Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2016, 8(6):77-83

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Study of carbon distribution at glycosylation sites in Arabidopsis thaliana

Shweta Sachan[1], Parul Johri[1]*, Mala Trivedi[1], Aditi Singh[1] and Mohammed Haris Siddiqui[2]

[1]Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Malhaur, Gomti Nagar Extension, Lucknow- 226028 (UP)
[2]Department of Bioengineering, Integral University Lucknow (UP)

ABSTRACT

The post-translational addition of carbohydrate to proteins plays a significant role in overall biochemical complexity in humans (as well as all eukaryotes). The magnitude of this protein modification can be emphasized by the fact that approximately 50% of all proteins are known to be glycosylated and at least 1% of the human genome is represented by glycan biosynthesis genes. In the present work we have tried to analyze this post translational modification in a new perspective – atomic level. All the reviewed proteins of Arabidopsis thaliana were retrieved and subjected to atomic analysis using a dynamic programming code. The carbon atom percentage was calculated and analyzed. The range of carbon in N linked glycosylated protein was found to be 32.1% to 32.5 % were as the non glycosylated protein of the same species tends to attain 30.26% to 32.8% of carbon atom. The present analysis generalizes the trend of carbon atom in the fate of glycosylation of proteins.

Key words: *Arabidopsis thaliana*, carbon, glycosylation, protein.

INTRODUCTION

Proteins are essential macromolecules present within cells and are the primary source of physiological function in all biological systems. The collection of proteins within a cell determines its health and function. Many proteins catalyze biochemical reactions and are vital to metabolic processes. They are responsible for so many task of cellular life as cell shape, inner organization, waste cleanup, product manufacture and routine maintenance. Proteins are important in immune responses, cell signalling, cell adhesion and the cell cycle [1]. They are very important in the growth, cellular maintenance and functioning of the human body.

To understand protein's role and function in the human body, it is important to understand its basic composition and structure. Amino acids are the fundamental building blocks of protein. The primary sequence of protein is made up of 20 amino acids linked together by peptide bond. Histidine, lysine, isoleucine, methionine, leucine, phenylalanine, threonine, tryptophan, and valine are 9 amino acids considered as essential amino acids, which body cannot produce; they must get from the food [2]. Nonessential is a slightly misleading label because these amino acids actually play essential roles, but they can be synthesized by human body so that they're not an essential part of the diet. Arginine, glutamine, tyrosine, cysteine, glycine, proline, serine, ornithine, alanine, asparagine and aspartate are not essential amino acids.

Amino acids are present in numerous combinations in different proteins. Its arrangement determines the structure and chemical properties of the protein. They consist of the following elements: carbon (C), oxygen (O), hydrogen (H), nitrogen (N), and sulphur (S). The difference in proteins and hence their function can thus be attributed to these various sequential combinations of amino acids [3]. Its general structure consists of a carbon centre and its four substituent's, which are amino group, organic acid group, a hydrogen atom and a fourth group, referred to as the R-group, they determines its structural identity and chemical properties. They are classified as hydrophilic and hydrophobic according to the properties of their side chains. Each amino acid is assigned a hydropathy index on the basis of hydrophobic/hydrophilic properties of the side chains [4]. Carbon is the main element that is responsible for the hydrophobic nature of proteins. The presence of carbon contributes to a higher hydropathy index [5].

Post-translational modification is covalent and generally enzymatic modification of proteins during or after protein biosynthesis. Glycosylation is a form of co-translational and post-translational modification. In the reaction of glycosylation a carbohydrate is attached to a hydroxyl or other functional group of another molecule. It usually refers in particular to the enzymatic process that attaches glycans to lipids, proteins or other organic molecules. One of the fundamental biopolymers found in cells are produced by these enzymatic processes. Glycosylation is a form of post-translational and co-translational modification. Glycans plays a variety of functional and structural roles in secreted proteins and membrane [6]. In the process of glycosylation majority of proteins synthesized in the rough ER. It is an enzyme-directed site-specific process. Glycosylation is depends on where the subunits attached in the side chains. Mainly it is of two types: in N- linked glycosylation subunits attached to N atom [7] and in O- linked glycosylation subunits attached to the hydroxyl oxygen of tyrosine, serine, threonine, hydroxylysine, or hydroxyproline side-chains [8].

In the current work, the authors have divided *Arabidopsis thaliana* reviewed protein into N-linked glycosylated, O-linked glycosylated and non glycosylated proteins and tried to analyze peptide sequences at atomic level taking mainly carbon atoms into consideration. The authors calculated the carbon percentage in the peptide sequences using a dynamic programming code.

EXPERIMENTAL SECTION

In the current analysis, Lipase protein sequences were obtained from Uniprot database (www.uniprot.org). 169 reviewed protein sequences of *Arabidopsis thaliana* were taken for analysis. All the sequences were analysed for their post translational modifications. Sequences were parsed using a perl script and subjected to atomic level sequence analysis using a dynamic programming code written in perl (Figure 1). The carbon percentage was calculated and further subjected to statistical analysis using MS Excel 2010.

SPLIT THE SEQUENCES INTO ATOMS
↓
TAKE WINDOW SIZE AS 10% OF ATOMS AND SLIDE BY 1 ATOM
↓
CALCULATE THE C% OF EACH WINDOW
↓
COUNT THE FREQUENCY FOR EACH WINDOW
↓
DIVIDE THE FREQUENCY BY TOTAL NUMBER OF WINDOWS
↓
WE WILL GET SOME VALUES FOR EACH C%
↓
REPEAT THE ABOVE STEPS FOR ALL PROTEINS
↓
CALCULATE AVERAGE
↓
PLOT % C vs. FREQUENCY

Figure – 1 Dynamic Programming code steps for carbon calculation

					A. Thaliana N- linked Glycosylation									
Entry ID	**Protein Name**	**Length**	**M. Weight**	**PI**	**Glycosylation time**	**Position**	**C' Atom**	**H' Atom**	**N' Atom**	**O' Atom**	**S' Atom**	**Total No.**	**% Carbon**	
Q7XA74	Inactive GDSL esterase/lipase-like protein 25	417	46082.8	6.26	3 times	160, 308, 311	2081	3230	534	611	18	6474	32.14	32.1
Q9FLN0	GDSL esterase/lipase 1	374	41706.2	6.6	5 times	34, 184, 203, 330, 360	1875	2873	497	557	13	5815	32.32	32.3
Q9LJG3	GDSL esterase/lipase ESM1	392	44060.2	7.59	3 times	146, 166, 290	2004	3054	522	572	14	6166	32.50	32.5
Q9SYF0	GDSL esterase/lipase 2	376	42013.8	8.53	4 times	36, 186, 205, 362	1902	2906	498	551	14	5871	32.39	32.4
Q9M8Y5	GDSL esterase/lipase LTL1	366	40148.9	5.6	2 times	117, 354	1810	2801	473	528	16	5628	32.16	32.2
Q1H583	GDSL esterase/lipase 22	391	43194	6.65	3 times	105, 165, 288	1931	2955	515	573	20	5994	32.21	32.2
Q8W4H8	Inactive GDSL esterase/lipase-like protein 23	386	43143.2	8.33	3 times	105, 165, 288	1941	2982	518	562	18	6021	32.24	32.2
Q9LZA6	Triacylglycerol lipase SDP1	825	92091.7	6.55	7 times	129, 327, 331, 601, 616, 637, 800	4039	6421	1163	1225	38	12886	31.34	31.3
Q93X94	GDSL esterase/lipase EXL6	343	38596	9.49	1 time	24	1777	2753	451	480	15	5476	32.45	32.4
Q0WUV7	GDSL esterase/lipase EXL4	343	37910.9	9.84	1 time	23	1715	2677	463	474	17	5346	32.08	32.1
Q71DJ5	Triacylglycerol lipase 1	393	44230.7	5.61	2 times	41, 261	2012	3052	508	573	22	6167	32.63	32.6
Q94CH5	GDSL esterase/lipase EXL5	358	39945.7	9.69	1 time	24	1829	2857	477	490	18	5671	32.25	32.2
Q94CH8	GDSL esterase/lipase EXL1	375	41333.7	8.43	2 times	47, 334	1880	2934	488	532	14	5848	31.15	31.1
Q94CH7	GDSL esterase/lipase EXL2	379	42126.5	8.64	1 time	42	1911	2962	492	545	18	5928	32.24	32.2
Q67ZU1	Triacylglycerol lipase 2	418	46047.8	6.7	3 times	158, 286, 342	2080	3217	555	594	17	6463	32.18	32.2
Q9M1I6	Triacylglycerol lipase SDP1L	801	89933.5	6.36	9 times,	130, 328, 332, 605, 620, 649, 653,708, 759	3956	6265	1137	1183	39	12580	31.45	31.4
Q9C5N8	GDSL esterase/lipase At1g54020	372	41800.8	8.06	2 times	161, 280	1888	2891	497	542	18	5836	32.35	32.3
Q94F40	GDSL esterase/lipase At1g28600	393	43610.3	8.77	3 times	133, 317, 382	1988	3052	506	560	19	6125	32.46	32.5
Q9C857	GDSL esterase/lipase At1g31550	394	43375.5	5.45	3 times	138, 290, 322	1959	2982	502	571	21	6035	32.46	32.5
Q9FFN0	GDSL esterase/lipase At5g03810	353	38354.6	5.95	5 times	100, 255, 256, 260, 320	1715	2653	453	514	16	5351	32.05	32
Q9LU14	GDSL esterase/lipase APG	353	38199.6	8.58	2 times	197, 320	1733	2684	438	512	11	5378	32.22	32.2
Q9SZW7	GDSL esterase/lipase At4g30140	348	37470.5	5.95	1 time	342	1680	2596	456	496	11	5239	32.07	32.1
Q9FXJ2	GDSL esterase/lipase At1g28580	390	43025	5.15	2 times	140, 322	1945	2944	494	569	21	5973	32.56	32.6
Q9LY84	GDSL esterase/lipase At5g14450	389	43570.6	8.72	2 times	125, 335	1947	2995	531	568	20	6061	32.13	32.1
Q3ECP6	GDSL esterase/lipase At1g54790	408	45380.7	5.9	3 times	273, 289, 361	2058	3147	521	605	16	6347	32.42	32.4
Q9SYF5	GDSL esterase/lipase 3	367	41109.7	9.11	4 times	175, 194, 321, 351	1859	2845	501	534	11	5750	32.33	32.3
Q9LJP1	GDSL esterase/lipase 4	377	42028.7	8.68	6 times	135, 188, 194, 207, 241, 364	1909	2907	501	550	11	5878	32.48	32.5
Q9C653	GDSL esterase/lipase At1g58480	349	38838.7	8.04	2 times	25, 316	1767	2738	454	505	13	5477	32.26	32.3
Q9FPE4	GDSL esterase/lipase At1g28660	383	42033.5	4.88	5 times	105, 138, 193, 240, 320	1897	2882	478	573	15	5845	32.45	32.4
Q9SRM5	GDSL esterase/lipase CPRD49	256	28584.2	4.87	3 times	49, 79, 243	1290	1960	332	386	9	3977	32.44	32.4
Q8LFJ9	GDSL esterase/lipase 7	364	40226.8	8.52	4 times	236, 237, 264, 351	1810	2787	483	527	15	5622	32.19	32.2
Q9SSA7	GDSL esterase/lipase 5	385	43457.7	9.32	5 times	45, 66, 194, 211, 289	1963	3028	532	558	14	6095	32.2	32.2
Q9SHP6	GDSL esterase/lipase At1g28610	383	42426.5	5.93	3 times	134, 184, 315	1933	2946	482	559	17	5937	32.56	32.6
Q8LB81	GDSL esterase/lipase At5g33370	366	40359.3	6.22	2 times	118, 355	1800	2823	487	529	19	5658	31.81	31.2
Q9ZUE4	GDSL esterase/lipase At1g23500	345	37472.3	8.96	1 time	103	1689	2639	443	482	19	5272	32.04	32
Q3EAQ9	GDSL esterase/lipase At3g43550	288	32179.3	9.14	1 time	25	1478	2290	382	405	9	4564	32.38	32.4
Q9LH73	GDSL esterase/lipase At3g14820	351	39194	6.71	1 time	25	1792	2775	445	518	11	5541	32.34	32.3
F4IBF0	GDSL esterase/lipase At1g59030	349	38857.7	8.33	2 times	25, 316	1772	2731	455	502	13	5473	32.38	32.4
P0DKJ6	GDSL esterase/lipase At1g20120	402	44221.7	7.48	3 times	73, 314, 367	2000	3085	505	586	20	6196	32.28	32.3
Q9FXJ1	GDSL esterase/lipase At1g28570	389	43300.6	5.03	2 times	137, 319	1973	3001	489	570	19	6052	32.6	32.6
Q9MAA1	GDSL esterase/lipase At3g05180	379	42344	6.5	2 times	294, 330	1922	2920	500	558	12	5912	32.51	32.5
O80522	GDSL esterase/lipase At1g09390	370	40476.5	7.99	2 times	90, 315	1821	2822	482	523	20	5668	32.13	32.1
Q3ECM4	GDSL esterase/lipase At1g58725	349	38857.7	8.33	2 times	25, 316	1772	2731	455	502	13	5473	32.38	32.4
Q9C996	GDSL esterase/lipase 6	362	40224.2	8.82	5 times	50, 103, 107, 195, 296	1820	2794	476	515	20	5625	32.36	32.4
Q9LIN2	GDSL esterase/lipase At3g26430	380	42062.2	9.1	3 times	97, 115, 183	1897	2950	514	541	14	5916	32.10	32.1
P0DI15	GDSL esterase/lipase At1g59406	349	38857.7	8.33	2 times	25, 316	1772	2731	455	502	13	5473	32.38	32.4

Q9SJA9	GDSL esterase/lipase At2g24560	363	40814.9	8.74	3 times	25, 103, 325	1830	2852	486	530	21	5719	31.99	32
Q9FVV1	GDSL esterase/lipase At1g71250	374	41192.2	9.1	1 time	162	1846	2887	507	533	15	5788	31.89	31.9
Q9LZC5	GDSL esterase/lipase At5g03820	354	38543.7	6.31	7 times	66, 100, 237, 256, 257, 261, 321	1721	2664	456	518	16	5375	32.02	32
Q9C9V0	GDSL esterase/lipase At1g73610	344	37502.8	5.69	1 time	30	1692	2600	442	493	15	5242	32.28	32.3
Q9FXB6	GDSL esterase/lipase LIP-4	373	41575.7	9.47	1 time	93	1868	2900	510	531	18	5827	32.06	32.1
Q9FIA1	GDSL esterase/lipase At5g55050	376	41447.4	8.77	2 times	134, 245	1857	2908	504	544	14	5827	31.87	31.9
Q67ZI9	GDSL esterase/lipase At2g42990	350	39199.5	5.9	3 times	98, 117, 141	1770	2693	461	518	15	5457	32.44	42.4
Q6NLP7	GDSL esterase/lipase At3g62280	365	40803.9	8.53	3 times	137, 178, 231	1848	2855	489	523	16	5731	32.25	32.2
Q8L5Z1	GDSL esterase/lipase At1g33811	370	41697.3	9.37	3 times	203, 241, 242	1865	2882	518	545	13	5823	32.03	32
Q9FYD3	GDSL esterase/lipase At3g43570	320	35761.2	8.36	2 times	25, 287	1639	2528	414	461	11	5053	32.44	32.4
Q9SF94	GDSL esterase/lipase At3g09930	354	39828.7	9.24	5 times	133, 233, 237, 256, 300	1822	2789	475	506	12	5604	32.51	32.5
Q8VY93	GDSL esterase/lipase At4g26790	351	39760.2	5.43	1 time	118	1806	2749	457	528	14	5554	32.52	32.5
Q38894	GDSL esterase/lipase At1g28670	384	42427.1	5.32	3 times	105, 138, 321	1931	2921	481	569	14	5916	32.64	32.6
Q9LMJ3	GDSL esterase/lipase At1g06990	360	39981.7	5.34	5 times	26, 31, 73, 126, 272	1776	2776	468	537	22	5579	31.83	31.8
Q3E7I6	GDSL esterase/lipase At1g28650	385	42841.7	5.52	3 times	107, 140, 322	1956	2951	487	567	15	5976	32.70	32.7
Q0WPI9	GDSL esterase/lipase At3g53100	351	38692.3	8.7	3 times	234, 254, 318	1739	2700	468	501	16	5424	32.06	32.1
Q9FHW9	GDSL esterase/lipase At5g42170	369	40529.6	9.18	4 times	28, 45, 203, 336	1818	2909	489	537	10	5763	31.55	31.5
Q9M153	GDSL esterase/lipase At4g01130	382	42013.3	9.09	4 times	118, 263, 275, 330	1913	2951	499	537	15	5915	32.34	32.3
Q8RWJ4	GDSL esterase/lipase At2g36325	356	39958.9	9.44	4 times	25, 165, 185, 240	1821	2815	473	515	12	5636	32.31	32.3
Q8RXT9	GDSL esterase/lipase At1g28590	403	44811.2	5.11	2 times	139, 323	2026	3104	512	594	21	6257	32.38	32.4
Q9FJ45	GDSL esterase/lipase At5g45910	372	41533.6	8.98	3 times	66, 101, 137	1875	2886	494	538	18	5811	32.27	32.3
P0C8Z7	GDSL esterase/lipase At1g28640	390	43509.1	5.07	4 times	105, 138, 321, 364	1970	2976	498	587	15	6046	32.58	32.6
Q9FFC6	GDSL esterase/lipase At5g22810	362	39540	5.84	4 times	159, 162, 264, 329	1786	2757	459	531	12	5545	32.21	32.2
Q9SIF3	GDSL esterase/lipase At2g04020	322	35962.3	8.91	1 time	260	1626	2523	425	466	15	5055	32.17	32.2
Q9LII9	GDSL esterase/lipase At3g27950	371	40778	9.1	5 times	82, 143, 178, 194, 315	1853	2881	481	527	14	5756	32.19	32.2
O23470	GDSL esterase/lipase At4g16230	368	40242.2	6	2 times	117, 286	1813	2860	474	535	12	5694	31.84	31.8
Q9SIQ2	GDSL esterase/lipase At2g31550	360	40203.4	8.03	3 times	26, 104, 326	1820	2823	469	520	19	5651	32.21	32.2
O64469	GDSL esterase/lipase At2g19060	349	38724.2	6.99	1 time	178	1747	2692	456	510	15	5420	32.23	32.2
Q9LMS5	GDSL esterase/lipase At1g18120	256	28969.3	9.08	1 time	181	1321	2036	348	368	9	4082	32.36	32.4
Q5PNZ0	GDSL esterase/lipase At5g18430	362	40123.2	9.31	2 times	117, 355	1797	2827	493	515	17	5649	31.80	31.8
Q9LZS9	GDSL esterase/lipase At5g03590	344	38427.2	9.5	3 times	126, 227, 238	1742	2701	459	492	15	5409	32.21	32.2
Q9SVU5	GDSL esterase/lipase At4g28780	367	39896.5	4.9	2 times	119, 356	1778	2789	473	537	16	5593	31.79	31.8
Q9STM6	GDSL esterase/lipase At3g48460	381	42235	7.61	3 times	112, 240, 258	1904	2929	491	566	15	5905	32.24	32.2
Q9LZS7	GDSL esterase/lipase At5g03610	359	40106	9.04	3 times	136, 236, 259	1840	2808	470	515	11	5644	32.60	32.6
O64468	GDSL esterase/lipase At2g19050	349	38577.1	8.76	1 time	49	1740	2667	459	501	17	5384	32.32	32.3
Q9ZQI3	GDSL esterase/lipase At2g27360	394	43853.8	5.16	4 times	136, 319, 371, 382	1977	3000	500	588	21	6086	32.48	32.5
Q9CA68	GDSL esterase/lipase At1g74460	366	40696.4	5.73	2 times	113, 260	1831	2822	480	539	16	5688	32.19	32.2
Q9SJB4	GDSL esterase/lipase At2g04570	350	38477.9	7.55	3 times	98, 117, 343	1742	2654	442	507	18	5363	32.48	32.5
O80470	GDSL esterase/lipase At2g23540	387	42083	8.37	3 times	139, 159, 380	1882	2957	505	561	14	5919	31.79	31.8
Q9SIQ3	GDSL esterase/lipase At2g31540	360	40121.3	7.54	2 times	104, 326	1821	2821	467	518	18	5645	32.26	32.3
O80443	GDSL esterase/lipase At2g38180	312	35405.2	5.42	1 time	79	1595	2463	431	465	9	4963	32.14	32.1
Q9FHQ1	GDSL esterase/lipase At5g37690	356	39210.7	8.84	2 times	116, 291	1760	2734	472	518	13	5497	32.12	32.1
Q9FJ41	GDSL esterase/lipase At5g45950	357	39854.3	8.34	2 times	37, 132	1813	2832	462	509	19	5635	32.17	32.2
Q9C648	GDSL esterase/lipase At1g58430	360	40402.4	8.93	3 times	22, 104, 326	1816	2824	474	529	20	5663	32.07	32.1
O22927	GDSL esterase/lipase At2g30310	359	39738.7	8.64	2 times	103, 325	1792	2795	471	516	17	5591	32.05	32
Q9M2R9	GDSL esterase/lipase At3g50400	374	41498.3	8.98	2 times	104, 125	1865	2909	511	541	11	5837	31.95	31.9
O65921	GDSL esterase/lipase At2g19010	344	38162.4	8.99	1 time	303	1701	2637	465	501	17	5321	31.97	32
Q9LZB2	GDSL esterase/lipase At5g03980	323	35694.5	8.31	1 time	77	1600	2442	434	466	15	4957	32.28	32.3
O22918	GDSL esterase/lipase At2g30220	358	39671.6	6.93	3 times	25, 102, 324	1801	2798	456	522	15	5592	32.21	32.2
Q9FNP2	GDSL esterase/lipase At5g08460	385	42417.4	5.08	4 times	218, 285, 368, 378	1898	2931	501	561	21	5912	32.10	32.1
Q9SIF5	GDSL esterase/lipase At2g03980	367	40757.7	7.52	3 times	25, 191, 303	1834	2837	475	538	19	5703	32.16	32.2
Q9FJ40	GDSL esterase/lipase At5g45960	375	42481.8	8.71	1 time	340	1951	2967	493	544	14	5969	32.69	32.7
O22527	Chlorophyllase-1	324	34854.8	5.44	1 time	36	1580	2433	401	462	13	4889	32.32	32.3

Q9FNA9	Phospholipid:diacylglycerol acyltransferase 1	671	74156.7	6.5	4 times	161, 381, 434, 647	3327	5100	892	961	37	10317	32.25	32.2
Q9FXE5	Alpha-L-fucosidase 3	372	40461.6	8.29	4 times	96, 114, 139, 182	1823	2756	492	530	13	5614	32.47	32.5

	A. Thaliana No glycosylation										
Entry ID	**Protein Name**	**Length**	**M. Weight**	**PI**	**C' Atom**	**H' Atom**	**N' Atom**	**O' Atom**	**S' Atom**	**Total No.**	**% Carbon**
Q9S745	Lipase-like PAD4	541	60984.6	6.01	2735	4261	723	811	23	8553	31.98%
Q9MA46	Galactolipase DONGLE, chloroplastic	471	52800	9.01	2347	3696	656	701	15	7415	31.65%
O49523	Phospholipase A1-Iigamma	419	47791.9	5.53	2139	3298	582	640	12	6671	32.06%
Q9C7N4	GDSL esterase/lipase At1g29670	363	39871.8	8.85	1768	2713	499	529	14	5523	32.01%
Q94CH6	GDSL esterase/lipase EXL3	364	39815.8	5.59	1806	2822	450	533	14	5625	32.11%
Q9C7N5	GDSL esterase/lipase At1g29660	364	40141.8	5.19	1782	2711	487	548	13	5541	32.16%
Q9LJP2	GDSL esterase/lipase At3g14220	363	40248.6	6.37	1803	2777	489	533	13	5615	32.11%
Q9SF78	GDSL esterase/lipase At1g71691	384	41724.6	5.34	1874	2891	499	548	17	5829	32.15%
Q93YW8	GDSL esterase/lipase At4g18970	361	39598.3	5.3	1737	2662	488	536	20	5443	31.91%
Q9SIZ6	GDSL esterase/lipase At2g40250	361	39884.9	8.76	1801	2837	473	524	12	5647	31.89%
Q9FMK6	GDSL esterase/lipase At5g63170	338	36912.5	8.47	1666	2622	440	480	13	5221	31.90%
O23469	GDSL esterase/lipase At4g16220	245	26592.4	8.93	1198	1874	316	352	8	3748	31.96%
Q9FK75	GDSL esterase/lipase At5g45670	362	39575.5	5.96	1745	2693	491	532	16	5477	31.86%
Q9SU72	Protein EDS1	623	71690.2	5.73	3212	4975	867	957	20	10031	32.02%
Q680C0	GDSL esterase/lipase At4g10955	350	39473.9	7.88	1782	2751	497	505	8	5543	32.15%
Q9FJ25	GDSL esterase/lipase At5g41890	369	40742.1	5.89	1823	2820	488	547	13	5691	32.03%
Q4F883	Senescence-associated carboxylesterase 101	537	62066.4	8.33	2798	4381	733	813	24	8749	31.98%
Q9FM04	GDSL esterase/lipase At5g62930	242	27156.6	5.99	1213	1862	334	359	9	3777	32.12%
Q9MAA7	Gibberellin receptor GID1A	345	38617.8	6.16	1736	2682	478	503	10	5409	32.09%
Q6NMR9	GDSL esterase/lipase At5g45920	241	27153.9	4.87	1219	1893	323	362	9	3806	32.03%
Q9SIN9	Phospholipase A1-Ialpha2, chloroplastic	484	54923.6	8.18	2444	3827	667	727	23	7688	31.79%
Q9LZS8	GDSL esterase/lipase At5g03600	322	35723.1	9.27	1610	2514	428	461	15	5028	31.02%
Q940G6	Gibberellin receptor GID1C	344	38425.6	6.67	1728	2682	478	501	8	5397	32.02%
Q9LYC1	Gibberellin receptor GID1B	358	40299.7	6.86	1819	2794	496	525	9	5643	32.24%
Q9SJI7	Phospholipase A1-Iidelta	412	46057	5.12	2056	3195	547	629	13	6440	31.93%
Q9XF23	Protein EDS1L	623	71572.1	5.77	3203	4969	869	955	20	10016	31.98%
O23522	Phospholipase A1-Ibeta2, chloroplastic	517	57640.4	9.33	2525	4041	751	763	16	8096	31.19%
Q3EBR6	Phospholipase A1-Igamma2, chloroplastic	529	60402.3	6.57	2661	4178	776	797	19	8431	31.56%
Q9C942	Caffeoylshikimate esterase	332	36975.2	6.12	1669	2543	419	491	20	5142	32.46%
Q941F1	Phospholipase A1-Igamma1, chloroplastic	515	58445.1	8.79	2586	4037	753	762	18	8156	31.71%
Q9C8J6	Phospholipase A1-Igamma3, chloroplastic	527	60321	6.18	2695	4147	733	808	18	8401	32.08%
O48723	Patatin-like protein 2	407	44239.3	5.57	1990	3135	525	602	6	6258	32.80%
Q948R1	Phospholipase A(1) DAD1, chloroplastic	447	49960	9.23	2211	3517	633	655	16	7032	31.44%
O22975	1-acylglycerol-3-phosphate O-acyltransferase	418	46525.8	9.32	2108	3223	575	599	10	6515	32.36%
Q9SU71	Protein EDS1B	629	71631.2	5.76	3213	4966	866	954	20	10019	32.07%
Q9LY09	Oleosin GRP-17	543	53192.9	10.34	2255	3752	688	721	36	7452	30.26%
Q4VCM1	Phospholipid--sterol O-acyltransferase	633	71687.7	5.94	3255	4930	840	935	28	9988	32.59%
O23179	Patatin-like protein 1	414	45823.6	6.36	2041	3285	539	625	14	6504	31.38%
O23180	Patatin-like protein 5	414	45642	6.06	2030	3250	538	629	12	6459	31.43%
Q9M7I7	Chlorophyllase-2, chloroplastic	318	34904	6.51	1564	2444	414	461	15	4898	31.93%
Q94AS5	Isoprenylcysteine alpha-carbonyl methylesterase ICME	427	47581.5	6.11	2142	3349	571	625	15	6702	31.96%
O82274	Phospholipase A1-Iibeta	414	47580.8	5.85	2143	3305	579	632	9	6668	32.14%
O23512	Probable pheophorbidase	262	29014	5.39	1324	2019	335	385	7	4070	32.53%
O23181	Patatin-like protein 3	428	47203.8	5.85	2094	3334	568	641	15	6652	31.48%
B3H7A9	Probable peroxygenase 7	210	23851.2	9.62	1066	1680	300	310	6	3362	31.71%
Q93V61	Phospholipase A(1) LCAT3	447	50348.6	5.72	2281	3505	589	660	19	7054	32.34%

F4HX15	Phospholipase A I	1309	145221.5	5.61	6404	10273	1769	1945	64	20455	31.31%
Q1PET6	Probable isoprenylcysteine alpha-carbonyl methylesterase ICMEL2	422	47009.7	7.68	2109	3280	584	605	16	6594	31.98%
P40602	Anther-specific proline-rich protein APG	534	58007.5	9.61	2657	4136	680	731	22	8226	32.30%
P29525	Oleosin 18.5 kDa	173	18569	9.43	816	1308	236	249	5	2614	31.22%
Q39165	Oleosin 21.2 kDa	199	21279.3	9.36	940	1497	265	280	9	2991	31.43%
Q8VYP9	Probable isoprenylcysteine alpha-carbonyl methylesterase ICMEL1	476	52718.7	6.36	2357	3633	637	703	18	7348	32.10%
Q9LNC2	Phospholipase A1-IIalpha	423	48122.5	5.95	2157	3352	584	644	11	6748	31.97%
Q9SMN0	Probable carboxylesterase 12	324	35567.3	5.23	1589	2461	417	484	13	4964	32.01%
Q9SMM9	Probable carboxylesterase 13	329	36099.5	5.36	1622	2489	427	494	7	5039	32.19%
Q42431	Oleosin 20.3 kDa	191	20313.1	6.91	901	1438	252	272	5	2868	31.42%
Q9FVW3	Putative methylesterase 14, chloroplastic	348	38854.7	8.97	1717	2734	472	512	21	5456	31.47%
Q9LMA7	Probable carboxylesterase 1	318	36081	5.45	1639	2496	430	471	10	5046	32.48%
Q8LED9	Probable carboxylesterase 16	446	49368.2	8.03	2206	3432	616	642	16	6912	31.92%
Q9LFR7	Probable carboxylesterase 17	344	37698	8.99	1674	2601	485	477	17	5254	31.86%
Q9SX25	Probable carboxylesterase 6	336	37149.7	6.46	1642	2605	457	485	20	5209	31.52%
O64640	Probable carboxylesterase 8	329	36424.2	6	1615	2513	443	488	15	5074	31.83%
Q9SX78	Probable carboxylesterase 2	314	35255.3	5.17	1593	2457	411	465	14	4940	32.25%
Q71N54	Lecithin-cholesterol acyltransferase-like 4	535	60430.6	5.04	2717	4166	708	809	23	8423	32.26%
Q43284	Oleosin 14.9 kDa	141	14852.2	9.86	658	1073	187	193	5	2116	31.10%
Q9FYC7	Putative phospholipid:diacylglycerol acyltransferase 2	665	73652.6	8.69	3292	5164	902	953	32	10343	31.83%
Q9FZI8	Lecithin-cholesterol acyltransferase-like 1	432	48307.6	8.73	2186	3416	578	625	16	6821	32.05%
Q9SS98	Oleosin 5	183	19754.7	7.11	880	1412	248	258	5	2803	31.40%

RESULTS AND DISCUSSION

In the current analysis, all the reviewed protein sequences of *Arabidopsis thaliana* from Uniprot database were considered for the analysis. In which 101 sequences were N-linked glycosylated and 69 were non glycosylated. No O-linked glycosylation and both (N & O) linked glycosylation were found. The atomic level analysis of the N linked glycosylated protein revealed that the total carbon atom percentage ranged from 32.1% to 32.5 % and in the non glycosylated sequences the range of carbon content was recorded between 30.26% to 32.8%.

CONCLUSION

The present study of glycosylated protein in Arabidopsis thaliana provides a comprehensive picture of the role of carbon atom in the post translational modifications of proteins. Carbon has been a vital atom in the protein structure that provides the basic hydrophobic or hydrophilic nature to the proteins. The demarcation in the carbon percentage in glycosylated and non glycosylated proteins benchmarks the fate of post translational modification. This represents a new perspective of sequence analysis.

Acknowledgement:
Authors are grateful to Dr. A. K. Chauhan, Founder President & Mr. Aseem Chauhan, Chancellor Amity University Haryana & Chairperson AMITY Lucknow for providing necessary facilities and support. We also extend our gratitude to Maj. Gen. K.K Ohri, AVSM (Retd.), Pro Vice Chancellor, Amity University, Uttar Pradesh Lucknow Campus for constant support and encouragements.

REFERENCES

[1] C Chothia; AM Lesk, *The EMBO Journal*., **1986**, 5(4), 823–836.
[2] P Johri, *Int. J. Comput. Bioinfo. In Silico Model*., **2013**, 2(4), 173-179.
[3] P Johri; M Gokhale, *J. Comput. Biol*., **2013**, 2 (1), 1-6.
[4] P Anandagopu; S Suhanya; V Jayaraj; E Rajasekaran, *Bioinformation*, **2008**, 2(7), 304-307.
[5] V Jayaraj; M Vijayasarathy; R Geerthana et al, *Int. J. Computl. Intell. in Bioinfo*., **2009**, 2(2), 99–102.
[6] A Varki. Essentials of Glycobiology, 2nd Edition, Cold Spring Harbor Laboratories Press, New York, US, **2008**, ISBN 978-0-87969-770-9.
[7] E Maverakis; K Kim; M Shimoda; M Gershwin et al, *J. Autoimmun*., **2015**, 57 (6), 1-13.
WG Flynne. Biotechnology and Bioengineering, 1st Edition, Nova Science Publishers, New York, US, **2010**, ISBN 978-1-60456-067-1.