

ORIGINAL RESEARCH

**OPEN ACCESS**

Full open access to this and thousands of other papers at <http://www.la-press.com>.

## Evolutionary Pattern of N-Glycosylation Sequon Numbers in Eukaryotic ABC Protein Superfamilies

R. Shyama Prasad Rao, Ole Thomsen Buus and Bernd Wollenweber

Aarhus University, Department of Genetics and Biotechnology, Forsøgsvej 1, Slagelse 4200, Denmark.  
Email: [rao@agrsci.dk](mailto:rao@agrsci.dk)

---

**Abstract:** Many proteins contain a large number of NXS/T sequences (where X is any amino acid except proline) which are the potential sites of asparagine (N) linked glycosylation. However, the patterns of occurrence of these N-glycosylation sequons in related proteins or groups of proteins and their underlying causes have largely been unexplored. We computed the actual and probabilistic occurrence of NXS/T sequons in ABC protein superfamilies from eight diverse eukaryotic organisms. The ABC proteins contained significantly higher NXS/T sequon numbers compared to respective genome-wide average, but the sequon density was significantly lower owing to the increase in protein size and decrease in sequon specific amino acids. However, mammalian ABC proteins have significantly higher sequon density, and both serine and threonine containing sequons (NXS and NXT) have been positively selected—against the recent findings of only threonine specific Darwinian selection of sequons in proteins. The occurrence of sequons was positively correlated with the frequency of sequon specific amino acids and negatively correlated with proline and the NPS/T sequences. Further, the NPS/T sequences were significantly higher than expected in plant ABC proteins which have the lowest number of NXS/T sequons. Accordingly, compared to overall proteins, N-glycosylation sequons in ABC protein superfamilies have a distinct pattern of occurrence, and the results are discussed in an evolutionary perspective.

**Keywords:** ABC proteins, evolution, N-glycosylation sequons, probability, proline

---

*Bioinformatics and Biology Insights* 2010:4 9–17

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



## Introduction

Proteins are frequently modified by the attachment of sugars (glycans)—a process known as glycosylation—which in turn affects a number of biological processes.<sup>1–6</sup> Based on the site of attachment, there can be two major types, namely, O-linked and N-linked glycosylation. While the former occurs on the hydroxyl group of serine (S) or threonine (T) amino acids, which are usually present in abundance, N-glycosylation occurs only on the amide group of asparagine (N) present in the three amino acid motifs NXS/T (where X is any amino acid except proline, which is known to limit the N-glycosylation due to conformational hindrance).<sup>1,4,7–10</sup> Although NXS/T sequons occur frequently and not all sequons are N-glycosylated due to a variety of reasons including conformational limitations,<sup>8–10</sup> the NXS/T sequons are essential, but could be limited by probabilistic occurrence.<sup>11</sup>

For instance, in an ideal protein which contains 20 types of amino acids in equal proportions, there is just a 0.00475 probability of obtaining a NXS/T sequon as against a 0.1 probability of observing a serine or threonine residue per amino acid. That is, it is over 21 times (0.1/0.00475) less likely to find an N-glycosylation sequon compared to an O-glycosylation site in a protein. This raises interesting questions. What is the number of N-glycosylation sequons found in a protein or groups of proteins? Do they follow the expected probability? Is there any trend in the occurrence of sequon numbers among proteins? What factors direct the evolution of such a pattern?

The prevalence of NXS/T sequons was previously studied in the sequence entries of Swiss-Prot database and based on the results it has been presumed that over two third of all proteins could be potential N-glycoproteins.<sup>1</sup> In a recent study, Cui et al<sup>12</sup> showed that N-glycosylation sequon numbers vary over four folds among phylogenetically diverse eukaryotes with N-glycan dependent quality control of the proteins. Further, a Darwinian selection for NXT (but not for NXS) sequons was observed in secreted proteins of eukaryotes and envelop proteins of viruses. This is due to the fact that there is an increased conditional probability that asparagine and threonine occur in sequons rather than elsewhere in the protein sequence.<sup>12,13</sup> However, it is not known, if a Darwinian selection may also be expected for NXS sequons in some proteins.

Our interest was to see the pattern of occurrence of N-glycosylation sequon numbers in closely related ATP Binding Cassette (ABC) proteins—which form a large group of proteins with important biological functions.<sup>14</sup> The majority of the members of this superfamily are transmembrane proteins and many (such as permeable glycoprotein and multidrug resistance protein) are well known N-glycoproteins.<sup>15</sup> Further, many ABC proteins contain an abundance of NXS/T sequons. For example, mouse and human ABCA13 proteins contain 77 and 80 NXS/T sequons, respectively. On the other hand, many ABC proteins in plants (for example, *Arabidopsis* and rice) contain no NXS/T sequon at all. It is interesting to see such a large difference in the NXS/T sequon numbers in closely related proteins. Is this an expected pattern? What are the underlying causes for this distinction? We answer this question by making a comparative analysis of NXS/T sequons in ABC protein superfamilies from eight diverse eukaryotic organisms and examine the factors which may affect the number of sequons in the proteins.

## Materials and Methods

### Sequence acquisition

We used the ABC protein sequences from eight diverse eukaryotic organisms for which a near complete ABC protein inventory exists and sequences are available in the database. This includes two plant species—*Arabidopsis thaliana* (Arath) and *Oryza sativa* (Rice),<sup>14</sup> two invertebrate species—*Caenorhabditis elegans* (Worm) and *Drosophila melanogaster* (Fly),<sup>16</sup> two vertebrate species—*Mus musculus* (Mouse) and *Homo sapiens* (Human),<sup>17</sup> and two unicellular fungal/yeast species—*Saccharomyces cerevisiae*<sup>18</sup> and *Candida albicans*,<sup>19</sup> for which ABC protein lists can be found in the respective references. The mouse ABC protein sequences were collected from the database based on the human orthologs (<http://nutrigene.4t.com/humanabc.htm>). The ABC protein sequences were obtained from Uniprot database (<http://www.uniprot.org/>). Further, all available protein entries for the said organisms were also retrieved from the database. These entries contained both reviewed and un-reviewed sequences.

### Frequency of NXS/T sequons

The number of NXS/T sequons (NXS and NXT, where X is any amino acid except proline) and



NPS/T sequences were counted programmatically in each protein sequence. The number of NXS/T per 100 amino acids was considered as the density of the sequons. The percentage of N, P, S and T amino acids in the protein sequences were also enumerated from their respective frequency over the total frequency of amino acids.

### Predicting NXS/T sequon number

By considering the protein sequence as a Markov chain,<sup>11</sup> the probabilistic occurrence of NXS/T sequons may be computed from the transition of amino acid frequencies. For example, an ideal protein with 400 residues containing 20 types of amino acids in equal proportions has a total NXS/T probability of 0.00475 and therefore 1.9 predicted NXS/T sequon number. This is the product of two separate transitions from N to X and NX to S or T ( $20/400 * 380/400 * 40/400$ ). For example, the mouse ABCA1 protein (P41233) is 2261 amino acids long and contains 23 NXS/T sequons (with just one known N-glycosylation at position 481). However, following the probabilistic calculation ( $104/2261 * 2151/2261 * (179 + 114)/2261 * 2261 = 12.82$ ) would give the mouse ABCA1 just 13 NXS/T sequons. The number of NPS/T sequences were also predicted using the above mentioned procedure.

### Data analysis

All the sequence analyses and data handling were done using programs written in the Python programming language (ver. 2.6, <http://www.python.org/>). The

Biopython (<http://biopython.org/>) tools were also used in sequence parsing. Data visualization, graphing and statistical analyses were performed on a spreadsheet (Excel 2003/2007, Microsoft, USA). A regression line was fitted to the scatter plots of mean ( $\pm$ SEM) values and Pearson's correlation coefficient ( $r$ ) was computed between two variables. Owing to the large sample size, parametric tests were favored and means were declared significant at  $p < 0.05$ . A paired test was used to see if the difference between actual and predicted mean values were significant. A Z-test for two proportions or two sample means was used wherever appropriate.

## Results

### Pattern of N-glycosylation sequon numbers in ABC proteins

The percentage of sequences with at least one NXS/T sequon and thus has the potential to be N-glycoprotein is significantly higher (Z test for two proportions:  $Z = 2.3$  to  $5.6$ ,  $p < 0.05$ ) in ABC protein superfamilies compared to the proportion in the respective Uniprot entries (Table 1). The proportion for Uniprot entries varies from ~59% (in rice) to ~87% (in *Candida*) with an average of ~75%. On the other hand, more than 95% (except rice, which has ~84%) of ABC proteins contain at least one NXS/T sequon.

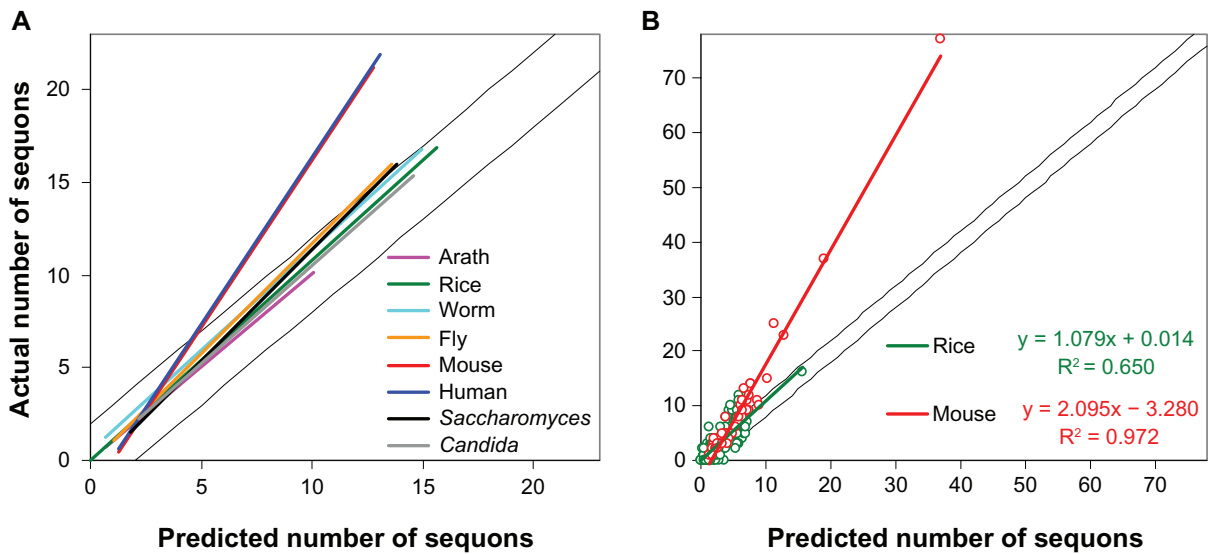
Figure 1A shows the regression lines for scatter plots of actual versus predicted number of sequons in ABC proteins. Although the actual number of sequons in some ABC proteins are much higher or lower than the predicted number (above or below the

**Table 1.** The number of total proteins and *potential* N-glycoproteins in ABC protein superfamilies and Uniprot entries.

Organism <sup>1</sup>	ABC superfamily		Uniprot entries	
	Total proteins	N-glycoproteins	Total proteins	N-glycoproteins
Arath	129	122 (94.6%)**	51598	38443 (74.5%)
Rice	126	106 (84.1%)**	142664	84768 (59.4%)
Worm	54	52 (96.3%)**	24642	19589 (79.5%)
Fly	56	56 (100%)**	33204	25481 (76.7%)
Mouse	47	47 (100%)**	67871	49300 (72.6%)
Human	49	49 (100%)**	100599	71282 (70.9%)
<i>Saccharomyces</i>	30	30 (100%)**	24025	19888 (82.8%)
<i>Candida</i>	28	28 (100%)*	14529	12580 (86.6%)

<sup>1</sup>*Arabidopsis thaliana* (Arath), *Oryza sativa* (Rice), *Caenorhabditis elegans* (Worm), *Drosophila melanogaster* (Fly), *Mus musculus* (Mouse), *Homo sapiens* (Human), *Saccharomyces cerevisiae* and *Candida albicans*.

\*\*significantly different (Z test for two proportions,  $p < 0.05$ ) compared to Uniprot entries (\*one tailed).

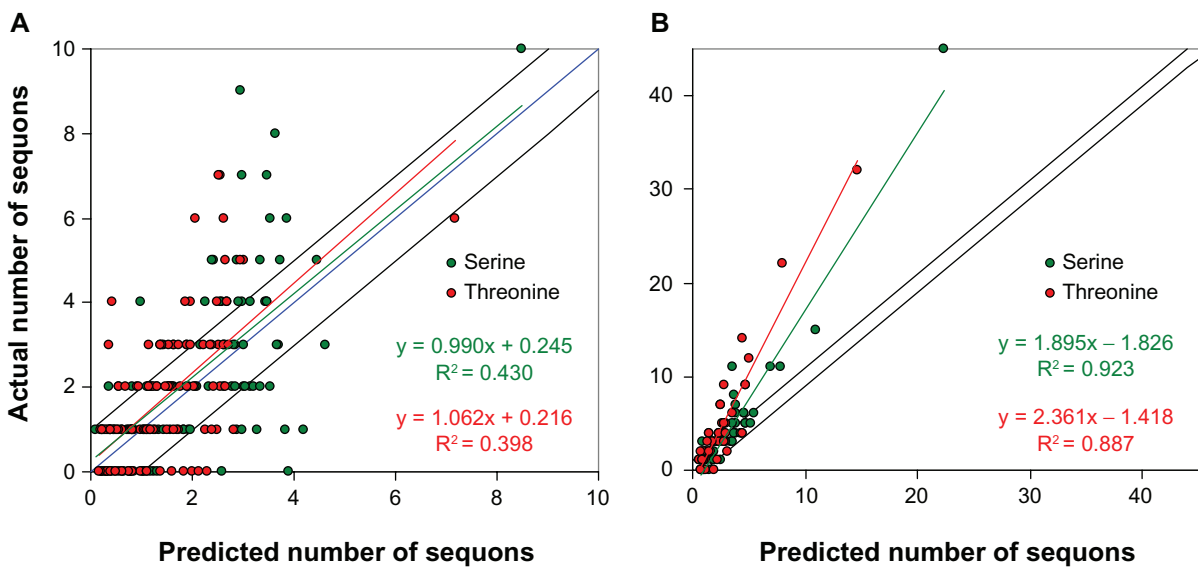


**Figure 1.** The N-glycosylation sequons in ABC proteins. **A)** The regression lines for the scatter plots (not shown) of actual versus predicted number of NXS/T sequons in most ABC proteins has a slope close to 1. However, most of the ABC proteins in mouse and human have actual sequon numbers far greater than predicted number (two outlying points in mouse and human were removed to make the scale smaller). **B)** The scatter plots for actual versus predicted number of sequons in rice and mouse ABC proteins are elaborated.

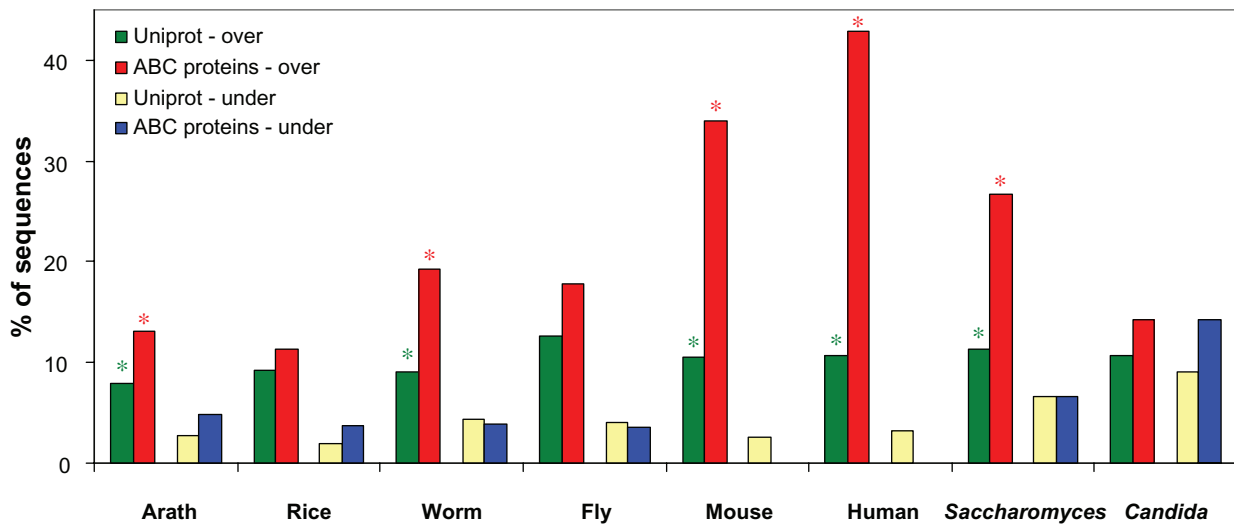
two diagonally parallel lines—a margin of  $\pm 2$  sequons was considered as the rounding error from predicted probability to predicted frequency for two separate sequons NXS and NXT), for example, as shown for rice ABC proteins (Fig. 1B), the overall trend line had a slope close to 1. Further, the actual versus predicted sequon numbers are highly correlated ( $r > 0.9$ ). However, many ABC proteins in mouse

and human contain NXS/T sequons far higher than the predicted number (Fig. 1A and 1B) and therefore the overall trend line strongly deviates from the normal slope of 1.

A similar pattern may be observed even when NXS and NXT sequons are considered separately, as shown, for instance in rice and mouse (Fig. 2A and 2B). For example, most sequences in rice ABC protein



**Figure 2.** The NXS and NXT sequons in rice and mouse ABC proteins. **A)** The scatter plot of actual versus predicted number of NXS and NXT sequons in rice shows many sequences with actual sequons higher or lower than the predicted number (above or below the  $\pm 1$  diagonally parallel lines), but the regression line has a slope close to 1. **B)** The scatter plot of sequons in mouse reveals that most sequences have actual sequon number much higher than the predicted number and therefore, the regression line has a slope far greater than 1. It may be noted that NXS sequons are also positively selected.

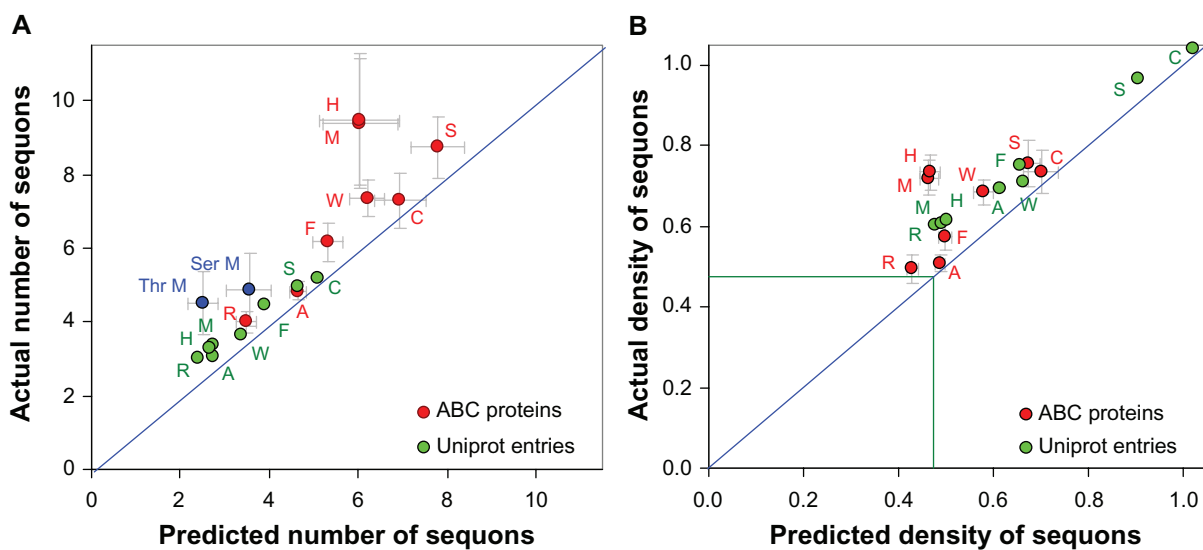


**Figure 3.** The over-representation of sequons in ABC proteins. The proportion of sequences which have over-representation (actual sequon number is more than predicted sequon number plus two) of NXS/T sequons is significantly higher (Z test for two proportions:  $Z = 2.0$  to  $7.0$ ,  $p < 0.05$ ) in majority of ABC protein superfamilies compared to the respective Uniprot entries (indicated by asterisks).

superfamily have NXS and NXT sequons close to prediction (bounded by  $\pm 1$  diagonally parallel lines) with only a few sequences deviating (above or below), and therefore the trend line is very close to diagonal line (Fig. 2A). However, many ABC proteins in mouse have far higher than the predicted number of sequons—both NXS and NXT (Fig. 2B). In fact, the percentage of sequences with NXS/T sequons much higher than expected (over-representation) is significantly higher in many ABC protein superfamilies compared to the

respective Uniprot entries (Fig. 3). On the other hand, there is no statistical difference in the percentage of sequences with sequons under-represented.

The mean number of NXS/T sequons varies from  $\sim 4.0$  (in rice) to  $\sim 9.5$  (in human) for ABC proteins and from  $\sim 3.0$  (in rice) to  $\sim 5.2$  (in *Candida*) for Uniprot entries. Further, the mean number of actual sequons is higher than the mean of predicted number of sequons (all points are above diagonal line in Fig. 4A). This difference is especially noticeable for



**Figure 4.** The sequon number and density. **A**) The mean number of NXS/T sequons is significantly higher (paired test,  $p < 0.05$ ) than the predicted number in most ABC protein superfamilies. Further, they are much higher than the mean number of sequons in the respective Uniprot entries. The mean number of NXS (Ser-M) and NXT (Thr-M) sequons are shown only for mouse ABC proteins. **B**) The mean sequon density in ABC proteins is lower (except for mouse and human) compared to the respective Uniprot entries. Green lines show the sequon density (0.475) in an ideal protein.



ABC proteins and significant (paired test:  $t=3.5$  and  $3.8$ ,  $p < 0.01$ ) for mouse and human ABC proteins. A very similar trend can be observed even when NXS and NXT sequons are considered separately, for instance, as shown in Figure 4A for mouse ABC proteins.

### Sequon density of ABC proteins

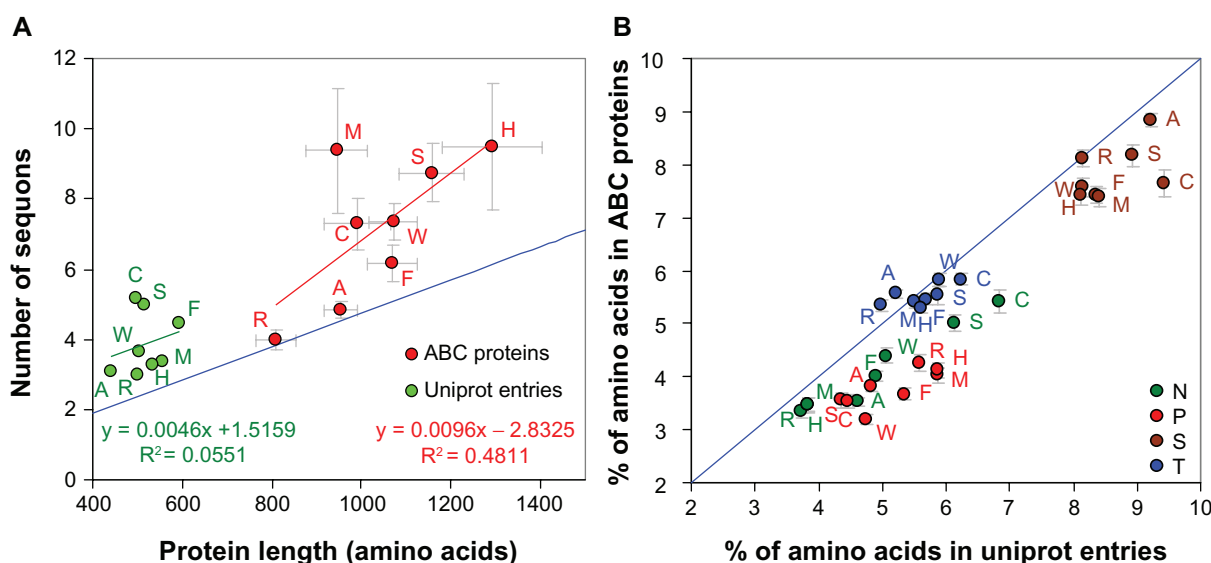
In contrast to high mean sequon number, ABC proteins showed lower sequon density (number of sequons per 100 amino acids) compared to the respective Uniprot entries (Fig. 4B). The mean sequon density varied from 0.50 (in rice) to 0.76 (in *Saccharomyces*) for ABC proteins and from 0.60 (in rice) to 1.04 (in *Candida*) for Uniprot entries. Noticeably, mouse and human ABC proteins have higher sequon density (from all sequences concatenated) compared to the density in respective Uniprot entries (Fig. 4B). It may be noted that an ideal protein with 20 types of amino acids in equal proportion has an expected NXS/T sequon density of 0.475 (green lines in Fig. 4B). On the contrary, we can find more than one sequon for every hundred amino acids in a typical protein from *Candida*.

Compared to Uniprot entries which have an average protein size of ~500 amino acids, the ABC proteins are nearly double in size (Fig. 5A). However, except for mouse and human, the number of sequons

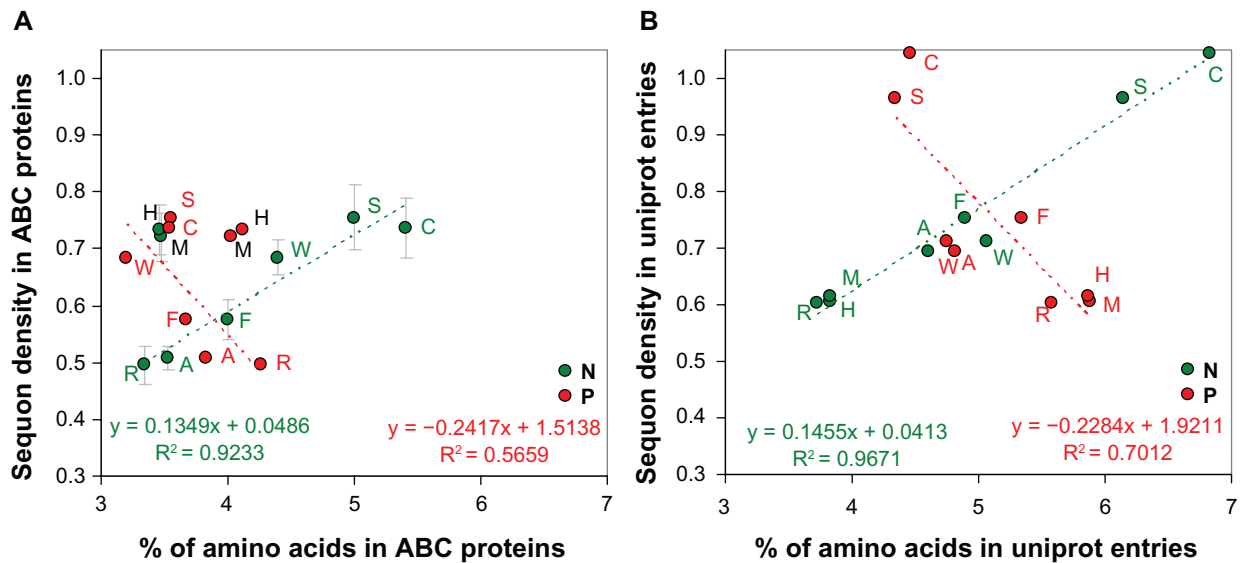
in ABC proteins is not doubled to keep pace with the increase in protein size. Further, the percentage of sequon specific amino acids—asparagine, serine and threonine, and proline are lower (below diagonal line) in most ABC protein superfamilies compared to the respective Uniprot entries (Fig. 5B).

### Sequon density is correlated to percentage of amino acids

The density of sequons is correlated with the amount of sequon specific amino acids—asparagine, serine and threonine. As shown in the Figure 6A, the sequon density in ABC proteins is positively correlated ( $r=0.96$ ) with percentage of asparagine and negatively correlated ( $r=-0.75$ ) with percentage of proline. For example, rice ABC protein superfamily which is low in sequon density has lower percentage of asparagine and higher percentage of proline. This trend is quite reverse for *Candida* which has high sequon density. A very similar, but stronger correlation ( $r=0.98$  for asparagine and  $r=-0.84$  for proline) has been found for sequons in Uniprot entries (Fig. 6B). It may be noted (Fig. 6A) that mouse and human ABC protein superfamilies are clear outliers—they have high sequon density in spite of low in asparagine and high in proline content.



**Figure 5.** The ABC sequences are longer and contain less sequon specific amino acids. **A)** Compared to Uniprot entries, ABC proteins are nearly double in size (amino acid chain length) but number of sequons is not doubled (except mouse and human ABC proteins). Inclined line shows the number of sequons (1.9 per 400 amino acids) in an ideal protein. **B)** Sequon specific amino acids (N, S and T) and proline in ABC protein superfamilies are lower compared to the respective Uniprot entries.



**Figure 6.** The sequon density is correlated to amino acid frequencies. **A**) The sequon density is positively correlated ( $r = 0.96$ ) to percentage of asparagine and negatively correlated ( $r = -0.75$ ) to proline in ABC proteins. However, mouse and human ABC protein superfamilies are clear outliers. **B**) A similar correlation ( $r = 0.98$  for asparagine and  $r = -0.84$  for proline) exists for sequon density in Uniprot entries.

## Role of NPS/T sequences

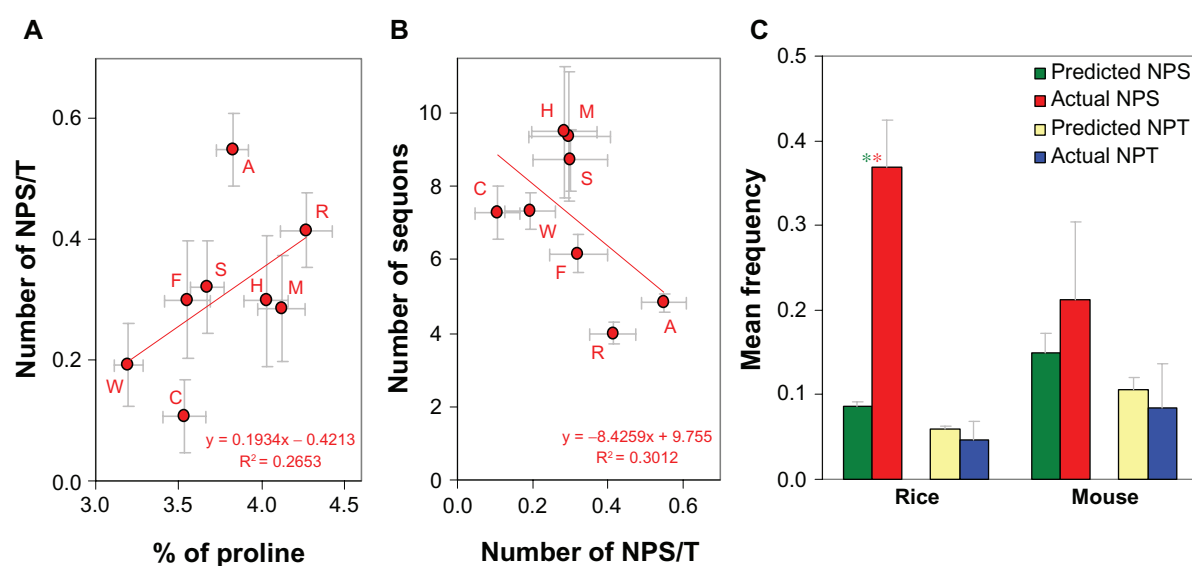
The mean number of NPS/T sequences in ABC proteins varies from 0.11 (in *Candida*) to 0.55 (in *Arabidopsis*) and is positively correlated ( $r = 0.52$ ) with the percentage of proline (Fig. 7A). Further, the number of NXS/T sequons is negatively correlated ( $r = -0.55$ ) with the mean number of NPS/T sequences in ABC proteins (Fig. 7B). It may be noted that the actual number of NPS/T sequences is significantly different (paired test,  $p < 0.05$ ) from predicted number in *Arabidopsis* (0.55 versus 0.18), rice (0.42 versus 0.15) and *Candida* (0.11 versus 0.25) ABC protein superfamilies (data not shown). While the actual number of NPS/T sequences is significantly higher in *Arabidopsis* and rice, it is significantly lower in *Candida*. Further, it is the number of NPS sequence, but not NPT, which is significantly different (paired test:  $t = 5.1$ ,  $p < 0.05$ ), for example, as shown for rice ABC proteins in Fig. 7C. The actual versus predicted number of NPS/T sequences are not significantly different in Uniprot entries.

## Discussion

The NXS/T sequons (where X is any amino acid except proline) have attracted much attention in recent years due to their role as potential N-glycosylation sites.<sup>2,20,21</sup> Several studies have focused on identifying the N-glycosylation sites either experimentally or

by predicting the likelihood of NXS/T sequons to be N-glycosylated.<sup>10,22–25</sup> However, studies are limited on the pattern of occurrence of N-glycosylation sequons in proteins or groups of proteins and in particular, the underlying causes responsible for such a pattern.<sup>12,13,26</sup> In this study, we made a comparative analysis of N-glycosylation sequons in ABC protein superfamilies and the respective genome-wide proteins from diverse eukaryotic organisms.

The ABC protein superfamilies contain significantly higher number of N-glycosylation sequons compared to the respective Uniprot entries. Many membrane N-glycoproteins and viral coat proteins were found to contain a high number of NXS/T sequons,<sup>12,13</sup> and most ABC sequences have been predicted to contain transmembrane domains (sequence annotation results in <http://www.uniprot.org/>) and thus likely to be membrane N-glycoproteins.<sup>1,15</sup> On the other hand, the low mean number of sequons in Uniprot entries is comparable to the mean sequon numbers found in Swiss-Prot entries by a previous study.<sup>1</sup> Further, a significantly higher proportion (~95%) of ABC sequences contains at least one NXS/T sequons compared to the respective Uniprot entries (~75%). This is expected, as most ABC proteins are known transmembrane N-glycoproteins as against Uniprot entries which also contain cytosolic proteins and unlikely to be N-glycoproteins.<sup>1,12</sup> Overall, plants contain lowest number of sequons and



**Figure 7.** The role of NPS/T sequences. **A)** The number of NPS/T sequences is positively correlated ( $r = 0.52$ ) with the percentage of proline. **B)** The number of NXS/T sequons is negatively correlated ( $r = -0.55$ ) with NPS/T sequences. **C)** The actual number of NPS/T sequences is significantly higher than predicted number, for example, in rice which has low sequon number, but not in mouse. Further, this difference is significant (paired test:  $t = 5.1$ ,  $p < 0.01$ ) only for NPS sequence and not for NPT sequence.

unicellular fungi contain highest number of sequons. The observed pattern of sequon numbers could very well be correlated with the amount of sequon specific amino acids. For instance, unicellular fungi contain highest amount of asparagine, serine and threonine and so is the number of sequons. In a recent study, the AT content of nucleotide sequence has been positively correlated with the content of sequon specific asparagine residue and thus sequons, as asparagine is encoded by AT-rich codons.<sup>12</sup> We did not quantify or correlate the AT content with asparagine, however, find a clear correlation between asparagine and sequon numbers.

In spite of high sequon numbers, ABC proteins showed low sequon density. This could be due to the fact that ABC proteins have longer chain length but the sequon specific amino acids are lower compared to the respective Uniprot entries. However, the mouse and human ABC proteins contain high sequon numbers compared to the respective genome-wide average. Further, a significantly higher sequon number and density compared to the expected number clearly indicates positive selection for N-glycosylation sequons in these organisms. In a recent study,<sup>12</sup> it has been shown that there is a Darwinian selection for NXT sequons in secreted and membrane glycoproteins of eukaryotes and viruses. This is due to the fact that the conditional probabilities of finding asparagine

and threonine in NXT sequons are much higher than other locations in the protein sequence.<sup>12,13</sup> Here, we showed that NXS sequons too have experienced a clear Darwinian selection, albeit to a slightly lesser extent compared to the NXT sequons, at least in mammalian ABC proteins. In fact, all ABC protein superfamilies contain a large number of sequences with NXT and NXS sequons higher than predicted number, indicating a selective advantage of N-glycosylation sequons in these proteins or groups of proteins.

The presence of proline in or around NXS/T sequences was known to obstruct the N-glycosylation due to conformational hindrance.<sup>1,9,10</sup> Consequently, the NPS/T sequence may be used as one of the partial mechanisms to modulate the number of sequons in a protein during the evolution. Here, we found a positive correlation between NPS/T sequences and proline, and a negative correlation between NXS/T sequons and NPS/T in ABC proteins. Further, the observed number of NPS/T sequences (with a clear bias for NPS) is shown to be significantly higher in ABC protein superfamilies with low sequon numbers.

In conclusion, we showed a distinct pattern in the occurrence of N-glycosylation sequon numbers in ABC proteins from eight diverse eukaryotic organisms. The high sequon number was correlated with the content of sequon specific amino acids. Proline and





NPS sequence may have a partial mechanistic role in modulating sequon numbers. Finally, a clear Darwinian selection for NXS sequons was also observed in ABC proteins.

## Disclosures

The authors report no conflicts of interest.

## References

1. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochem Biophys Acta*. 1999;1473:4–8.
2. Delos SE, Burdick MJ, White JM. A single glycosylation site within the receptor-binding domain of the avian sarcoma/leukosis virus glycoprotein is critical for receptor binding. *Virology*. 2002;294:354–63.
3. Dennis JW, Granovsky M, Warren CE. Protein glycosylation in development and disease. *Bioessays*. 1999;21:412–21.
4. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002;12:43R–56R.
5. Verki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 1993;3:97–130.
6. Weber ANR, Morse MA, Gay NJ. Four N-linked glycosylation sites in human toll-like receptor 2 cooperate to direct efficient biosynthesis and secretion. *J Biol Chem*. 2004;279:34589–94.
7. Yan A, Lennarz WJ. Unraveling the mechanism of protein N-glycosylation. *J Biol Chem*. 2005;280:3121–4.
8. Ben-Dor S, Esterman N, Rubin E, et al. Biases and complex patterns in the residues flanking protein N-glycosylation sites. *Glycobiology*. 2004;14:95–101.
9. Gavel Y, von Heijne G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng*. 1990;3:433–42.
10. Petrescu A-J, Milac A-L, Petrescu SM, et al. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology*. 2004;14:103–14.
11. Wu G. Frequency and Markov chain analysis of amino acid sequences of human glutathione reductase. *Biochem Biophys Res Commun*. 2000;268:823–6.
12. Cui J, Smith T, Robbins PW, et al. Darwinian selection for sites of Asn-linked glycosylation in phylogenetically disparate eukaryotes and viruses. *Proc Natl Acad Sci U S A*. 2009;106:13421–6.
13. Bushkin GG, Ratner DM, Cui J, et al. Suggestive evidence for Darwinian selection against asparagine-linked glycans of *Plasmodium* and *Toxoplasma*. *Eukaryotic Cell*. 2009;doi:10.1128/EC.00197–09.
14. Verrier PJ, Bird D, Burla B, et al. Plant ABC proteins—a unified nomenclature and updated inventory. *Trends Plant Sci*. 2008;13:151–9.
15. Hipfner DR, Almquist KC, Leslie EM, et al. Membrane topology of the multidrug resistance protein (MRP)—a study of glycosylation-site mutants reveals an extracytosolic NH<sub>2</sub> terminus. *J Biol Chem*. 1997;272:23623–30.
16. Sturm A, Cunningham P, Dean M. The ABC transporter gene family of *Daphnia pulex*. *BMC Genomics*. 2009;10:170.
17. Dean M. The genetics of ATP-binding cassette transporters. *Methods Enzymol*. 2005;400:409–29.
18. Bauer BE, Wolfger H, Kuchler K. Inventory and function of yeast ABC proteins: about sex, stress, pleiotropic drug and heavy metal resistance. *Biochim Biophys Acta*. 1999;1461:217–36.
19. Gaur M, Choudhury D, Prasad R. Complete inventory of ABC proteins in human pathogenic yeast, *Candida albicans*. *J Mol Microbiol Biotechnol*. 2005;9:3–15.
20. Wojczyk BS, Takahashi N, Levy MT, et al. N-glycosylation at one rabies virus glycoprotein sequon influences N-glycan processing at a distant sequon on the same molecule. *Glycobiology*. 2005;15:655–66.
21. Sasaki K, Nagamine N, Sakakibara Y. Support vector machine prediction of N- and O-glycosylation sites using whole sequence information and subcellular localization. *IPSJ Transactions on Bioinformatics*. 2009;2:25–35.
22. Caragea C, Sinapov J, Silvescu A, et al. Glycosylation site prediction using ensembles of Support Vector Machine classifiers. *BMC Bioinformatics*. 2007;8:438.
23. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing*. 2002;7:310–22.
24. Hamby SE, Hirst JD. Prediction of glycosylation sites using random forests. *BMC Bioinformatics*. 2008;9:500.
25. Zhang H, Loriaux P, Eng J, et al. UniPeP—a database for human N-linked glycosites: a resource for biomarker discovery. *Genome Biol*. 2006;7:R73.1–R73.12.
26. Belair M, Dovat M, Foley B, et al. The polymorphic nature of HIV type 1 env V4 affects the patterns of potential N-glycosylation sites in proviral DNA at the intrahost level. *Aids Res Hum Retroviruses*. 2009;25:199–206.

**Publish with Libertas Academica and every scientist working in your field can read your article**

*“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”*

*“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”*

*“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”*

### Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>