

ContPro: A web tool for calculating amino acid contact distances in protein from 3D -structure at different distance threshold

Ahmad Firoz^{1*}, Adeel Malik¹, Obaid Afzal², Vivekanand Jha^{1,3}

¹Biomedical Informatics Center, PGIMER, Chandigarh, India; ²Faculty of Pharmacy, Jamia Hamdard, New Delhi, India; ³Department of Nephrology, PGIMER, Chandigarh, India; Ahmad Firoz – E-mail: firoz@bmi.icmr.org.in; *Corresponding Author

Received May 8, 2010; accepted June 8, 2010; published July 6, 2010

Abstract:

To investigate the functional sites on a protein and the prediction of binding sites (residues) in proteins, it is often required to identify the binding site residues at different distance threshold from protein three dimensional (3D) structures. For the study of a particular protein chain and its interaction with the ligand in complex form, researchers have to parse the output of different available tools or databases for finding binding-site residues. Here we have developed a tool for calculating amino acid contact distances in proteins at different distance threshold from the 3D-structure of the protein. For an input of protein 3D-structure, ContPro can quickly find all binding-site residues in the protein by calculating distances and also allows researchers to select the different distance threshold, protein chain and ligand of interest. Additionally, it can also parse the protein model (in case of multi model protein coordinate file) and the sequence of selected protein chain in Fasta format from the input 3D-structure. The developed tool will be useful for the identification and analysis of binding sites of proteins from 3D-structure at different distance thresholds.

Availability: It can be accessed at: <http://procarb.org/contpro/>

Background:

The function of proteins depends on their interaction with other molecules like proteins, DNA, RNA, carbohydrates and other ligands [1]. Therefore, identifying amino acid contacts is important for understanding the biological processes. In order to understand the mechanism of these interactions it is important to calculate the amino acid contacts at different distance thresholds [2-4]. Binding site residues of proteins can also be identified from pictorial databases [6], visualization tools like Ligplot [7], or many other web servers developed earlier [1, 5], but this becomes overwhelmingly imposing when a large set of proteins have to be

analyzed. With the help of ContPro (Figure 1), user can identify a binding residue by selecting a protein chain and ligand of interest and retrieve the results in the form of different output files. Additionally, it can also parse the multi model PDB file, sequence of selected protein chain from the 3D-structure of protein and gives the atomic details of contacts including distance as compared with the previous developed tools for calculating binding-site residue from PDB structures [1, 5, 9]. Protein Data Bank (PDB) is repository for 3D structures of biological macromolecules which contains coordinates of its atoms [8], and by using these coordinates of two atoms, one can compute the distance between them.



Figure 1: Screen shot of ContPro Home page. From this home page user can upload PDB file.

Methodology:

A residue is defined as a binding residue if the distance between atoms of the interacting partner is less than a certain distance cutoff [1, 9]. Upon uploading the protein 3D-structure file of interest and option selected for interacting partners by user, ContPro searches the PDB file for the protein chains, DNA chain and the number of protein models (if multi model protein). If more than one model is present, ContPro gives option to select and parse the desired model present in the uploaded PDB file. Distance threshold in angstroms, protein chain and ligand of interest can also be selected by the user. Then ContPro calculates the distance between selected protein chain residue atoms and interacting partner atoms, and when this distance falls below or equal to the selected distance threshold, this residue is considered as binding residue. The overall methodology is illustrated in figure (Figure 2).

Web interface:

Web interface of current version of ContPro was developed using HTML and CGI-PERL scripting languages. Help & Documentation, Sample Input file, Sample output file are provided at the ContPro web site.

Program input:

The input to the Contpro is a protein 3D coordinate file like PDB, modeled protein or a docked complex file. The user can select between protein-protein, protein-DNA and protein-ligand interaction type as well as protein chain, ligand, distance threshold, model number if structure has more than one model after uploading file.

Program output:

At the top of the program's result page (Figure 3), the uploaded file name and distance threshold selected for the calculations are displayed. The calculated distance between the two atoms, its residue, protein chain and the interacting atoms is displayed in a tabular form (Figure 3). The page summary section at the end has three downloadable output files. These three downloadable files are:- a) the protein sequence file in the Fasta format which was extracted from the structure of selected protein chain, b) result.txt for details about contact and c) Con.txt which has three column data. First column has residue in one letter code of selected protein chain, second column has chain name of protein and third column has 0 or 1 where 0 indicates non binding and 1 indicates binding to respective residue.

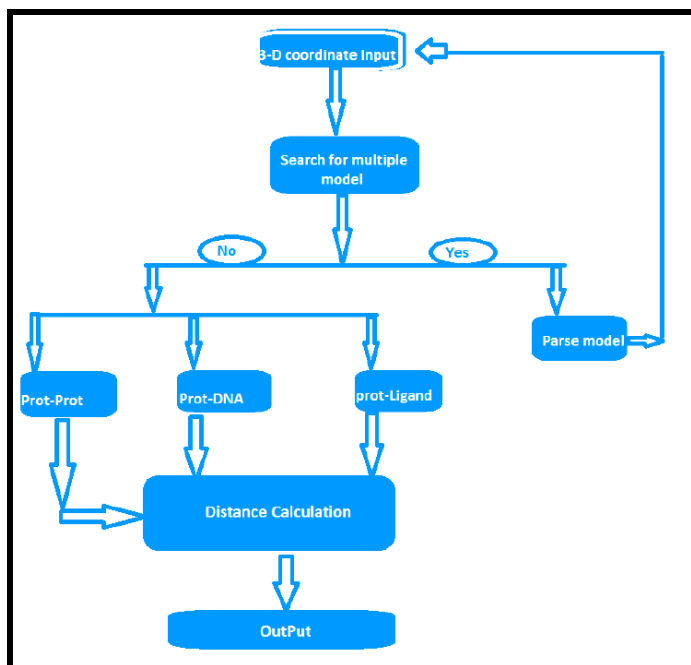


Figure 2: Flow chart of ContPro, where overall methodology is illustrated

Atom 1	Chain 1	Atom 2	Chain 2	Distance (Å)
1000	PROT	1001	PROT	1.5
1002	PROT	1003	PROT	2.0
1004	PROT	1005	PROT	1.8
1006	PROT	1007	PROT	2.2
1008	PROT	1009	PROT	1.9
1010	PROT	1011	PROT	2.1
1012	PROT	1013	PROT	1.7
1014	PROT	1015	PROT	2.3
1016	PROT	1017	PROT	1.6
1018	PROT	1019	PROT	2.4
1020	PROT	1021	PROT	1.5
1022	PROT	1023	PROT	2.0
1024	PROT	1025	PROT	1.8
1026	PROT	1027	PROT	2.2
1028	PROT	1029	PROT	1.9
1030	PROT	1031	PROT	2.1
1032	PROT	1033	PROT	1.7
1034	PROT	1035	PROT	2.3
1036	PROT	1037	PROT	1.6
1038	PROT	1039	PROT	2.4
1040	PROT	1041	PROT	1.5
1042	PROT	1043	PROT	2.0
1044	PROT	1045	PROT	1.8
1046	PROT	1047	PROT	2.2
1048	PROT	1049	PROT	1.9
1050	PROT	1051	PROT	2.1
1052	PROT	1053	PROT	1.7
1054	PROT	1055	PROT	2.3
1056	PROT	1057	PROT	1.6
1058	PROT	1059	PROT	2.4
1060	PROT	1061	PROT	1.5
1062	PROT	1063	PROT	2.0
1064	PROT	1065	PROT	1.8
1066	PROT	1067	PROT	2.2
1068	PROT	1069	PROT	1.9
1070	PROT	1071	PROT	2.1
1072	PROT	1073	PROT	1.7
1074	PROT	1075	PROT	2.3
1076	PROT	1077	PROT	1.6
1078	PROT	1079	PROT	2.4
1080	PROT	1081	PROT	1.5
1082	PROT	1083	PROT	2.0
1084	PROT	1085	PROT	1.8
1086	PROT	1087	PROT	2.2
1088	PROT	1089	PROT	1.9
1090	PROT	1091	PROT	2.1
1092	PROT	1093	PROT	1.7
1094	PROT	1095	PROT	2.3
1096	PROT	1097	PROT	1.6
1098	PROT	1099	PROT	2.4
1100	PROT	1101	PROT	1.5
1102	PROT	1103	PROT	2.0
1104	PROT	1105	PROT	1.8
1106	PROT	1107	PROT	2.2
1108	PROT	1109	PROT	1.9
1110	PROT	1111	PROT	2.1
1112	PROT	1113	PROT	1.7
1114	PROT	1115	PROT	2.3
1116	PROT	1117	PROT	1.6
1118	PROT	1119	PROT	2.4
1120	PROT	1121	PROT	1.5
1122	PROT	1123	PROT	2.0
1124	PROT	1125	PROT	1.8
1126	PROT	1127	PROT	2.2
1128	PROT	1129	PROT	1.9
1130	PROT	1131	PROT	2.1
1132	PROT	1133	PROT	1.7
1134	PROT	1135	PROT	2.3
1136	PROT	1137	PROT	1.6
1138	PROT	1139	PROT	2.4
1140	PROT	1141	PROT	1.5
1142	PROT	1143	PROT	2.0
1144	PROT	1145	PROT	1.8
1146	PROT	1147	PROT	2.2
1148	PROT	1149	PROT	1.9
1150	PROT	1151	PROT	2.1
1152	PROT	1153	PROT	1.7
1154	PROT	1155	PROT	2.3
1156	PROT	1157	PROT	1.6
1158	PROT	1159	PROT	2.4
1160	PROT	1161	PROT	1.5
1162	PROT	1163	PROT	2.0
1164	PROT	1165	PROT	1.8
1166	PROT	1167	PROT	2.2
1168	PROT	1169	PROT	1.9
1170	PROT	1171	PROT	2.1
1172	PROT	1173	PROT	1.7
1174	PROT	1175	PROT	2.3
1176	PROT	1177	PROT	1.6
1178	PROT	1179	PROT	2.4
1180	PROT	1181	PROT	1.5
1182	PROT	1183	PROT	2.0
1184	PROT	1185	PROT	1.8
1186	PROT	1187	PROT	2.2
1188	PROT	1189	PROT	1.9
1190	PROT	1191	PROT	2.1
1192	PROT	1193	PROT	1.7
1194	PROT	1195	PROT	2.3
1196	PROT	1197	PROT	1.6
1198	PROT	1199	PROT	2.4
1200	PROT	1201	PROT	1.5
1202	PROT	1203	PROT	2.0
1204	PROT	1205	PROT	1.8
1206	PROT	1207	PROT	2.2
1208	PROT	1209	PROT	1.9
1210	PROT	1211	PROT	2.1
1212	PROT	1213	PROT	1.7
1214	PROT	1215	PROT	2.3
1216	PROT	1217	PROT	1.6
1218	PROT	1219	PROT	2.4
1220	PROT	1221	PROT	1.5
1222	PROT	1223	PROT	2.0
1224	PROT	1225	PROT	1.8
1226	PROT	1227	PROT	2.2
1228	PROT	1229	PROT	1.9
1230	PROT	1231	PROT	2.1
1232	PROT	1233	PROT	1.7
1234	PROT	1235	PROT	2.3
1236	PROT	1237	PROT	1.6
1238	PROT	1239	PROT	2.4
1240	PROT	1241	PROT	1.5
1242	PROT	1243	PROT	2.0
1244	PROT	1245	PROT	1.8
1246	PROT	1247	PROT	2.2
1248	PROT	1249	PROT	1.9
1250	PROT	1251	PROT	2.1
1252	PROT	1253	PROT	1.7
1254	PROT	1255	PROT	2.3
1256	PROT	1257	PROT	1.6
1258	PROT	1259	PROT	2.4
1260	PROT	1261	PROT	1.5
1262	PROT	1263	PROT	2.0
1264	PROT	1265	PROT	1.8
1266	PROT	1267	PROT	2.2
1268	PROT	1269	PROT	1.9
1270	PROT	1271	PROT	2.1
1272	PROT	1273	PROT	1.7
1274	PROT	1275	PROT	2.3
1276	PROT	1277	PROT	1.6
1278	PROT	1279	PROT	2.4
1280	PROT	1281	PROT	1.5
1282	PROT	1283	PROT	2.0
1284	PROT	1285	PROT	1.8
1286	PROT	1287	PROT	2.2
1288	PROT	1289	PROT	1.9
1290	PROT	1291	PROT	2.1
1292	PROT	1293	PROT	1.7
1294	PROT	1295	PROT	2.3
1296	PROT	1297	PROT	1.6
1298	PROT	1299	PROT	2.4
1300	PROT	1301	PROT	1.5
1302	PROT	1303	PROT	2.0
1304	PROT	1305	PROT	1.8
1306	PROT	1307	PROT	2.2
1308	PROT	1309	PROT	1.9
1310	PROT	1311	PROT	2.1
1312	PROT	1313	PROT	1.7
1314	PROT	1315	PROT	2.3
1316	PROT	1317	PROT	1.6
1318	PROT	1319	PROT	2.4
1320	PROT	1321	PROT	1.5
1322	PROT	1323	PROT	2.0
1324	PROT	1325	PROT	1.8
1326	PROT	1327	PROT	2.2
1328	PROT	1329	PROT	1.9
1330	PROT	1331	PROT	2.1
1332	PROT	1333	PROT	1.7
1334	PROT	1335	PROT	2.3
1336	PROT	1337	PROT	1.6
1338	PROT	1339	PROT	2.4
1340	PROT	1341	PROT	1.5
1342	PROT	1343	PROT	2.0
1344	PROT	1345	PROT	1.8
1346	PROT	1347	PROT	2.2
1348	PROT	1349	PROT	1.9
1350	PROT	1351	PROT	2.1
1352	PROT	1353	PROT	1.7
1354	PROT	1355	PROT	2.3
1356	PROT	1357	PROT	1.6
1358	PROT	1359	PROT	2.4
1360	PROT	1361	PROT	1.5
1362	PROT	1363	PROT	2.0
1364	PROT	1365	PROT	1.8
1366	PROT	1367	PROT	2.2
1368	PROT	1369	PROT	1.9
1370	PROT	1371	PROT	2.1
1372	PROT	1373	PROT	1.7
1374	PROT	1375	PROT	2.3
1376	PROT	1377	PROT	1.6
1378	PROT	1379	PROT	2.4
1380	PROT	1381	PROT	1.5
1382	PROT	1383	PROT	2.0
1384	PROT	1385	PROT	1.8
1386	PROT	1387	PROT	2.2
1388	PROT	1389	PROT	1.9
1390	PROT	1391	PROT	2.1
1392	PROT	1393	PROT	1.7
1394	PROT	1395	PROT	2.3
1396	PROT	1397	PROT	1.6
1398	PROT	1399	PROT	2.4
1400	PROT	1401	PROT	1.5
1402	PROT	1403	PROT	2.0
1404	PROT	1405	PROT	1.8
1406	PROT	1407	PROT	2.2
1408	PROT	1409	PROT	1.9
1410	PROT	1411	PROT	2.1
1412	PROT	1413	PROT	1.7
1414	PROT	1415	PROT	2.3
1416	PROT	1417	PROT	1.6
1418	PROT	1419	PROT	2.4
1420	PROT	1421	PROT	1.5
1422	PROT	1423	PROT	2.0
1424	PROT	1425	PROT	1.8
1426	PROT	1427	PROT	2.2
1428	PROT	1429	PROT	1.9
1430	PROT	1431	PROT	2.1
1432	PROT	1433	PROT	1.7
1434	PROT	1435	PROT	2.3
1436	PROT	1437	PROT	1.6
1438	PROT	1439	PROT	2.4
1440	PROT	1441	PROT	1.5
1442	PROT	1443	PROT	2.0
1444	PROT	1445	PROT	1.8
1446	PROT	1447	PROT	2.2
1448	PROT	1449	PROT	1.9
1450	PROT	1451	PROT	2.1
1452	PROT	1453	PROT	1.7
1454	PROT	1455	PROT	2.3
1456	PROT	1457	PROT	1.6
1458	PROT	1459	PROT	2.4
1460	PROT	1461	PROT	1.5
1462	PROT	1463	PROT	2.0
1464	PROT	1465	PROT	1.8
1466	PROT	1467	PROT	2.2
1468	PROT	1469	PROT	1.9
1470	PROT	1471	PROT	2.1
1472	PROT	1473	PROT	1.7
1474	PROT	1475	PROT	2.3
1476	PROT	1477	PROT	1.6
1478	PROT	1479	PROT	2.4
1480	PROT	1481	PROT	1.5
1482	PROT	1483	PROT	2.0
1484	PROT	1485		

Conclusion:

The developed tool will be useful for the identification and analysis of binding sites of protein from 3D-structure at different distance threshold.

Acknowledgements:

The financial support under Project "Biomedical Informatics Centre of ICMR" is gratefully acknowledged.

References:

- [1] H Jing & Y Changhuan, *BMC Struct Biol* (2009) **9**: 52 [PMID: 19650927]
- [2] A Malik & S Ahmad, *BMC Struct Biol* (2007) **7**:1 [PMID: 17201922]
- [3] T Liu & RB Altman, *BMC Struct Biol* (2009) **9**:7 [PMID: 20003365]
- [4] M Kumar *et al. Proteins: Struct Func Bioinform* (2007), **71**:189. [PMID: 17932917].
- [5] AL Mancini *et al. Bioinformatics* (2004) **20**: 2145 [PMID: 15073001]
- [6] RA Laskowski *et al. Nucleic Acids Res* (2001) **29**: 221 [PMID: 11125097].
- [7] CA Wallace *et al. Protein Eng Design Selec* (1995) **8**: 127 [PMID: 7630882]
- [8] H. Berman *et al. Nucleic Acid Research* (2007) **35**: D301 [PMID: 17142228]
- [9] V Sobolev *et al. Bioinformatics* (1999) **15**: 327 [PMID: 10320401]

Edited by P. Kanguane

Citation: Firoz *et al.* Bioinformation 5(2) 55-57 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for noncommercial purposes, provided the original author and source are credit.