Computational Biology and Chemistry 70 (2017) 107–115

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

An in silico analysis of primary and secondary structure specificity determinants for human peptidylarginine deiminase types 2 and 4

Justin S. Olson[a,1], Joshua M. Lubner[b,1], Dylan J. Meyer[a], Jennifer E. Grant[a,*]

[a] Department of Biology, University of Wisconsin – Stout, 410 10th Avenue E., Menomonie, WI 54751, United States
[b] Department of Physiology and Neurobiology, University of Connecticut, 75 North Eagleville Rd, Storrs, CT 06269, United States

ARTICLE INFO

Article history:
Received 23 November 2016
Received in revised form 23 March 2017
Accepted 1 August 2017
Available online 9 August 2017

Keywords:
Citrullination
Motif-x
pLogo
Glycine
Arginine
Secondary structure

ABSTRACT

Human peptidylarginine deiminases (hPADs) are a family of five calcium-dependent enzymes that facilitate citrullination, which is the post-translational modification of peptidyl arginine to peptidyl citrulline. The isozymes hPAD2 and hPAD4 have been implicated in the development and progression of several autoimmune diseases, including rheumatoid arthritis and multiple sclerosis. To better characterize the primary and secondary structure determinants of citrullination specificity, we mined the literature for protein sequences susceptible to citrullination by hPAD2 or hPAD4. First, protein secondary structure classification (α-helix, β-sheet, or coil) was predicted using the PSIPRED software. Next, we used *motif-x* and pLogo to extract and visualize statistically significant motifs within each data set. Within the data sets of peptides predicted to lie in coil regions, both hPAD2 and hPAD4 appear to favor citrullination of glycine-containing motifs, while distinct hydrophobic motifs were identified for hPAD2 citrullination sites predicted to reside within α-helical and β-sheet regions. Additionally, we identified potential substrate overlap between coil region citrullination and arginine methylation. Together, these results confirm the importance and offer some insight into the role of secondary structure elements for citrullination specificity, and provide biological context for the existing hPAD specificity and arginine post-translational modification literature.

1. Introduction

Citrullination is the conversion of peptidyl arginine to peptidyl citrulline, resulting in a loss of positive charge that disrupts the hydrogen and ionic bonding patterns that support protein structure (Vossenaar et al., 2003a). Extensive protein citrullination may lead to protein unfolding, loss of function (Tarcsa et al., 1996) and increased susceptibility to proteolytic degradation (Inagaki et al., 1989).

Citrullination of peptidyl arginine is catalyzed by the peptidyl arginine deiminase (PAD) family of enzymes. The PADs represent a class of five calcium-dependent enzymes, PAD 1-4 and PAD6, each demonstrating distinct specificities against cellular substrates (Darrah et al., 2012) as well as unique tissue-specific expression (Balandraud et al., 2005; Nachat et al., 2005; Vossenaar et al., 2003b)

PAD1 is expressed primarily in the epidermis and uterus (Guerrin et al., 2003; Nachat et al., 2005; Rus'd et al., 1999; Terakawa et al., 1991), whereas PAD3 is found primarily in hair follicles (Kanno et al., 2000; Nachat et al., 2005; Nishijyo et al., 1997; Rogers et al., 1997; Terakawa et al., 1991). PAD2 has been detected in a variety of tissues and therefore appears to be expressed ubiquitously (Vossenaar et al., 2003b), PAD4 is expressed in eosinophils, neutrophils, and granulocytes (Asaga et al., 2001; Bhattacharya and Nicholas, 2014; Nakashima et al., 2002; Nishijyo et al., 1997; Takahara et al., 1989; Vossenaar et al., 2004), and has been observed in specific neuronal populations within the cortex and hippocampus (Acharya et al., 2012). Furthermore, PAD4 is the only isoform that is localized in the nucleus of granulocytes (Nakashima et al., 2002), where it participates in histone modifications (Mastronardi et al., 2006) believed to be involved in gene regulation (Denman, 2005; Mastronardi et al., 2006; Wang et al., 2004). The role of citrullination in modulating other post-translational arginine modifications (particularly arginine methylation) remains contentious, with conflicting evidence indicating that methylarginine can be "reversed" by citrullination (Wang et al., 2004), that methylarginine prevents citrullination (Kearney et al., 2005), or that

* Corresponding author.
E-mail address: grantje@my.uwstout.edu (J.E. Grant).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.compbiolchem.2017.08.001
1476-9271/

arginine methylation and citrullination are antagonistic modifications (Cuthbert et al., 2004; Thompson and Fast, 2006).

In addition to facilitating chromatin modifications, PAD-mediated citrullination regulates the interactions of several structural proteins. During terminal epidermal differentiation of keratinocytes, the inherent charge neutralization due to citrullination lowers the isoelectric point (reduces the positive charge) of filaggrin and keratin proteins, facilitating binding and keratin matrix organization (György et al., 2006), which provides structure to the outer skin layer. Similarly, citrullination facilitates keratin-binding to desmoplakin (György et al., 2006; Ishida-Yamamoto et al., 2000, p.) a protein involved in cell adhesion. During apoptosis, citrullination destabilizes the head domain interaction of vimentin, inducing depolymerization (Witalison et al., 2015; György et al., 2006; Inagaki et al., 1989), and promotes nucleosome collapse and nuclear lamina disorganization (Mizoguchi et al., 1998).

Aberrant protein citrullination has been observed pathologically. Citrullination of myelin basic protein (MBP) and glial fibrillary acidic protein (GFAP) has been associated with multiple sclerosis (Bradford et al., 2014; Moscarello et al., 1994; Nicholas et al., 2004). Furthermore, citrullination of vimentin, fibrinogen, and other synovial proteins has been detected in the synovial fluid of rheumatoid arthritis patients (Kearney et al., 2005; Moscarello et al., 2007; van Beers et al., 2010; Vossenaar et al., 2003a, 2003b), and indeed, the human peptidylarginine deiminases (hPADs) hPAD2 and hPAD4 have been identified as targets of therapeutic intervention (Bradford et al., 2014; Nicholas et al., 2004; Vossenaar et al., 2003a). Furthermore, the inhibition of hPAD2 and hPAD4 has been shown to reverse protein-hypercitrullination and to attenuate disease progression in mouse models of multiple sclerosis (Moscarello et al., 2007). Given the role of hPAD2 and hPAD4 in human disease, characterizing the influence of primary and secondary structure on these hPAD's arginine selectivity represents a critical step towards identifying potential targets of hPAD2 and hPAD4 in disease states, and the development of highly specific hPAD inhibitors.

Nomura et al. (Nomura, 1992) investigated citrullination by PAD from rabbit skeletal muscle using small peptides. They found that an arginine in-between two prolines (P-*R*-P) was not citrullinated, while two arginines in-between two prolines (P-*R*-*R*-P) were moderately citrullinated. Additionally, they observed that arginine at or near the NH_2-terminal end of the peptide showed a reduced susceptibility to citrullination when compared with arginine located at or near the COOH-terminal end. In some cases, arginine located at the COOH-terminal had a higher susceptibility for citrullination than arginine located in the middle of the peptide.

Identifying influences of local secondary structure on the rate of citrullination, Tarcsa et al. (Tarcsa et al., 1996) examined citrullination of two known substrates, trichohyalin and filaggrin, by rabbit PAD2. They reported that the highly α-helical protein trichohyalin was citrullinated more slowly and less completely than filaggrin, a protein with little structural order. From these findings, they proposed that the extent and rate of citrullination depends on the local secondary structure of the substrate, as well as the peptide sequence surrounding the target arginine. However, this study was limited to just two proteins, representing only a small sample of potential substrates. Furthermore, the substrate specificity of rabbit muscle PAD2 appears to deviate from human PAD2 (Moscarello et al., 2007), and thus, any cross-species comparison of substrate specificity should be conducted with prudence.

More recently, Stensland et al. (Stensland et al., 2008) reported the influence of primary sequence on citrullination of nearby arginine residues. Using matrix-assisted laser desorption-ionization time-of-flight mass spectrometery (MALDI-TOF MS), the authors measured centroid mass shifts of filaggrin- or histone H3-derived peptides to PAD4-mediated citrullination. After systematically substituting each amino acid from the −2 to +2 positions on both peptides, residues favorable and disfavorable to citrullination were determined. Glycine was found to positively influence citrullination at the +1 position for both proteins, while it was favorable at the −2 position only for histone. Notably, when the amino acid preferences for filaggrin were compared to histone, the authors found that an individual amino acid could prove favorable for one protein, but not the other.

These results suggest that citrullination preferences are not only influenced by individual primary structure elements (i.e. specific residues at specific positions), but that enzyme preference for these primary structure elements is influenced by the greater context of the peptide or protein (i.e. the surrounding secondary structure). These differences in citrullination patterns for histone and filaggrin seem to endorse the premise of the Tarcsa data, i.e. that secondary structure elements, together with the surrounding primary structure components, influence citrullination substrate discrimination.

Assohou-Luty et al. (2014) conducted a detailed investigation into the substrate specificity of hPAD2 and hPAD4. They analyzed citrullination of transfected COS-1 and HEK293 cell lysate material by hPAD2 and hPAD4, and plotted each enzyme's substrate specificity for the four positions flanking either side of the central arginine. Flanking residues were listed as 'preferred' or 'disfavored' at specific positions based on the frequency of citrullination in the PAD-transfected cell lysate. Significant trends in citrullination patterns by hPAD2 and hPAD4 emerged from this analysis. The authors observed that glycine at the +1 position and tyrosine at the +3 position promote citrullination by both isotypes, while histidine at −2 and +4 and tryptophan at +2 and +3 disfavor citrullination. Their results indicate that there are differences in specificity between hPAD2 and hPAD4, and suggest that hPAD4 has more rigid specificity requirements than hPAD2.

Arita et al. (2006) solved the structure of hPAD4 bound to three histone N-terminal peptides using X-ray crystallography. They discovered that the enzyme induces a β-turn-like bent conformation composed of five consecutive residues at the molecular surface, while the remaining five residues maintain a highly disorganized and flexible structure. Furthermore, based on binding properties, the authors suggested that residues at the −2 position should be small to avoid causing steric hindrance that would prevent the enzyme from recognizing the arginine. However, it should be noted that this study uses peptides derived from only two related proteins, and the complexes are likely uninformative regarding the influence of secondary structure elements.

Despite these efforts, questions remain regarding hPAD substrate specificity, and the role(s) of primary and secondary structure motifs. To address this knowledge gap, we consolidated multiple hPAD2/hPAD4 citrullination data sets, applied more consistent analysis parameters, and investigated both primary and secondary structure components. We utilized motif identification software to rigorously extract statistically significant motifs and visualized the primary structure specificity preferences for hPAD2 and hPAD4 within different secondary structure contexts. Our results help to reconcile the existing hPAD specificity literature, and offer biological context for citrullination in both disease states and its role in arginine post-translational modifications.

2. Materials and methods

2.1. Peptide data set assembly

A combined total of 461 (285 hPAD2 and 176 hPAD4) R-centered 11-mer peptides with known citrullination sites were curated from