

Supplementary File

Implementation of Pseudoreceptor-based Pharmacophore Queries in the Prediction of Probable Protein Targets: Explorations on *Zea mays* Protein Structural Profile

Sivakumar Prasanth Kumar¹, Prakash C. Jha², Himanshu A. Pandya¹ & Yogesh T. Jasrai*¹

¹Department of Bioinformatics, Applied Botany Centre (ABC), University School of Sciences, Gujarat University, Navrangpura, Ahmedabad – 380009, Gujarat, India.

²School of Chemical Sciences, Central University of Gujarat, Sector-30, Gandhinagar-382030, India

*Corresponding Author.

Tel: +91 079 26302578; Fax: +91 079 26306194; E-mail: yjasrai@yahoo.com

Supplementary files information

Fig. S1 The flowchart of the proposed methodology

Fig. S2 The binding conformations of co-crystallized and GRIP docking poses of 3 ligands with its experimental proteins illustrating RMSD.

Fig. S3 The energy plot of common targets complexed with the ligands for the timescale of 2 ns

Table S1 The evaluative step of selected GRIP poses of the protein target-ligand dataset

Table S2 List of amino acids available in protein cavities of the *Zea mays* dataset

Table S3 Distribution of various types of amino acids present in protein cavities of the *Zea mays* dataset

Table S4 Enrichment factor study of probable protein targets retrieval rate

Citation details for R1-R15 of Table 7

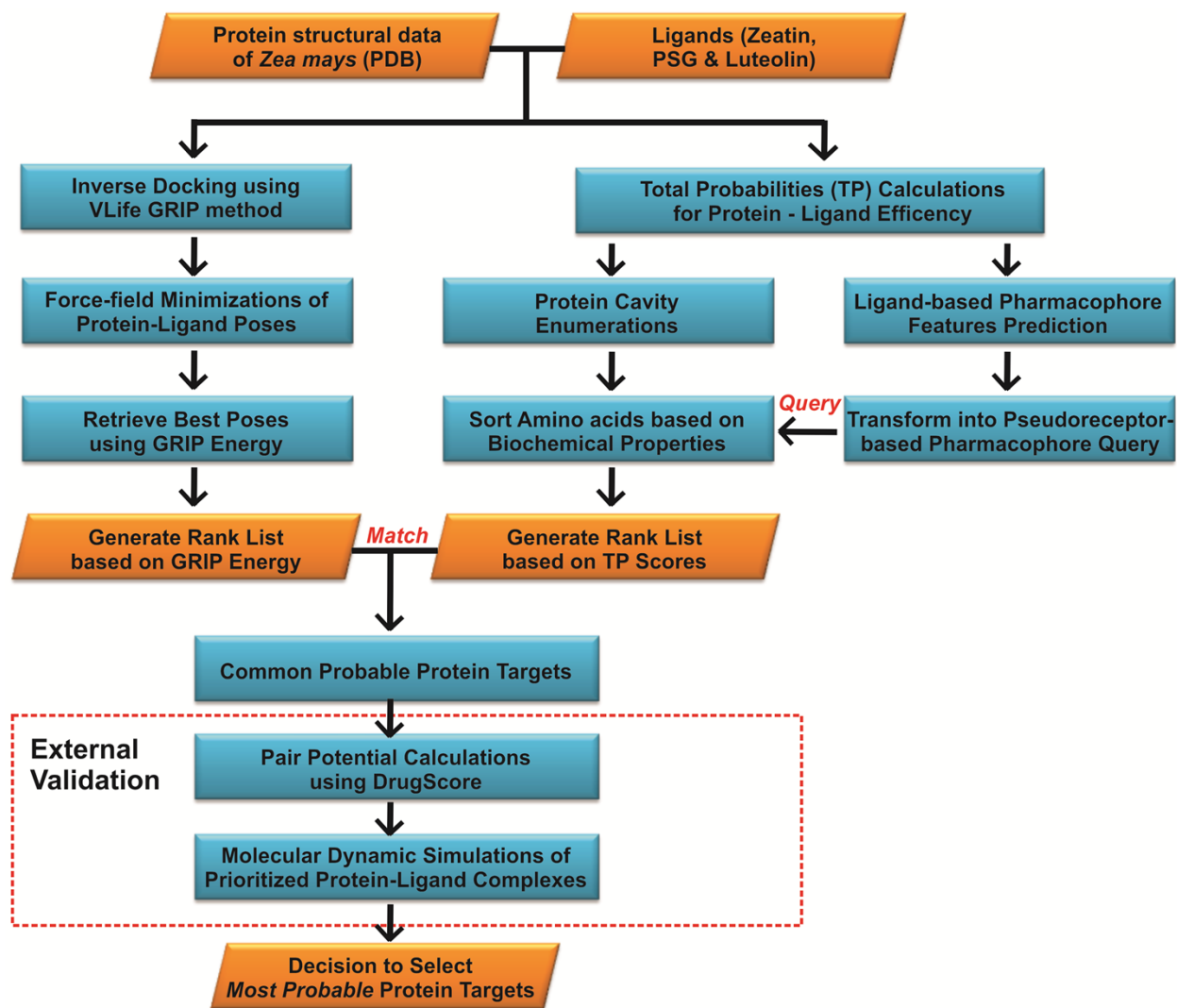


Fig. S1 The flowchart of the proposed methodology

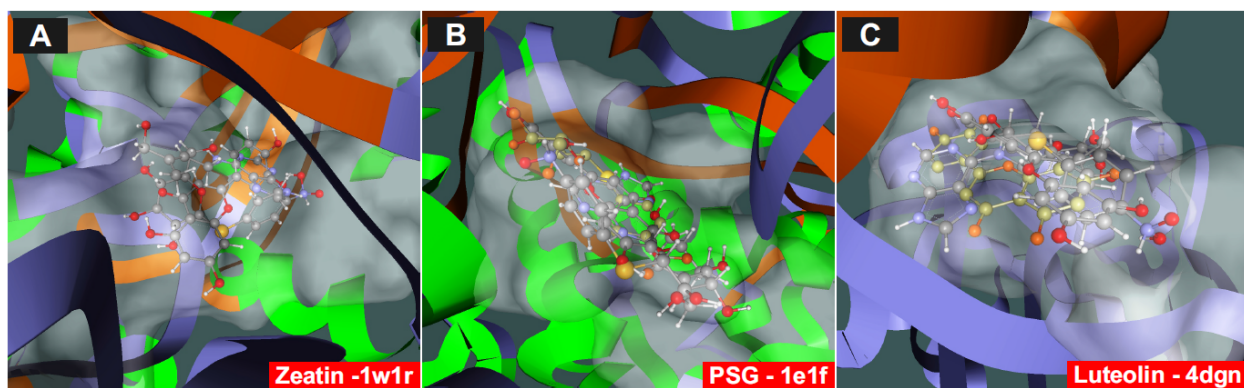


Fig. S2 The binding conformations of co-crystallized and GRIP docking poses of 3 ligands with its experimental proteins illustrating RMSD.

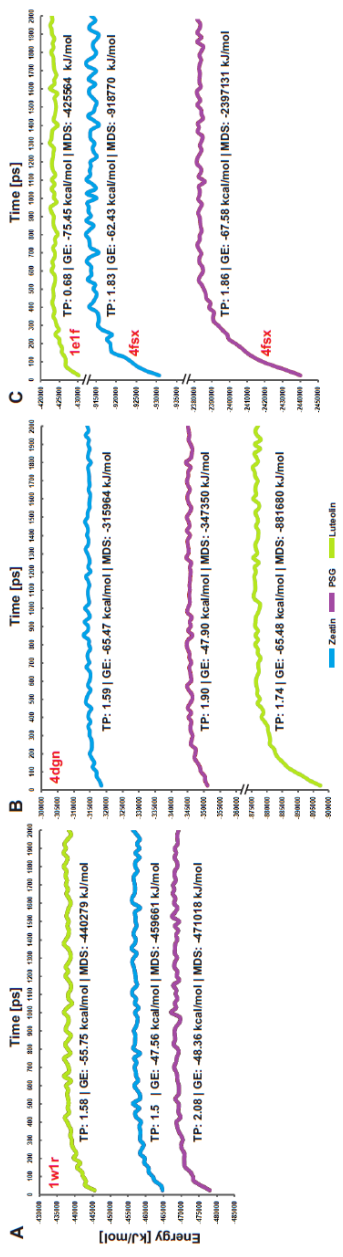


Fig. S3 The energy plot of common targets complexed with the ligands for the timescale of 2 ns.

Table S1 The evaluative step of selected GRIP poses of the protein target-ligand dataset

No.	Protein	PDB ID	GRIP evaluative step		
			Zeatin	PSG	Luteolin
1	DNA (cytosine-5)-methyltransferase 1	4fsx	27	60	73
2	4-Hydroxyphenylpyruvate dioxygenase	1sp9	71	62	74
3	Aminoaldehyde dehydrogenase 1	4i8p	43	34	21
4	Beta-expansin 1a	2hcz	46	69	16
5	Non-symbiotic hemoglobin 1	2r50	64	68	85
6	Zeamatin	1du5	37	16	63
7	Nonspecific lipid transfer protein	1mzm	58	81	26
8	Chloroplast group II intron splicing factor CRS2	1ryb	5	18	37
9	Cytokinin dehydrogenase 1 ^a	1w1r	12	76	6
10	Nucleoside N-ribohydrolase 3	4kpo	70	96	31
11	Transketolase	1itz	49	29	34
12	Glutathione S-transferase I	1axd	92	34	53
13	Bifunctional amylase/serine protease inhibitor	1bea	74	51	1
14	BX1 / Chloroplastic indole-3-glycerol phosphate lyase	1tjr	34	33	78
15	Pyruvate phosphate dikinase	1vbh	81	7	96
16	Histidine-containing phosphotransfer protein	1wn0	21	6	49
17	Adenylate kinase	1zak	61	50	6
18	NADH-dependent nitrate reductase	2cnd	12	36	40
19	Ribosome-inactivating protein 3	2pqj	12	51	74
20	Casein kinase II alpha subunit ^a	4dgn	87	11	66
21	Ferredoxin-NADP reductase	3lo8	14	21	37
22	Chloroplast pentatricopeptide repeat protein 10	4m59	25	89	5
23	Polyamine oxidase	1b37	41	92	69
24	Phosphoenolpyruvate carboxylase	1jqo	74	95	79
25	Auxin-binding protein 1	1lrh	94	24	97
26	Beta-glucosidase ^a	1e1f	5	65	41
27	Chloroplastic 9-cis-epoxycarotenoid dioxygenase 1	3npe	76	63	46
28	Glutamine synthetase	2d3a	16	8	45

^a Experimental protein targets.

Table S2 List of amino acids available in protein cavities of the *Zea mays* dataset

No.	PDB code	Type of protein structure ^a	Amino acids	Total
1	4fsx	K	Tyr347A, Ser348A, Gly349A, Cys350A, Gly351A, Gly352A, Val374A, Asp375A, Phe376A, Asn377A, Ala380A, Glu396A, Lys397A, Ala398A, Asp399A, Gly514A, Pro516A, Gln540A, Glu559A, Asn851A, Ala852A, Val853A	22
2	1sp9	P	His226B, Val228B, Glu252B, Ser263B, Leu265B, Asn266B, Ser267B, Ala268B, Val269B, Pro280B, Ile281B, Asn282B, Gln307B, His308B, Ile160A, Thr161A, Pro162A, Trp163A, Asn164A, Met169A, Trp172A, Lys187A, Pro188A, Ser189A, Glu190A, Gly218A, Leu219A, Gly220A, Pro221A, Gly224A, Ala225A, Ser228A, Phe238A, Thr239A, Gly240A, Ser241A, Phe242A, Thr244A, Lys247A, Ile248A, Glu262A, Leu263A, Gly264A, Gly265A, Cys296A, Glu395A, Phe397A, Leu423A, Trp461A, Ser467A, Glu472A	14
3	4i8p	K	Arg139X, Cys140X, Lys141X, Tyr142X, Pro143X, Ala144X, Gly145X, Gln146X, Lys147X, Ala169X, Asp170X, Phe40A, Val43A, Ala50A, Met53A, Phe54A, Lys69A, His73A, Ser76A, Val77A, Met80A, Thr99A, Arg103A, Leu104A, Thr107A, His108A, Tyr111A, Val113A, His117A, Phe118A, Thr121A, Tyr150A, Leu153A, Val154A, Ile157A	37
4	2hcz	P	Ala42B, Ala43B, Arg44B, Tyr75B, Glu84B, Tyr85B, Ala86B, Phe95B, Phe96B, Asp97B	11
5	2r50	K	Ile11A, Cys14A, Ile15A, Ala18A, Val33A, Leu36A, Asn37A, Ala40A, Arg46A, Ala49A, Leu53A, Ala56A, Ala57A, Val60A, Ala68A, Ile71A, Pro72A, Val77A, Ile79A, Pro80A, Tyr81A, Ile83A, Val92A	24
6	1du5	P	Leu102A, Arg103A, Leu104A, Gln105A, Lys106A, Lys107A, Gly108A, Asp169A, Ile184A, Val378A, Glu381A, Trp391A, Trp397A, Asn399A, Pro427A, Ile429A, Tyr431A, Ser456A, Leu458A, Tyr491A, Leu492A, Asp8A, Asp12A, Asp13A, Asn37A, Ala77A, Phe79A, Val80A, His81A, Leu123A, Gly124A, Leu149A, Ala156A, Gly157A, Asn158A, Ala159A, Pro161A, Glu164A, Ala165A, Asn166A, Leu189A, Thr193A, Tyr223A, Trp226A, His227A, Tyr231A, Phe237A, His239A, Asp240A, Leu41A, Ala76A, His78A, Gly127A, Pro128A, Leu129A, Gly167A, Asp168A, Gly169A, Cys170A, Glu173A, Asp196A, Asn198A, Ile200A, Ser201A, Ile202A, Ile261A, His275A, Ala391B, Asp392B, Leu393B, Val421B, Glu423B, Thr446B, Phe447B, Phe450B, Tyr453B, His486B	10
7	1mzm	K	Asn13A, Arg16A, Gln53A, Val54A, Pro55A, Glu66A, Ser67A, Arg68A, Glu100B, Asn104B	23
8	1ryb	P	Pro13A, His14A, Asn15A, Pro16A, Leu17A, Pro18A, Arg46A, Glu47A, Leu48A, Asp50A, Ile51A	7
9	1wlr	K	Tyr107A, Thr109A, Glu134A, Gly136A, Val137A, Pro138A, Cys139A, Pro142A, Ile144A, Asp145A, Ile149A, Val153A, Leu184A, Ser185A, Tyr186A, Ile207A, Pro209A, Leu233A, Tyr255A, Val257A, Thr263A, Gly264A, Pro265A, Gly291A, Phe292A, Gly293A, Ile294A, Ser295A, Ile312A, Ile313A, Gly314A, Ser315A, Ala316A	14
10	4kpo	P	Leu562A, Arg564A, Glu566A, Arg621A, Asp624A, Arg671A, Met748A, Glu750A, Phe770A, Gly771A, Thr772A, Asn773A, Asp774A, Cys836A, Gly837A	28
11	1itz	K	Glu25A, Gln26A, Gln28A, Gln29A, Met32A, Glu8C, Gln9C, Asn11C, Ala12C, Ala19D, Ser20D, Gly21D	28
12	1axd	K	Ala13A, Pro14A, Ala15A, Ser16A, Gly17A, Lys18A, Gly19A, Thr20A, Gln21A, Ala36A, Gly37A, Leu40A, Arg41A, Phe57A, Met58A, Gly61A,	10
13	1bea	P		11
14	1tjr	P		33
15	1vbh	K		15
16	1wn0	P		12
17	1zak	K		42

			Gln62A,Leu63A,Val64A,Val69A,Gly91A,Tyr92A,Pro93A,Arg94A, Gln98A,Arg125A,Val126A,Gly128A,Arg129A,Ile138A,Tyr139A, His140A,Tyr143A,Ser144A,Arg158A,Asp160A,Arg169A,Gly193A, Ala195A,Thr196A,Val197A,Val200A	
18	2cnd	K	His48A,Cys60A,Arg62A,Ala63A,Tyr64A,Thr65A,Leu79A,Val80A, Lys81A,Tyr83A,Phe84A,Glu87A,His88A,Phe91A,Gly94A,Gly95A, Leu96A,Met97A,Thr98A,Gln99A,Ser146A,Thr149A,Pro150A, Cys242A,Phe270A	25
19	2pqj	K	Leu93A,Tyr94A,Leu95A,Phe108A,Gly128A,Arg129A,Tyr130A, Val202A,Cys206A,Glu207A,Arg210A	11
20	4dgn	K	Val45A,Gly46A,Ser51A,Val53A,Ile66A,Lys68A,Glu81A,Val95A, Phe113A,Glu114A,Tyr115A,Val116A,Asn117A,Asn118A,Met163A, Ile174A,Asp175A,Trp176A	18
21	3lo8	K	Ser71A,Arg91A,Leu92A,Tyr93A,Ser94A,Cys112A,Val113A,Arg114A, Ala116A,Val117A,Tyr118A,Ser129A,Lys130A,Asn131A,Gly132A, Val133A,Cys134A,Ser135A,Asn136A,Thr175A,Ala178A,Glu314A, Tyr316A	23
22	4m59	P	Tyr219A,Gly220A,Arg221A,Met222A,Gly223A,Arg224A,Ser225A, Trp226A,Pro227A,Ile229A,Ala254A,Cys255A,Arg257A,Asp258A, Leu260A	15
23	1b37	K	Val10A,Gly11A,Ala12A,Gly13A,Met14A,Ser15A,Gly16A,Leu34A, Glu35A,Ala36A,Thr37A,Gly41A,Gly42A,Arg43A,Met44A,Leu56A, Gly57A,Ala58A,Asn59A,Trp60A,Glu62A,Tyr205A,Lys235A, Val236A,Val237A,Ser265A,Ala266A,Ser267A,Val270A,Leu275A, Ile276A,Tyr298A,Lys300A,Trp393A,Phe398A,Tyr399A,Thr402A, Phe403A,Gly429A,Glu430A,Gly438A,Tyr439A,Val440A,His441A, Ala443A	45
24	1jqo	P	Tyr601B,Arg610B,Ala613B,Ala614B,Trp615B,Leu617B,Tyr618B, Gln621B,Arg641B,Val645B,Gly646B,Gly648B,Gly649B,Gly650B, Pro651B,Leu654B,Ala655B,Ser658B,Gln659B,Pro660B,Val790B, Phe821B,Val824B,Thr825B,Leu826B,Leu828B,Leu829B,Val832B	28
25	1lrh	K	Leu68A,Gln94A,Asn95A,Ile12D,Ser13D,Gln17D,Ile28D,Leu45D, Lys69D,Gln118D	10
26	1e1f	K	Ser35A,Gln38A,His142A,Trp143A,Asn190A,Glu191A,Thr194A, Phe198A,Phe205A,Met263A,Asn331A,Tyr333A,Trp378A,Glu406A, Trp457A,Ser458A,Asn462A,Glu464A,Trp465A,Phe466A,Tyr473A	21
27	3npe	K	Glu132A,Arg133A,Pro134A,Pro135A,Arg193A,Asp276A,Asp278A	7
28	2d3a	K	Glu129A,Glu131A,Tyr158A,Asn190A,Glu192A,Val193A,Gln197A, Glu199A,Asn244A,Gly245A,Ala246A,Gly247A,His249A,Arg291A, His296A,Glu297A,Thr298A,Arg311A,Arg316A,Glu330A,Arg332A, Asp56B,Ser59B	23

^a K –Known binding site P-Predicted binding site

Table S3 Distribution of various types of amino acids present in protein cavities of the *Zea mays* dataset

Proteins	Type of amino acid					
	Negative charged (Acidic)	Positive charged (Basic)	Amino acid with aromatic side chain	HBA	HBD	HBA/HBD (pH dependent)
4fsx	4	1	2	4	1	5
1sp9	1	2	2	1	0	7
4i8p	4	2	6	4	5	8
2hcz	1	3	1	1	3	2
2r50	0	5	8	0	2	9
1du5	2	1	4	2	1	2
1mzm	0	1	1	0	1	2
1ryb	0	3	0	0	3	1
1w1r	2	0	4	2	2	4
4kpo	5	3	8	5	1	9
1itz	5	3	6	5	0	7
1axd	2	2	0	2	2	4
1bea	2	2	1	2	1	2
1tjr	2	0	4	2	0	8
1vbh	4	3	1	4	3	2
1wn0	2	0	0	2	0	6
1zak	1	8	5	1	7	11
2cnd	1	4	7	1	2	9
2pqj	1	2	3	1	2	2
4dgn	3	1	3	3	2	4
3lo8	1	3	3	1	3	10
4m59	1	3	2	1	4	2
1b37	3	4	9	3	5	11
1jqo	0	2	4	0	3	6
1lrh	0	1	0	0	1	5
1e1f	3	1	10	3	4	10
3npe	3	2	0	3	2	0
2d3a	7	6	3	7	4	8

Table S4 Enrichment factor study of probable protein targets retrieval rate

Statistic	Zeatin	PSG	Luteolin
Database molecules (D)	28	28	28
Actives (A)	6	11	3
Total hits (Ht)	19	20	17
Active hits (Ha)	1	1	1
% Yield of actives [(Ha/Ht)*100]	5.2631	5	5.8823
% Ratio of actives [(Ha/A)*100]	16.6667	9.0909	33.3333
Enrichment factor (EF) [(Ha*D)/(Ht*A)]	0.2456	0.1273	0.5490

Citation details for R1-R15 of Table 7

R1. I. Frébort, M. Kowalska, T. Hluska, J. Frébortová and P. Galuszka, *J. Exp. Bot.*, 2011, **62**, 2431-2452.

R2. B. Sugiharto, J. N. Burnell and T. Sugiyama, *Plant Physiol.*, 1992, **100**, 153-156.

R3. J. Veselý, L. Havlicek, M. Strnad, J. J. Blow, A. Donella-Deana, L. Pinna, D. S. Letham, J. Kato, L. Detivaud, S. Leclerc and L. Meijer, *Eur. J. Biochem.*, 1994, **224**, 771-786.

R4. M. E. Maffei, G. Arimura and A. Mithöfer, 2012. *Nat. Prod. Rep.*, 2012, **29**, 1288-1303.

R5. B. Gillissen, L. Bürkle, B. André, C. Kühn, D. Rentsch, B. Brandl and W. B. Frommer, *Plant Cell*, 2000, **12**, 291-300.

R6. K. M. Schnorr, C. Gaillard, E. Biget, P. Nygaard and M. Laloue, *Plant J.*, 1996, **9**, 891-898.

R7. L. Hecquet, J. Bolte and C. Demuyneck, *Tetrahedron*, 1994, **50**, 8677-8684.

R8. Chemical similarity searching using similarity ensemble approach (SEA). ADK_TOXGO_10000 with e value = 2.03 e-26 and MaxTC =0.31, Calculated 15 January 2014, <http://sea.bkslab.org/>

R9. J. Du, X. Zhong, Y. V. Bernatavichute, H. Stroud, S. Feng, E. Caro, A. A. Vashisht, J. Terragni, H. G. Chin, A. Tu, J. Hetzel, J. A. Wohlschlegel, S. Pradhan, D. J. Patel and S. E. Jacobsen, *Cell*, 2012, **151**, 167-80.

R10. P. Hidalgo, V. Garretón, C. G. Berríos, H. Ojeda, X. Jordana and L. Holuigue, *Plant Physiol.*, 2001, **125**, 396-405.

R11. W. Peters, B. Füchtbauer and E. Beck, *J. Plant Physiol.*, 1995, **147**, 401-407.

R12. T. Filipi, P. Mazura, L. Janda, N. S. Kiran and B. Brzobohatý, *Phytochemistry*, 2012, **74**, 40-48.

R13. G. R. Moran, *Arch. Biochem. Biophysics*, 2005, **433**, 117–128.

R14. A. Weinhold, K. Shaker, M. Wenzler, B. Schneider and I. T. Baldwin, *J. Chem. Ecol.*, 2011, **37**, 1091-1098.

R15. J. Frébortová, M. W. Fraaije, P. Galuszka, M. Sebel, P. Pec, J. Hrbác, O. Novák, K. D. Bilyeu, J. T. English and I. Frébort, *Biochem. J.*, 2004, **380(Pt 1)**, 121-130.