Journal of Biomolecular Structure and Dynamics, 2017
Vol. 35, No. 11, 2337–2350, https://doi.org/10.1080/07391102.2016.1217275

Physicochemical sequence characteristics that influence *S*-palmitoylation propensity

Krishna D. Reddy[a,†], Jashwanth Malipeddi[a,†], Shelly DeForte[a], Vikas Pejaver[c], Predrag Radivojac[c], Vladimir N. Uversky[a,b] and Robert J. Deschenes[a]*

[a]*Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd., MDC 07, Tampa, FL 33612, USA;* [b]*Johnnie B. Byrd Alzheimer's Research Institute, University of South Florida, Tampa, FL 33612, USA;* [c]*Department of Computer Science and Informatics, Indiana University, Bloomington, IN 47405, USA*

Communicated by Ramaswamy H. Sarma

(*Received 17 June 2016; accepted 18 July 2016*)

Over the past 30 years, several hundred eukaryotic proteins spanning from yeast to man have been shown to be *S*-palmitoylated. This post-translational modification involves the reversible addition of a 16-carbon saturated fatty acyl chain onto the cysteine residue of a protein where it regulates protein membrane association and distribution, conformation, and stability. However, the large-scale proteome-wide discovery of new palmitoylated proteins has been hindered by the difficulty of identifying a palmitoylation consensus sequence. Using a bioinformatics approach, we show that the enrichment of hydrophobic and basic residues, the cellular context of the protein, and the structural features of the residues surrounding the palmitoylated cysteine all influence the likelihood of palmitoylation. We developed a new palmitoylation predictor that incorporates these identified features, and this predictor achieves a Matthews Correlation Coefficient of .74 using 10-fold cross validation, and significantly outperforms existing predictors on unbiased testing sets. This demonstrates that palmitoylation sites can be predicted with accuracy by taking into account not only physiochemical properties of the modified cysteine and its surrounding residues, but also structural parameters and the subcellular localization of the modified cysteine. This will allow for improved predictions of palmitoylated residues in uncharacterized proteins. A web-based version of this predictor is currently under development.

Keywords: acyltransferase; palmitoylation; palmitoylomics; post-translational modification; intrinsically disordered protein

Introduction

S-palmitoylation is the enzyme-catalyzed, reversible addition of a 16-carbon palmitoyl moiety to a cysteine residue via a thioester linkage (the term 'palmitoylation' will refer to *S*-palmitoylation hereafter, unless otherwise indicated). Palmitoylation can affect protein function in several ways: it can promote membrane association and/or distribution within membrane subdomains, alter protein conformation through tilting of transmembrane domains (TMDs), or affect protein–protein interactions (Linder & Deschenes, 2007). While other types of lipidation, such as myristoylation and prenylation, are irreversible, palmitoylation is a reversible post-translational modification through an active balance of action of protein *S*-acyltransferases (PATs) and putative protein thioesterases, and is therefore considered important in dynamic subcellular trafficking of proteins between membranes (Mitchell, Vasudevan, Linder, & Deschenes, 2006). In the past 10 years, several new proteomic methods to detect palmitoylated proteins have been deployed, primarily acyl-biotin exchange (ABE)-based methods and 'Click'-based methods. ABE is a more indirect method based on replacement palmitates on palmitoylated cysteines with a group that enables separation, such as biotin or polyethylene glycol (Forrester et al., 2011; Percher et al., 2016; Roth, Wan, Bailey, et al., 2006; Wan, Roth, Bailey, & Davis, 2007). More direct labeling methods such as 'click' chemistry involves incorporation of palmitate analogs such as either ω–alkynyl-palmitate or an azido-palmitate into cells, and these analogs can be subsequently labeled with tagged azides or triarylphosphines, respectively (Hannoush, 2012; Kostiuk et al., 2008; Martin & Cravatt, 2009; Peng & Hang, 2015; Yap et al., 2010). Several thousands of new palmitoylated substrates have been experimentally identified by deploying these methods on specific organisms, subcellular compartments, cell types, and other classifications. This increase in known substrates has led to comprehensive compendiums of palmitoylated proteins, which will surely be essential for future development of the field (Blanc et al., 2015; Sanders et al., 2015). Despite these innovations, a clearly recognizable palmitoylation 'motif' has not been identified, as is the case for other types of lipidation.

In the case of myristoylation (C14 acylation), there are two highly conserved N-myristoyl transferases

*Corresponding author. Email: rdeschen@health.usf.edu
†These authors contributed equally to this work.

(NMT1 and NMT2) with well-characterized enzymatic mechanisms based on high-resolution crystal structures, as well as clearly defined preference for the α-amino group of an N-terminal glycine (Farazi, Waksman, & Gordon, 2001; Martin, Beauchamp, & Berthiaume, 2011). Similarly, farnesylation is carried out by a single farnesyltransferase (FTase) and geranylgeranylation by two enzymes (GGTase I and II), which share a common subunit with the FTase (Casey & Seabra, 1996; Zverina, Lamphear, Wright, & Fierke, 2012). The presence of protein prenyl modifications can be easily recognized by specific features of primary sequence at the modification site, which consist of highly conserved CaaX, CC, and CXC motifs. In contrast, palmitoylation is catalyzed by two distinct families of transferases. Secreted proteins, such as the morphogens Sonic Hedgehog and Wnt, are *N*-palmitoylated and *O*-acylated, respectively. Both of these modifications are performed by MBOATs (membrane bound *O*-acyl transfersases) of which there are 11 known members, such as Hhat and Porcupine (Buglino & Resh, 2008; Gao & Hannoush, 2014). Soluble and membrane proteins are *S*-palmitoylated by zDHHC enzymes, of which there are 23 expressed in mammals (Mitchell et al., 2006).

The large number of proteins that undergo palmitoylation, along with the large family of palmitoyltransferase enzymes, may contribute to the difficulty of identifying a simple consensus sequence. It also suggests that a more sophisticated search algorithm must be developed that takes into account more than primary sequence, also including factors such as subcellular localization, membrane topology, secondary structure, and other physiochemical properties of a polypeptide chain. For example, cysteines both within and adjacent to transmembrane domains appear to be prone to palmitoylation (Roth, Wan, Bailey, et al., 2006). In addition, prior farnesylation and myristoylation promote sequential palmitoylation of proximal cysteines, such as with Ras and Gα, respectively (Aicart-Ramos, Valero, & Rodriguez-Crespo, 2011; Linder & Deschenes, 2007). Overall, it appears that while PATs may have some inherent specificity toward substrates, there may be overlap between related zDHHC PATs (Greaves & Chamberlain, 2011; Ohno et al., 2012; Rocks et al., 2010). However, attempts to determine whether individual zDHHC enzyme or subfamilies of zDHHC enzymes recognize common targets have had limited success. The development of better computational tools, along with larger data-sets of experimentally determined palmitoylated proteins should clarify whether zDHHC-dependent consensus sequences and/or recognition features exist.

In this study, we use a carefully selected and manually curated data-set of palmitoylated cysteines to characterize the physicochemical properties that define the surrounding environment of palmitoylatable residues. We find that the likelihood of a cysteine to be palmitoylated is dependent not only on its membrane association state and structural context, but also on its surrounding sequence composition. Inclusion of context-based information led to the development of a palmitoylation predictor that outperforms existing predictors on unbiased data-sets. Overall, we hypothesize that these observations will not only lead to improvements in palmitoylation prediction, but also a better understanding of the complex dynamics between PATs and their substrates.

Methods

Data-set assembly

Our data-set was compiled by retrieving palmitoylated proteins from UniProt and CSS-Palm 4.0 (Ren et al., 2008). Palmitoylated proteins from the UniProt database (release 2013_10) were composed by searching the keywords 'Sequence annotation [FT]' under 'Field,' 'Lipidation' under 'Topic,' '*S*-palmitoyl cysteine' under 'Term,' and 'Experimental' under 'Confidence.' After duplicates were removed, an 80% sequence identity cut-off was implemented using the ElimDupes web server in order to remove redundant proteins. Cysteines ± 10 amino acids were extracted, and another 90% sequence identity cut-off was implemented at the 21-mer level in order to eliminate remaining redundancy. Sites were manually divided into palmitoylated (P) and non-palmitoylated (NP) sets, based on the following requirements:

- Palmitoylation must be directly labeled, through either [^{3}H] palmitate incorporation, ABE-based methods, 'Click' chemistry, or other experimental methods

  Studies where cysteine mutation led to change in localization, gel shift other than acyl-PEGyl exchange, or other method not directly relevant to palmitoylation were not considered. While the sensitivity of direct detection of palmitoylation varies by method, we believe that these differences are likely to have minimal impact.
- Proteins must be *S*-palmitoylated, shown by either hydroxylamine cleavage or 2-bromopalmitate inhibition

  Several instances of palmitoylation occur via different mechanisms, involving enzymes other than zDHHCs and linkages other than thioesters. In this study, we were only interested in zDHHC catalyzed, thioester-linked palmitoylation. For our purposes, either hydroxylamine cleavage (demonstrates thioester linkage) and/or 2-BP inhibition (demonstrates probable zDHHC catalysis) was considered sufficient.
- Specific palmitoylated sites must be validated experimentally

Instances where multiple cysteines were mutated simultaneously were not included, unless the sites were within three residues of one another.

- Palmitoylation must not be detectable when all known palmitoylated cysteines are mutated

  If mutation of a subset of cysteines on a protein of interest still had detectable palmitoylation levels, it is possible that other cysteines on the protein may also be palmitoylated, which would therefore bias the NP set.
- Previously modified cysteines are not included

  As *S*-palmitoylation does not occur in the lumen, all disulfide-bonded cysteines that were both annotated by UniProt and predicted by CYSPRED (Fariselli, Riccobelli, & Casadio, 1999)were removed in order to reduce bias in our negative data-set. Prenylated cysteines were also excluded from our data-set.

Although these rules may miss some palmitoylated sites due to experimental variation, they likely represent the highest possible stringency while retaining enough proteins for the analysis.

Annotation

Membrane association state

Proteins that are lipidated are more likely to undergo sequential *S*-palmitoylation. Therefore, known palmitoylation-promoting modifications such as N-myristoylated glycines (N-terminal, MGC motif) and prenylated cysteines (C-terminal, -CaaX box) were annotated. While *O*-palmitoleoylation is known to promote subsequent *S*-palmitoylation in proteins such as Wnt, these were not annotated, as secreted proteins undergo a different mechanism of palmitoylation (Resh, 2013). Additionally, palmitoylation occurs at integrally associated membrane proteins, often directly adjacent to a TMD. We applied the TMHMM 2.0 algorithm to our data-sets (Krogh, Larsson, von Heijne, & Sonnhammer, 2001), cross-checked this analysis with UniProt annotations of subcellular localization, and classified each protein as either not transmembrane (NONTMP) or transmembrane (TMP). Within the TMP set, sites were classified according to the output of TMHMM 2.0, which factors the probability that the N-terminus is on the cytoplasmic side of the membrane. Therefore, the correct number of TMDs and the orientation of the termini are predicted with reasonable accuracy, with some exceptions occurring due to factors such as signal peptides being mistaken for TMDs or reversed direction of a protein with only one TMD. Based on the TMHMM 2.0 output, the TMP set was further subdivided into: within the helix (TMHELIX), predicted to be cytosolic (TMP-IN), or predicted to be organelle luminal or outward facing (TMP-OUT).

Intrinsic disorder features

Structural characteristics such as intrinsic disorder have been implemented into palmitoylation prediction algorithms with varying success (Hu et al., 2011; Kumari, Kumar, & Kumar, 2014). However, we hypothesized that accurate intrinsic disorder predictors may help generalize aspects of structural context. Therefore, we utilized outputs from PONDR-FIT (Xue, Dunbrack, Williams, Dunker, & Uversky, 2010), a highly accurate metapredictor which incorporates the outputs of FoldIndex (Prilusky et al., 2005), TopIDP (Campen et al., 2008), PONDR® VLXT (Romero et al., 2001), PONDR® VSL2 (Peng, Radivojac, Vucetic, Dunker, & Obradovic, 2006), PONDR® VL3 (Peng et al., 2005), and IUPred (Dosztanyi, Csizmok, Tompa, & Simon, 2005).

All cysteines were broadly classified based on their intrinsic disorder status (Figure 1). For a given protein, intrinsically disordered (>.5) residues from the first instance until the first instance of an ordered residue were considered to be the N-terminal tail (N-DIS). The reverse principle was used for the C-terminal tail (C-DIS); that is, intrinsically disordered residues from the last residue toward the amino terminus, until the first instance of an ordered residue. Cysteines within the remaining internal residues were further classified based on predicted intrinsic disorder, where sites with a disorder score of greater than or equal to .5 were considered 'disordered middle' (M-DIS), whereas sites less than .5 were considered 'ordered' (ORD).

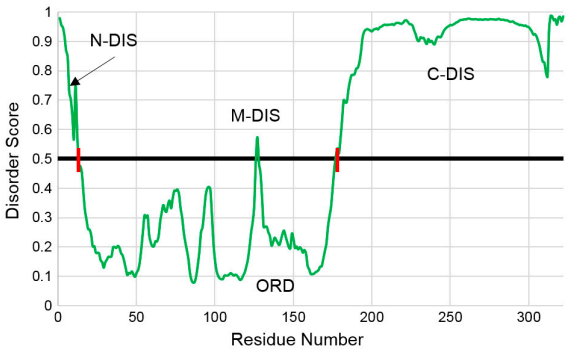

Figure 1. Classification of predicted termini based on disorder score.

Notes: In this model PONDR-FIT plot (Ras2, UniProt ID: P01120), all sites until the first red tick mark were considered N-DIS, all sites until the next red tick mark were considered ORD or M-DIS (if the site had a disorder score less than or greater than .5, respectively), and all sites after were considered C-DIS.

Disulfide bond filtering

Cysteines identified as disulfide-bonded by both UniProt annotations and CYSPRED prediction (Fariselli et al., 1999) were removed from the data-set. This filter removed 0 cysteines from the P set and 592 cysteines from the NP set, which was expected considering our initial assumptions.

Analysis

WebLogo

The WebLogo software is able to generate graphical representations of consensus sequences (Crooks, Hon, Chandonia, & Brenner, 2004; Schneider & Stephens, 1990). The default values were generally used, with minimal exceptions. The default color scheme based on chemical properties was used, where small/polar residues (G, S, T, Y, C) are green, neutral residues (Q, N) are purple, basic residues (K, R, H) are blue, acidic residues (D, E) are red, and hydrophobic residues (A, V, L, I, P, W, F, M) are black.

Two-sample logos

The two-sample WebLogo (2SWL) was utilized in order to display the differences between P and NP sets (Vacic, Iakoucheva, & Radivojac, 2006). In this tool, residues are separated into two groups: (a) enriched in the P set, or (b) depleted in the P set. The color scheme used was the WebLogo default mentioned previously. Statistically significant residues are displayed proportionally to the difference between the two sets. The *p*-value was calculated using a *t*-test, and *p*-values calculated via binomial distribution did not yield significantly different results (data not shown).

Composition profiler

In order to determine the differences of physicochemical properties between the P and NP sets, the Composition Profiler tool was used (Vacic, Uversky, Dunker, & Lonardi, 2007). Like the two-sample logo, this tool displays differences between two data-sets, except using specific amino acid physicochemical properties such as charge, hydrophobicity, size, and flexibility.

Model training and feature selection

For design of a predictive algorithm, the open-source WEKA tool was used (Frank, Hall, Trigg, Holmes, & Witten, 2004). Default values were generally used unless otherwise specified. The 'SimpleLogistic' (linear regression) classifier was used, although different classifiers did not yield significantly lower values.

Generally, the output of a predictor can have four potential output values: known palmitoylated cysteine is predicted to be palmitoylated (true positive, TP), known not-palmitoylated cysteine is not predicted to be palmitoylated (true negative, TN), known palmitoylated cysteine is not predicted to be palmitoylated (false negative, FN), and known not-palmitoylated cysteine is predicted to be palmitoylated (false positive, FP). These values can be interpreted in several different ways: Sensitivity is the true positive rate (TP/[TP + FN]), specificity is the true negative rate (TN/[FP + TN]), and accuracy is the combination of sensitivity and sensitivity ([TP + TN]/[TP + TN + FP + FN]). The Matthews Correlation Coefficient (MCC) is generally regarded as the most balanced measure of predictive capability, and can be represented by the equation below. The resulting value is from −1 to + 1, where −1 is completely imperfect prediction, 0 is completely random prediction, and 1 is completely perfect prediction.

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN} - \mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

We began with 356 features in our data-set, which included per-residue and cumulative scores of cysteines and surrounding residues (Atchley, Zhao, Fernandes, & Druke, 2005; Petersen, Petersen, Andersen, Nielsen, & Lundegaard, 2009; Romero et al., 1998; Xue et al., 2010), classification tags annotating structural context (Krogh et al., 2001; Xue et al., 2010), binary myristoylation/prenylation annotation, evolutionary features (Altschul et al., 1997), and a previous palmitoylation predictor (Pejaver et al., 2014). The entire list is available upon request. We performed feature selection using the in-built Weka attribute evaluator 'CfsSubsetEval' and search method 'Exhaustive.'

Results

The final data-set (Supplementary Table S1) consists of 244 proteins containing 473 P sites and 1973 NP sites. These sites were further subdivided into transmembrane (TMP) versus non-transmembrane (Non-TMP) based on the annotated membrane association state of the protein. Finally, the sequences of the TMP were further subdivided based on TMHMM 2.0 predictions, where cysteines were predicted to be either within a transmembrane helix (TMhelix), predicted to be luminal facing (TMP-OUT), or predicted to be cytosolic facing (TMP-IN). The distribution of the data-set across organisms is skewed toward vertebrates, with 192/244 of the sequences falling under this category. The final composition of the data-set after assembly and annotation is graphically represented in Figure 2. Examination of our entire data-set using a WebLogo suggests that there is no inherent sequence bias

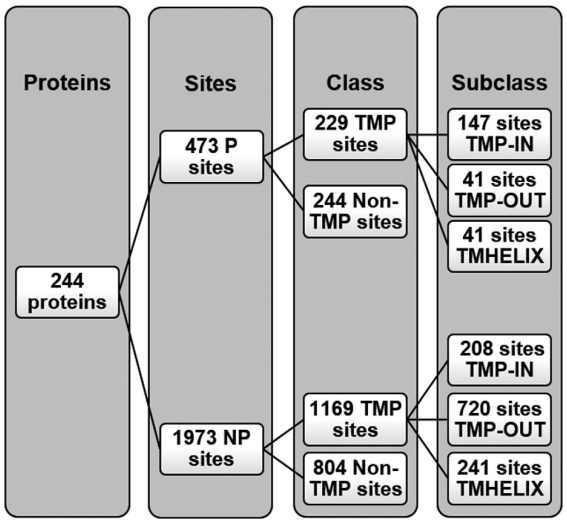

Figure 2. Characteristics of the entire data-set. (A) Flow chart of transmembrane filtering of palmitoylated substrates according to likely palmitoylation state (P, NP), membrane association state (TMP, NON-TMP), and predicted transmembrane helix orientation (TMhelix, TMP-OUT, TMP-IN).

around cysteines independent of palmitoylation (Figure 3(A))

Sequence composition directly surrounding cysteines influences likelihood of palmitoylation

Analysis of our entire data-set using a two-sample WebLogo (2SWL) revealed an apparent enrichment of hydrophobic residues on the amino side, and basic residues on the carboxyl side of palmitoylated cysteines (Figure 3(B)). Further separation of our entire data-set based on transmembrane association state reveals that this physicochemical contribution appears to be specific to transmembrane proteins (Figure 3(C)) but not cytosolic proteins (Figure 3(D)), which is in agreement with a previous study (Blanc et al., 2015).

Cysteine position relative to a transmembrane helix determines physicochemical properties of regions surrounding palmitoylatable sites

To investigate more closely the apparent 'sidedness' of hydrophobic and basic residues, we further parsed the TMP set using a combination of UniProt annotation and the TMHMM 2.0 algorithm (described further in Methods), which predicts transmembrane domains as well as orientation of a protein relative to the lumen and cytosol. Cysteines located within a predicted transmembrane helix (TMhelix) displayed a greater ratio of carboxyl-terminal basic residues compared to amino-terminal hydrophobic residues (Figure 4(A)), with this enrichment being significantly greater than that of the TMP set.

The predicted cytosolic (TMP-IN) set was used as a comparison to the TMhelix set, as the ratio of P/NP sites were more significant in the TMP-IN set compared to the predicted luminal (TMP-OUT) set (.71–.06, respectively). This ~12.4-fold increase in ratio is likely reflective of the relevance of biological context, since cytosolic proteins are palmitoylated whereas luminal proteins are not. Within the TMP-IN set, there is a clear increase in enrichment of hydrophobic residues on the amino-terminal side of the site, whereas the predominance of basic residues on the carboxyl-terminal side not only appears to be greatly depleted, but also dampened by the presence of residues with different physicochemical properties (Figure 4(B)). In other words, cysteines prone to palmitoylation within a transmembrane protein will either have surrounding basic or hydrophobic residues, and this is dependent on whether or not the cysteine is within a transmembrane helix.

It should be noted that in these data-sets, there are instances that have surrounding sequence patterns that are reverse of what would be predicted by our results (i.e. amino-terminal basic residues and carboxyl-terminal hydrophobic residues). This is readily explained by taking into consideration that cytosolic-facing, juxtamembrane cysteines can occur before or after the TMD in the primary sequence, depending on the transmembrane topology of the protein.

Palmitoylated cysteines associated with myristoyl/prenyl modifications likely have no additional physicochemical properties that contribute to palmitoylation propensity

It is well known that peripheral lipid modifications such as myristoylation and prenylation can promote palmitoylation of nearby cysteines. However, it is unclear if surrounding sequence-based characteristics influence likelihood of palmitoylation in the presence of other lipidations. The N-DIS and C-DIS displayed minimal enrichment with regard to local amino acid composition, besides the presence of myristoylation and prenylation motifs (MGC- and -CaaX, respectively) which are known to promote palmitoylation (data not shown).

These sets were further divided into myristoylated and prenylated P sets, as no NP cysteines were found at the predicted N-terminus of myristoylated or prenylated proteins. In these cases, sequences were aligned to the myristoyl or prenyl motifs, rather than a putative palmitoylated cysteine, in order to determine if there was a relationship between myristoylation/prenylation and putative palmitoylated cysteines. In these sets, there are still no significant physicochemical features displayed. Additionally, the palmitoylated cysteine appears to be more conserved when located closer to the myristoylated/prenylated residue (Figure 5(A) and (B), respectively), indicating there

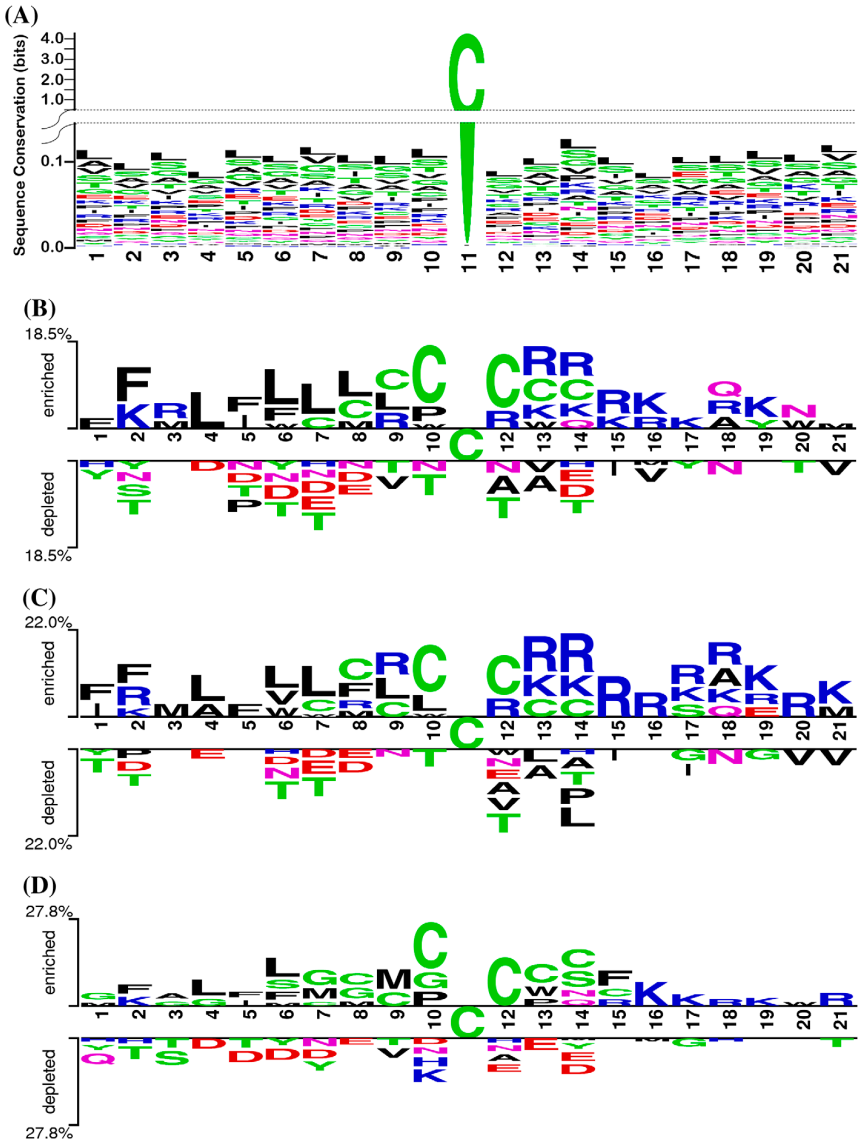

Figure 3. Properties of residues surrounding palmitoylated sites. (A) represents the distribution of the entire data-set using a WebLogo. (B) 2SWL analysis of the entire data-set, which containes 473 P sites and 1973 NP sites. (C) 2SWL analysis of the TMP data-set, which consist of 1169 NP sites and 229 P sites. (D) 2SWL analysis of the NON-TMP data-set, which consists of 804 NP sites and 244 P sites.

may be a correlation between palmitoylation likelihood and distance to the myristoyl or prenyl modification.

Improvement of predictive accuracy by incorporation of physicochemical factors, contextual classification, evolution, and data-set annotation

As classification of sites into categories revealed trends in physicochemical properties of surrounding sequence, we hypothesized that these annotations would significantly improve accuracy of palmitoylation algorithms. Using our manually curated data-set, classifications of membrane association and disorder propensity, values for secondary structure/electrostatic charge/average surface area (Atchley et al., 2005; Petersen et al., 2009), ModPred palmitoylation output values (Pejaver et al., 2014), and position-specific scoring matrix values (Altschul et al., 1997), we created a predictive model based on a simple logistic regression learning algorithm. Based on the 22 selected features (Table 1), our predictor accounts for

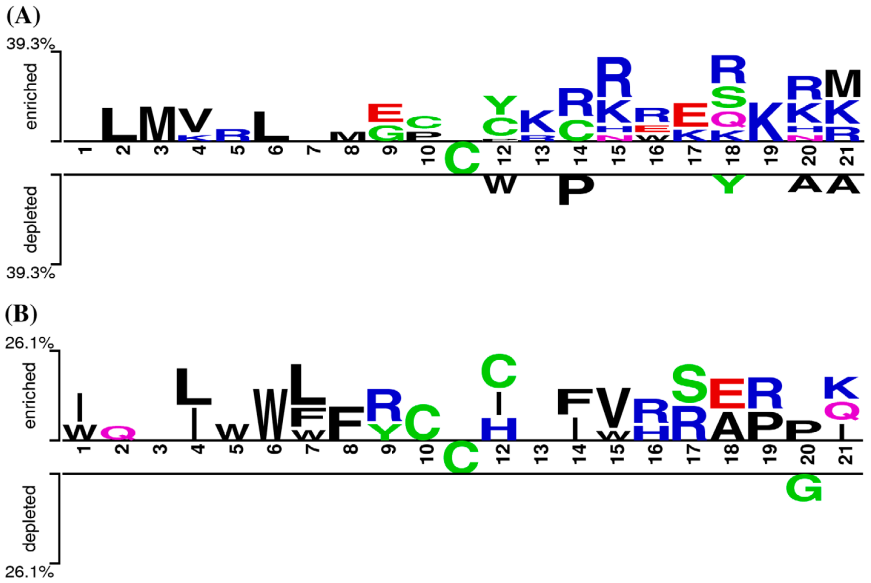

Figure 4. Enrichment of sequences surrounding cysteines is dependent on proximity to a transmembrane domain (A) 2SWL analysis of the TMhelix data-set, which consists of 241 NP sites and 41 P sites. (B) 2SWL analysis of the TMPIN sets, which consists of 208 NP sites and 147 P sites.

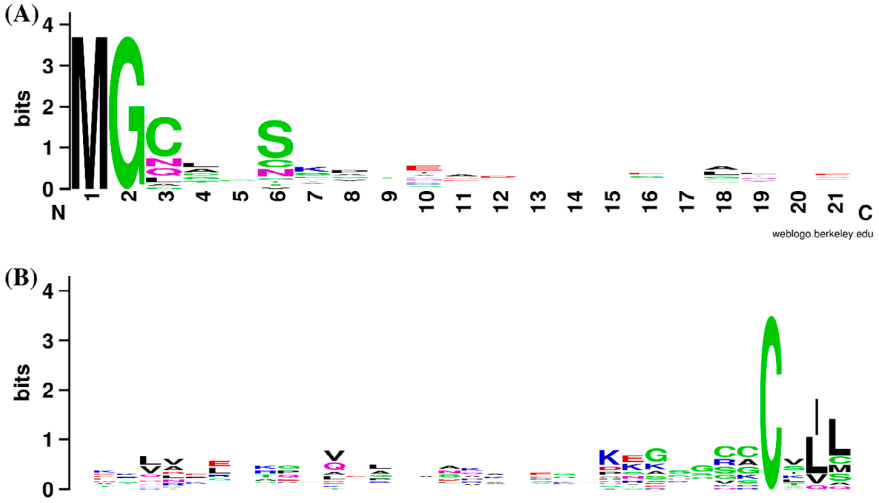

Figure 5. Correlation of myristoylation and prenylation motifs to the composition of nearby palmitoylated cysteine sites. (A) WebLogo analysis of 23 myristoylated proteins that are also classified as N-DIS and palmitoylated. (B) WebLogo analysis of 17 prenylated proteins that are classified as C-DIS and also palmitoylated.

some but not all of the annotations that we hypothesized were important, most significantly sidedness of charge/hydrophobicity and the predicted transmembrane position (TMP-OUT). The previously discussed ratio of P/NP sites between the TMP-IN and TMP-OUT sets made the presence of this factor unsurprising, indicating that simply considering the context of a cysteine in the cell may be a powerful predictive factor. A notable absence is myristoyl/prenyl motifs, which we assume to be the case because of the relatively low presence of these proteins in our training set (24 and 17 proteins, respectively). More notable absences are TMhelix and TMP-IN as predictive factors, which may be due to either redundancy in training data (i.e. secondary structure as an indicator of

Table 1. A combination of the 22 minimal physicochemical, evolutionary, and classification features of cysteines and their surrounding residues that contribute to palmitoylation prediction of our predictor described in this study.

Feature	Properties	Source	Reference
BF: 8	Secondary Structure (Factor II) of amino acid #8	N/A	Atchley et al. (2005)
EF: 6, 16	Electrostatic Charge (Factor V) amino acid #6, #16	N/A	Atchley et al. (2005)
ORD	Classification Tag	PONDR-FIT, this study	Xue et al. (2010)
C-DIS	Classification Tag	PONDR-FIT, this study	Xue et al. (2010)
TMP-OUT	Predicted Transmembrane Position	TMHMM 2.0	Krogh et al. (2001)
ASA: 7	Average surface area of amino acid #7	NetSurfP	Petersen et al. (2009)
Beta: 2, 6	Propensity for beta sheets at the #2, #6 residue	NetSurfP	Petersen et al. (2009)
Coil: 5, 15	Propensity for coil sheets at #5, #15 residue	NetSurfP	Petersen et al. (2009)
ModPred Score (PSSM)	ModPred final score for palmitoylation	ModPred	Pejaver et al. (2014)
PSSM: 16, 21, 26, 46, 56, 77, 94, 100, 140, 145	Raw PSSM scores (out of 164 evolutionary features)	PSI-BLAST	Altschul et al. (1997), Pejaver et al. (2014)

Note: For a given 21-mer, the cysteine is constantly at position #11, the first residue is position #1, and the last residue is position #21.

TMDs), or simply weak correlations that require refinement. Future predictors may consider creating separate models for proteins predicted to undergo additional lipid modifications, integral transmembrane proteins, and peripherally associated membrane proteins.

To determine the overall predictive ability of our algorithm, we used the Matthews Correlation Coefficient (MCC) value determined by 10-fold cross validation, which provides a value from −1 (completely imperfect prediction) to 1 (perfect prediction), with 0 being random prediction. In 10-fold cross-validation, the data-set is randomly separated into 10 equally sized subsamples. Nine subsamples are used for training, and 1 sample is used for testing. This process is repeated 10 times (folds), and the final output values are the average of the 10 folds. A more detailed explanation of the terms used hereafter can be found in the Methods section (Model training and feature selection). Our predictor achieved an MCC of .74 via 10-fold cross validation (Table 2), a marginally higher value than two recently published predictors, WAP-Palm and PalmPred (Kumari et al., 2014; Shi et al., 2013), which were reported to achieve .73 and .71, respectively.

However, MCC values can be easily biased by similar or redundant proteins in the data-set – in other words, the same site could potentially be found in both the training and testing sets. Therefore, a better method to compare predictive capability between algorithms is to use proteins that none of the algorithms have been trained on. We tested 13 proteins not used for the training of any published predictor, including our own (Supplementary Table S2). Our observations are limited by the size of our data-set, which is an unavoidable issue as relatively few unique palmitoylated proteins have been described that have not been used for training of any algorithm. To account for the fact that other predictors do not consider other modifications of cysteines, we did not remove disulfide-bonded and prenylated cysteines from our testing datasets. We found that our algorithm performed significantly better than published predictors, with a .43 and .34 increase in MCC over WAP-Palm and PalmPred, respectively (Table 3).

Table 2. 10-fold cross validation values of the new predictive algorithm described in this study.

10-fold cross validation	Value
Sensitivity	.732
Specificity	.967
Accuracy	.922
MCC	.739

We then directly compared our output when considering these 13 proteins to the outputs of CSS-Palm 4.0, WAP-Palm, PalmPred, and ModPred (Supplementary Table S3). In this limited data-set, our predictor outperforms others by having a similar if not better true positive (sensitivity) rate as others and a slightly higher true negative (specificity) rate while maintaining a relatively low false positive rate (Table 3). When predictions of individual sites are compared to those of the most-cited predictor in the field (CSS-Palm) and one of the more recently published predictors (PalmPred), it can be seen that the true negatives predicted are mostly common between predictors while there is variation between true positive, false positive, and false negative predictions. Again, these comparisons suggest our predictor is an

Table 3. Comparison of the new predictive algorithm to existing predictors. Predictions of disulfide-bonded and prenylated cysteines are included in these datasets.

	13 proteins				13 proteins + RyR1
Tool	Sensitivity (true positive rate)	Specificity (true negative rate)	Accuracy (Sn+Sp)	MCC	MCC
CSS-Palm 4.0	.50	.91	.84	.42	.18
WAP-Palm	.50	.65	.62	.11	.04
ModPred	.44	.83	.76	.25	.17
PalmPred	.22	.93	.81	.20	.06
This study	.44	.98	.89	.54	.34

Note: The initial 13 proteins consist of 18 P sites and 88 NP sites. RyR consists of 18 P sites and 82 NP sites.

improvement through relatively low false positive rates, slightly higher true negative rates, and similar or better true positive rates (Figure 6)

Interestingly, our results changed dramatically upon inclusion of a unique protein, the ryanodine receptor (RyR1) (Supplementary Table S4). RyR1 likely represents a challenge for present palmitoylation predictors, as it is a multipass transmembrane protein that is extremely large (~5000 amino acids) and contains many cysteines (18 P, 82 NP) that undergo palmitoylation as well as *S*-oxidation and *S*-nitrosylation due to reactive oxygen and nitrogen species (Chaube et al., 2014). Including the RyR in our testing set significantly decreased all MCC values, suggesting that more factors may have to be considered for optimal predictive accuracy (Table 3). Development of a web-based version of this predictor is currently in progress, and updates on release progress can be requested at descheneslab@gmail.com.

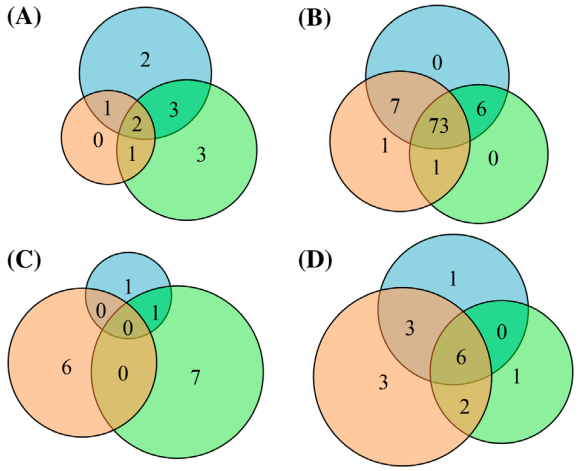

Figure 6. Comparison of predictors on the unbiased 13 protein testing set. The Venn Diagrams represent the (A) true positive, (B) true negative, (C) false positive, and (D) false negative rates of PalmPred (orange, bottom left), CSS-Palm 4.0 (green, bottom right), and the predictor described in this study (blue, top middle).
Notes: Within each Venn Diagram, each number represents the number of sites either unique to a predictor, in common with one predictor, or in common with two predictors. The area of each circle is directly proportional to the total number of sites within each circle.

Discussion

It is widely assumed that *S*-palmitoylated substrates, unlike those of many other post-translational modifications, lack a recognizable consensus motif. While studies have identified residues proximal to palmitoylated cysteines that appear critical for efficient palmitoylation (Khanal et al., 2015; Nadolski & Linder, 2009), a single consensus sequence has not emerged. This has led some to conclude that palmitoylation is promiscuous and there is no sequence preference (Rocks et al., 2010). Another possibility is that physiological context, with or without sequence specificity, dictates the likelihood of palmitoylation (Fukata & Fukata, 2010). For example, cysteines adjacent to transmembrane domains appear to increase the propensity of a cysteine being palmitoylated.

In this study, we addressed these questions using computational techniques. First, we constructed a high-quality data-set of palmitoylated proteins through an exhaustive critical analysis of the literature, similar to the publicly available database SwissPalm (Blanc et al., 2015). This data-set only consisted of unique proteins with very well-characterized palmitoylated sites, using a set of guidelines comprehensible and reasonable to the experimental biologist. Secondly, working under the hypothesis that perhaps there are differential characteristics guiding palmitoylation, we divided this data-set into different groups based on predicted and known structural and spatial context. While we assume that our data-set is appropriately distributed among the many types of palmitoylated proteins, the distribution is biased by the limits of the techniques in the field. The majority of palmitoylated sites in our data-set were identified by

^{3}[H] palmitate, which lacks the sensitivity of more recent techniques such as ABE, and therefore our data-set may be biased against transient or less favorable sites (Drisdel & Green, 2004; Martin & Cravatt, 2009; Roth, Wan, Green, Yates, & Davis, 2006). Additionally, identified sites are not certain to be physiologically relevant, as the identification techniques used throughout the literature are generally based on the non-physiological conditions of enzyme–substrate overexpression analysis (Fukata, Iwanaga, & Fukata, 2006). Nevertheless, these exceptions likely represent a minor fraction compared to the entire data-set.

Palmitoylated cysteines in transmembrane proteins appear to be enriched in both amino-terminal hydrophobic and carboxyl-terminal basic residues. We will discuss our results in the context of cytosolic cysteines carboxyl-terminal to a transmembrane domain, since our data-set appears to be skewed toward those instances. Our results suggest that the surrounding sequence of a palmitoylated cysteine found in a predicted transmembrane helix tends to be enriched in primarily basic residues, whereas palmitoylated cysteines outside of TMDs are enriched in primarily hydrophobic residues. The most likely reason for this difference may be because palmitoylated cysteines predicted to be within the transmembrane helix are specifically juxtamembrane, and the adjacent residues promote palmitoylation through interaction with the zDHHC PAT, membrane interactions, or some other mechanism. This mechanism has been seen in reovirus p10, which requires basic residues adjacent to the palmitoylated cysteine for p10-mediated membrane fusion (Shmulevitz, Salsman, & Duncan, 2003). Alternatively, basic residues could alter the accessibility of the cysteine to the zDHHC PAT catalytic cysteine-rich domain (CRD). Since polar residues within transmembrane helices contribute to charge–charge repulsions or interactions (Eilers, Shekar, Shieh, Smith, & Fleming, 2000), loose helix packing of multipass transmembrane proteins through repulsions of positively charged residues may increase accessibility of the cysteine to the CRD, which may be partially embedded in the membrane bilayer to prevent hydrolysis of the autopalmitoylation state by water (Mitchell, Mitchell, Ling, Budde, & Deschenes, 2010; Mitchell et al., 2012). Conversely, cysteines outside of transmembrane domains could be enriched in hydrophobic residues because they are either near a TMD, or require helix-promoting residues for either recognition or membrane-binding.

In the case of cysteines proximal to myristoylated and prenylated sites, no amino acid preference emerged in our results. The only noticeable information that could be extracted was a greater conservation of cysteines closer to the myristoyl/prenyl motif of the palmitoylated protein. Interestingly, no NP cysteines were found that were predicted to be at the disordered N-terminus of a myristoylated protein, or that were at the disordered C-terminus of a prenylated protein. Taken together, we hypothesize that the presence of a myristoyl or prenyl motif is sufficient to promote palmitoylation of a nearby cysteine. Consistent with this hypothesis, it has been previously observed that alterations of residues flanking the palmitoylated cysteines of H-Ras have no effect on the palmitoylation state (Willumsen, Cox, Solski, Der, & Buss, 1996). However, these mutants did not correctly localize to the plasma membrane, indicating there may be some targeting function of these residues independent of palmitoylation. Also in H-Ras, a mutant H-Ras which is myristoylated but not farnesylated is still a substrate for palmitoylation, and properly localizes to the plasma membrane (Cadwallader, Paterson, Macdonald, & Hancock, 1994), perhaps hinting that the exact chemical composition of the prior lipid modification is not especially important.

While cytosolic palmitoylated proteins were not a primary focus of this study, an interesting future direction of this work will be determining physicochemical parameters of these proteins. Our studies in peripherally bound membrane proteins suggest that properties promoting membrane association are critical in determining palmitoylation propensity. One possibility is that cytosolic proteins may be weakly interacting with membranes, through transient secondary structure and/or electrostatic interactions, a phenomenon seen with intrinsically disordered proteins such as α-synuclein (Jo, McLaurin, Yip, St George-Hyslop, & Fraser, 2000). This hypothesis is supported by a number of studies in the Ras family of GTPases. In the case of *Saccharomyces cerevisiae* Ras2, methods of membrane association appear interchangeable – replacement of the C-terminal Ras2 farnesylation motif (-CIIS) with a nonprenylated mutant rich in hydrophobic and basic residues (-CIIKLIKRK) is able to complement the loss of Ras2 (Mitchell, Farh, Marshall, & Deschenes, 1994), and a similar result is observed in mammalian H-Ras (Booden et al., 1999).

Through our studies and various observations in the literature, we propose a model where in order for most proteins to be palmitoylated, it requires an initial membrane association in order to increase the likelihood of interacting with an active zDHHC enzyme (Figure 7). This would suggest that most, if not all palmitoylated proteins are either juxtaposed to or within transmembrane domains, located near previously lipidated residues, or are surrounded by hydrophobic and/or basic residues which would likely have affinities toward phospholipid membranes. After this initial membrane association, there are multiple possibilities – either the protein is palmitoylated stochastically by the closest available enzyme, the enzyme specifically palmitoylates the protein through a substrate recognition mechanism, or a balance between both. Previous research supports all of these possibilities. Redundancy of PATs has been

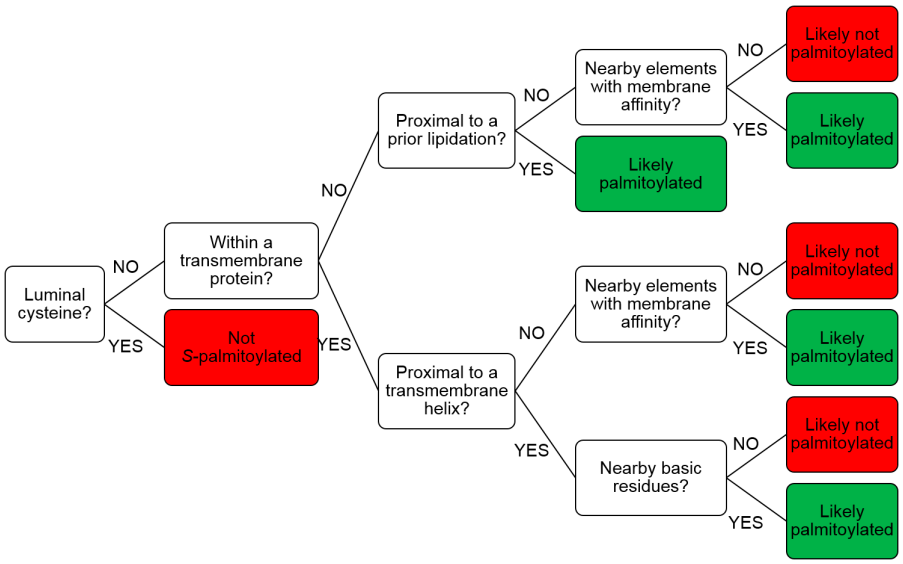

Figure 7. A decision tree model of factors that influence *S*-palmitoylation propensity of non-luminal cysteines in eukaryotes. Note: Likelihood of other cysteine modifications and enzyme regulation factors are not considered in this model.

documented – knockdown of zDHHC9, a PAT specific for H-Ras and N-Ras when using recombinant protein (Swarthout et al., 2005), did not significantly affect H-Ras membrane association kinetics in HeLa cells (Rocks et al., 2010), supporting a model of stochastic palmitoylation where any PAT can palmitoylate any substrate. Conversely, substrate recognition sequences have been found on a few PATs, such as the ankyrin repeats of zDHHC13 and zDHHC17 (Lemonidis, Gorleku, Sanchez-Perez, Grefen, & Chamberlain, 2014; Lemonidis, Sanchez-Perez, & Chamberlain, 2015), or the PDZ-binding motif of zDHHC8 (Thomas, Hayashi, Chiu, Chen, & Huganir, 2012; Thomas, Hayashi, Huganir, & Linden, 2013). However, it remains to be seen if these are a general phenomenon of zDHHC PATs, or simply outliers which recognize sequences distant to residues surrounding palmitoylated cysteines.

We have applied our findings toward the improvement of *S*-palmitoylation predictive algorithms. Several algorithms already exist and have increased in predictive accuracy over time (Kumari et al., 2014; Li et al., 2015; Pejaver et al., 2014; Ren et al., 2008), but most do not account for the structural and spatial parameters we have proposed. In general, our predictor appears to predict palmitoylated sites with a similar or better rate than other predictors, while avoiding overprediction of palmitoylation, shown by the lower false positive rate. Therefore, overall predictive capability toward unique palmitoylated proteins appears to increase when a combination of structural and spatial contexts, sequence enrichments, and manual curation is taken into account. Which combination of these factors is most important is unclear at this time; however, we expect that future iterations of palmitoylation predictors will continue to be improved by including only experimentally validated and curated data-sets such as SwissPalm (Blanc et al., 2015). Additionally, the extremely poor prediction of the unique RyR1 protein suggests that new factors may also need to be considered, such as propensity of *S*-nitrosylation and *S*-oxidation due to proximity of the mitochondria. These technical and contextual considerations, combined with continued investigation of the physicochemical properties of palmitoylated cysteines and zDHHC PAT regulation, should lead to better understanding of palmitoylation.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2016.1217275.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Aicart-Ramos, C., Valero, R. A., & Rodriguez-Crespo, I. (2011). Protein palmitoylation and subcellular trafficking. *Biochimica et Biophysica Acta (BBA) – Biomembranes, 1808*, 2981–2994. doi:10.1016/j.bbamem.2011.07.009

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research, 25*, 3389–3402.

Atchley, W. R., Zhao, J., Fernandes, A. D., & Druke, T. (2005). Solving the protein sequence metric problem. *Proceedings of the National Academy of Sciences, 102*, 6395–6400. doi:10.1073/pnas.0408677102

Blanc, M., David, F., Abrami, L., Migliozzi, D., Armand, F., Burgi, J., & van der Goot, F. G. (2015). SwissPalm: Protein palmitoylation database. *F1000Res, 4*, 261. doi:10.12688/f1000research.6464.1

Booden, M. A., Baker, T. L., Solski, P. A., Der, C. J., Punke, S. G., & Buss, J. E. (1999). A non-farnesylated Ha-Ras protein can be palmitoylated and trigger potent differentiation and transformation. *Journal of Biological Chemistry, 274*, 1423–1431.

Buglino, J. A., & Resh, M. D. (2008). Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *Journal of Biological Chemistry, 283*, 22076–22088. doi:10.1074/jbc.M803901200

Cadwallader, K. A., Paterson, H., Macdonald, S. G., & Hancock, J. F. (1994). N-terminally myristoylated Ras proteins require palmitoylation or a polybasic domain for plasma membrane localization. *Molecular and Cellular Biology, 14*, 4722–4730.

Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N., & Dunker, A. K. (2008). TOP-IDP-scale: A new amino acid scale measuring propensity for intrinsic disorder. *Protein & Peptide Letters, 15*, 956–963.

Casey, P. J., & Seabra, M. C. (1996). Protein prenyltransferases. *Journal of Biological Chemistry, 271*, 5289–5292.

Chaube, R., Hess, D. T., Wang, Y. J., Plummer, B., Sun, Q. A., Laurita, K., & Stamler, J. S. (2014). Regulation of the skeletal muscle ryanodine receptor/Ca2+-release channel RyR1 by *S*-palmitoylation. *Journal of Biological Chemistry, 289*, 8612–8619. doi:10.1074/jbc.M114.548925

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research, 14*, 1188–1190. doi:10.1101/gr.849004

Dosztanyi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005). IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics, 21*, 3433–3434. doi:10.1093/bioinformatics/bti541

Drisdel, R. C., & Green, W. N. (2004). Labeling and quantifying sites of protein palmitoylation. *BioTechniques, 36*, 276–285.

Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O., & Fleming, P. J. (2000). Internal packing of helical membrane proteins. *Proceedings of the National Academy of Sciences, 97*, 5796–5801.

Farazi, T. A., Waksman, G., & Gordon, J. I. (2001). The biology and enzymology of protein N-myristoylation. *Journal of Biological Chemistry, 276*, 39501–39504. doi:10.1074/jbc.R100042200

Fariselli, P., Riccobelli, P., & Casadio, R. (1999). Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. *Proteins: Structure, Function, and Genetics, 36*, 340–346.

Forrester, M. T., Hess, D. T., Thompson, J. W., Hultman, R., Moseley, M. A., Stamler, J. S., & Casey, P. J. (2011). Site-specific analysis of protein S-acylation by resin-assisted capture. *The Journal of Lipid Research, 52*, 393–398. doi:10.1194/jlr.D011106

Frank, E., Hall, M., Trigg, L., Holmes, G., & Witten, I. H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics, 20*, 2479–2481. doi:10.1093/bioinformatics/bth261

Fukata, Y., & Fukata, M. (2010). Protein palmitoylation in neuronal development and synaptic plasticity. *Nature Reviews Neuroscience, 11*, 161–175. doi:10.1038/nrn2788

Fukata, Y., Iwanaga, T., & Fukata, M. (2006). Systematic screening for palmitoyl transferase activity of the DHHC protein family in mammalian cells. *Methods, 40*, 177–182. doi:10.1016/j.ymeth.2006.05.015

Gao, X., & Hannoush, R. N. (2014). Single-cell imaging of Wnt palmitoylation by the acyltransferase porcupine. *Nat Chem Biol, 10*, 61–68. doi:10.1038/nchembio.1392

Greaves, J., & Chamberlain, L. H. (2011). DHHC palmitoyl transferases: Substrate interactions and (patho)physiology. *Trends in Biochemical Sciences, 36*, 245–253. doi:10.1016/j.tibs.2011.01.003

Hannoush, R. N. (2012). Profiling cellular myristoylation and palmitoylation using omega-alkynyl fatty acids. *Methods in Molecular Biology, 800*, 85–94. doi:10.1007/978-1-61779-349-3_7

Hu, L. L., Wan, S. B., Niu, S., Shi, X. H., Li, H. P., Cai, Y. D., & Chou, K. C. (2011). Prediction and analysis of protein palmitoylation sites. *Biochimie, 93*, 489–496. doi:10.1016/j.biochi.2010.10.022

Jo, E., McLaurin, J., Yip, C. M., St George-Hyslop, P., & Fraser, P. E. (2000). alpha-Synuclein membrane interactions and lipid specificity. *Journal of Biological Chemistry, 275*, 34328–34334. doi:10.1074/jbc.M004345200

Khanal, N., Pejaver, V., Li, Z., Radivojac, P., Clemmer, D. E., & Mukhopadhyay, S. (2015). Position of proline mediates the reactivity of *S*-palmitoylation. *ACS Chem Biol, 10*, 2529–2536. doi:10.1021/acschembio.5b00429

Kostiuk, M. A., Corvi, M. M., Keller, B. O., Plummer, G., Prescher, J. A., Hangauer, M. J., … Berthiaume, L. G. (2008). Identification of palmitoylated mitochondrial proteins using a bio-orthogonal azido-palmitate analogue. *FASEB J, 22*, 721–732. doi:fj.07-9199com[pii]10.1096/fj.07-9199com

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology, 305*, 567–580. doi:10.1006/jmbi.2000.4315

Kumari, B., Kumar, R., & Kumar, M. (2014). PalmPred: An SVM based palmitoylation prediction method using sequence profile information. *PLoS One, 9*, e89246. doi:10.1371/journal.pone.0089246

Lemonidis, K., Gorleku, O. A., Sanchez-Perez, M. C., Grefen, C., & Chamberlain, L. H. (2014). The Golgi S-acylation machinery comprises zDHHC enzymes with major differences in substrate affinity and S-acylation activity. *Molecular Biology of the Cell, 25*, 3870–3883. doi:10.1091/mbc.E14-06-1169

Lemonidis, K., Sanchez-Perez, M. C., & Chamberlain, L. H. (2015). Identification of a novel sequence motif recognized by the ankyrin repeat domain of zDHHC17/13 S-acyltransferases. *Journal of Biological Chemistry, 290*, 21939–21950. doi:10.1074/jbc.M115.657668

Li, S., Li, J., Ning, L., Wang, S., Niu, Y., Jin, N., … Xi, L. (2015). In silico identification of protein *S*-palmitoylation sites and their involvement in human inherited disease. *Journal of Chemical Information and Modeling, 55*, 2015–2025. doi:10.1021/acs.jcim.5b00276

Linder, M. E., & Deschenes, R. J. (2007). Palmitoylation: Policing protein stability and traffic. *Nature Reviews Molecular Cell Biology, 8*, 74–84. doi:10.1038/nrm2084

Martin, D. D., Beauchamp, E., & Berthiaume, L. G. (2011). Post-translational myristoylation: Fat matters in cellular life

and death. *Biochimie, 93*, 18–31. doi:10.1016/j.biochi.2010.10.018

Martin, B. R., & Cravatt, B. F. (2009). Large-scale profiling of protein palmitoylation in mammalian cells. *Nature Methods, 6*, 135–138. doi:10.1038/nmeth.1293

Mitchell, D. A., Farh, L., Marshall, T. K., & Deschenes, R. J. (1994). A polybasic domain allows nonprenylated Ras proteins to function in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry, 269*, 21540–21546.

Mitchell, D. A., Hamel, L. D., Ishizuka, K., Mitchell, G., Schaefer, L. M., & Deschenes, R. J. (2012). The Erf4 subunit of the yeast Ras palmitoyl acyltransferase is required for stability of the Acyl-Erf2 intermediate and palmitoyl transfer to a Ras2 substrate. *Journal of Biological Chemistry, 287*, 34337–34348. doi:10.1074/jbc.M112.379297

Mitchell, D. A., Mitchell, G., Ling, Y., Budde, C., & Deschenes, R. J. (2010). Mutational analysis of *Saccharomyces cerevisiae* Erf2 reveals a two-step reaction mechanism for protein palmitoylation by DHHC enzymes. *Journal of Biological Chemistry, 285*, 38104–38114. doi:10.1074/jbc.M110.169102

Mitchell, D. A., Vasudevan, A., Linder, M. E., & Deschenes, R. J. (2006). Protein palmitoylation by a family of DHHC protein S-acyltransferases. *The Journal of Lipid Research, 47*, 1118–1127. doi:10.1194/jlr.R600007-JLR200

Nadolski, M. J., & Linder, M. E. (2009). Molecular recognition of the palmitoylation substrate Vac8 by its palmitoyltransferase Pfa3. *Journal of Biological Chemistry, 284*, 17720–17730. doi:10.1074/jbc.M109.005447

Ohno, Y., Kashio, A., Ogata, R., Ishitomi, A., Yamazaki, Y., & Kihara, A. (2012). Analysis of substrate specificity of human DHHC protein acyltransferases using a yeast expression system. *Molecular Biology of the Cell, 23*, 4543–4551. doi:10.1091/mbc.E12-05-0336

Pejaver, V., Hsu, W. L., Xin, F., Dunker, A. K., Uversky, V. N., & Radivojac, P. (2014). The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci, 23*, 1077–1093. doi:10.1002/pro.2494

Peng, T., & Hang, H. C. (2015). Bifunctional fatty acid chemical reporter for analyzing S-palmitoylated membrane protein-protein interactions in mammalian cells. *Journal of the American Chemical Society, 137*, 556–559. doi:10.1021/ja502109n

Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., & Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics, 7*, 208. doi:10.1186/1471-2105-7-208

Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., Dunker, A. K., & Obradovic, Z. (2005). Optimizing long intrinsic disorder predictors with protein evolutionary information. *Journal of Bioinformatics and Computational Biology, 3*, 35–60.

Percher, A., Ramakrishnan, S., Thinon, E., Yuan, X., Yount, J. S., & Hang, H. C. (2016). Mass-tag labeling reveals site-specific and endogenous levels of protein S-fatty acylation. *Proceedings of the National Academy of Sciences, 113*, 4302–4307. doi:10.1073/pnas.1602244113

Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., & Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology, 9*, 51. doi:10.1186/1472-6807-9-51

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., … Sussman, J. L. (2005). FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics, 21*, 3435–3438. doi:10.1093/bioinformatics/bti537

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., & Yao, X. (2008). CSS-Palm 2.0: An updated software for palmitoylation sites prediction. *Protein Engineering Design and Selection, 21*, 639–644. doi:10.1093/protein/gzn039

Resh, M. D. (2013). Covalent lipid modifications of proteins. *Current Biology, 23*, R431–R435. doi:10.1016/j.cub.2013.04.024

Rocks, O., Gerauer, M., Vartak, N., Koch, S., Huang, Z. P., Pechlivanis, M., … Bastiaens, P. I. (2010). The palmitoylation machinery is a spatially organizing system for peripheral membrane proteins. *Cell, 141*, 458–471. doi:10.1016/j.cell.2010.04.007

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., & Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing, 1998*. 437–448.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Structure, Function, and Genetics, 42*, 38–48.

Roth, A. F., Wan, J., Bailey, A. O., Sun, B., Kuchar, J. A., Green, W. N., … Davis, N. G. (2006). Global analysis of protein palmitoylation in yeast. *Cell, 125*, 1003–1013. doi:10.1016/j.cell.2006.03.042

Roth, A. F., Wan, J., Green, W. N., Yates, J. R., & Davis, N. G. (2006). Proteomic identification of palmitoylated proteins. *Methods, 40*, 135–142. doi:10.1016/j.ymeth.2006.05.026

Sanders, S. S., Martin, D. D., Butland, S. L., Lavallee-Adam, M., Calzolari, D., Kay, C., … Hayden, M. R. (2015). Curation of the mammalian palmitoylome indicates a pivotal role for palmitoylation in diseases and disorders of the nervous system and cancers. *PLoS Computational Biology, 11*, e1004405. doi:10.1371/journal.pcbi.1004405

Schneider, T. D., & Stephens, R. M. (1990). Sequence logos: A new way to display consensus sequences. *Nucleic Acids Research, 18*, 6097–6100.

Shi, S. P., Sun, X. Y., Qiu, J. D., Suo, S. B., Chen, X., Huang, S. Y., & Liang, R. P. (2013). The prediction of palmitoylation site locations using a multiple feature extraction method. *Journal of Molecular Graphics and Modelling, 40*, 125–130. doi:10.1016/j.jmgm.2012.12.006

Shmulevitz, M., Salsman, J., & Duncan, R. (2003). Palmitoylation, membrane-proximal basic residues, and transmembrane glycine residues in the reovirus p10 protein are essential for syncytium formation. *Journal of Virology, 77*, 9769–9779.

Swarthout, J., Lobo, S., Farh, L., Croke, M., Greentree, W., Deschenes, R., & Linder, M. (2005). DHHC9 and GCP16 constitute a human protein fatty acyltransferase with specificity for H- and N-Ras. *Journal of Biological Chemistry, 280*, 31141–31148. doi:10.1074/jbc.M504113200

Thomas, G. M., Hayashi, T., Chiu, S. L., Chen, C. M., & Huganir, R. L. (2012). Palmitoylation by DHHC5/8 targets GRIP1 to dendritic endosomes to regulate AMPA-R trafficking. *Neuron, 73*, 482–496. doi:10.1016/j.neuron.2011.11.021

Thomas, G. M., Hayashi, T., Huganir, R. L., & Linden, D. J. (2013). DHHC8-dependent PICK1 palmitoylation is required for induction of cerebellar long-term synaptic depression. *Journal of Neuroscience, 33*, 15401–15407. doi:10.1523/JNEUROSCI.1283-13.2013

Vacic, V., Iakoucheva, L. M., & Radivojac, P. (2006). Two sample logo: A graphical representation of the differences between two sets of sequence alignments. *Bioinformatics, 22*, 1536–1537. doi:10.1093/bioinformatics/btl151

Vacic, V., Uversky, V. N., Dunker, A. K., & Lonardi, S. (2007). Composition profiler: A tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics, 8*, 211. doi:10.1186/1471-2105-8-211

Wan, J., Roth, A. F., Bailey, A. O., & Davis, N. G. (2007). Palmitoylated proteins: Purification and identification. *Nature Protocols, 2*, 1573–1584. doi:10.1038/nprot.2007.225

Willumsen, B. M., Cox, A. D., Solski, P. A., Der, C. J., & Buss, J. E. (1996). Novel determinants of H-Ras plasma membrane localization and transformation. *Oncogene, 13*, 1901–1909.

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., & Uversky, V. N. (2010). PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics, 1804*, 996–1010. doi:10.1016/j.bbapap.2010.01.011

Yap, M. C., Kostiuk, M. A., Martin, D. D., Perinpanayagam, M. A., Hak, P. C., Siddam, A., … Berthiaume, L. G. (2010). Rapid and selective detection of fatty acylated proteins using {omega}-alkynyl-fatty acids and click chemistry. *Journal of Lipid Research, 51*, 1566–1580. doi:jlr.D002790[pii]10.1194/jlr.D002790

Zverina, E. A., Lamphear, C. L., Wright, E. N., & Fierke, C. A. (2012). Recent advances in protein prenyltransferases: Substrate identification, regulation, and disease interventions. *Current Opinion in Chemical Biology, 16*, 544–552. doi:10.1016/j.cbpa.2012.10.015