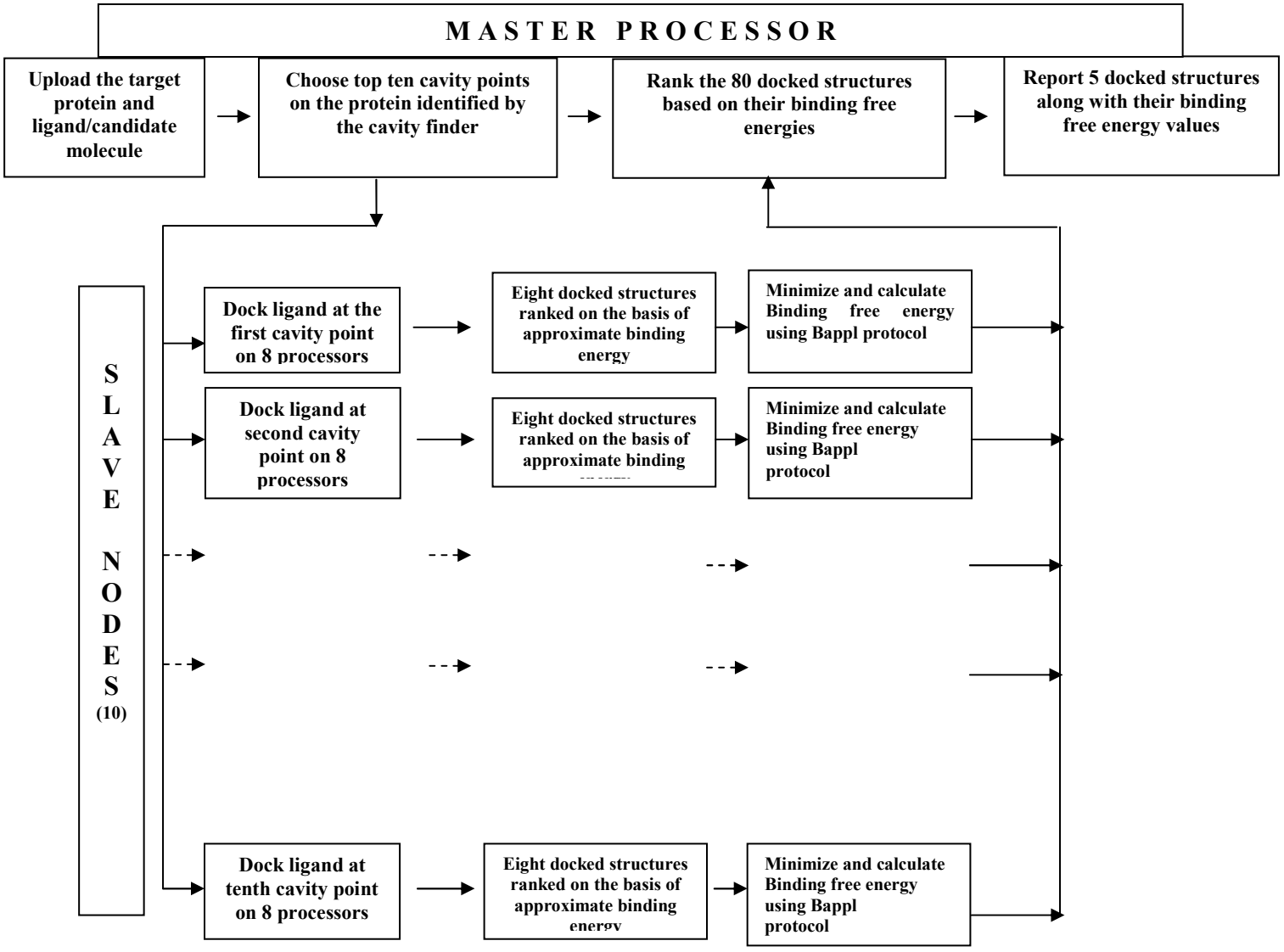
[a]Distance of the cavity point detected from any of the atom of the ligand molecule in the crystal structure by Active site identifier. (url: www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp)

[b]ASF: Active site identifier program.

[c]Rank: Rank of the cavity point generated by ASF and the center of mass of the site generated by Pocket Picker[7] and Ligsite[csc6].



Supplementary Figure 1. A schematic representation of active site identification protocol



Supplementary Figure 2. Flow chart for a parallel implementation of docking and scoring