RESEARCH ARTICLE

Structural and posttranslational analysis of human calcium-binding protein, spermatid-associated 1

Marcelo Marcet-Palacios1,2 | Eduardo Reyes-Serratos1 | Aron Gonshor3 | Robert Buck3,4 | Paige Lacy1 | A. D. Befus1

1Department of Medicine, Alberta Respiratory Centre, University of Alberta, Edmonton, Alberta, Canada
2Northern Alberta Institute of Technology, Biological Sciences, Edmonton, Alberta, Canada
3Fluids iQ Inc., Ottawa, Ontario, Canada
4GB Diagnostics, Kingman, Arizona

Correspondence
Marcelo Marcet-Palacios. University of Alberta, Heritage Medical Research Center, Edmonton AB T6G 2S2, Canada. Email: marcelo@ualberta.ca

Funding information
AllerGen NCE Inc, Grant/Award Number: 16B&B_MSI-C5

Abstract
Recently, we detected a novel biomarker in human saliva called calcium-binding protein, spermatid-associated 1 (CABS1). CABS1 protein had previously been described only in testis, and little was known of its characteristics other than it was considered a structurally disordered protein. Levels of human CABS1 (hCABS1) in saliva correlate with stress, whereas smaller sized forms of hCABS1 in saliva are associated with resilience to stress. Interestingly, hCABS1 also has an anti-inflammatory peptide sequence near its carboxyl terminus, similar to that of a rat prohormone, submandibular rat 1. We performed phylogenetic and sequence analysis of hCABS1. We found that from 72 CABS1 sequences currently annotated in the National Center for Biotechnology Information protein database, only 14 contain the anti-inflammatory domain “TxIFELL,” all of which are primates. We performed structural unfoldability analysis using PONDER and FoldIndex and discovered three domains that are highly disordered. Predictions of three-dimensional structure of hCABS1 using RaptorX, IonCom, and I-TASSER software agreed with these findings. Predicted neutrophil elastase cleavage density also correlated with hCABS1 regions of high structural disorder. Ligand binding prediction identified Ca2+, Mg2+, Zn2+, leucine, and thiamine pyrophosphate, a pattern observed in enzymes associated with energy metabolism and mitochondrial localization. These new observations on hCABS1 raise intriguing questions about the interconnection between the autonomic nervous system, stress, and the immune system. However, the precise molecular mechanisms involved in the complex biology of hCABS1 remain unclear. We provide a detailed in silico analysis of relevant aspects of the structure and function of hCABS1 and postulate extracellular and intracellular roles.

KEYWORDS
anti-inflammatory peptides, CABS1, disordered domains, in silico analysis, IonCom, I-TASSER, posttranslational processing, RaptorX, resilience, stress
1 | INTRODUCTION

The central nervous system (CNS) has bidirectional communications with the peripheral nervous system, which in turn has functional connections with the immune/inflammatory system. To investigate the interaction between the sympathetic nervous system and the inflammatory response, we developed an animal model that involved bilateral decentralization of the superior cervical ganglia in rats, followed by an assessment of its effects on inflammation. Among the observed outcomes was a marked reduction of pulmonary inflammation, including a reduction in alveolar macrophage and neutrophil influx into the airways. In experiments to understand the mechanisms involved following decentralization, we found that removal of submandibular glands (SMG) in male rats abolished the anti-inflammatory effects. It was postulated that the SMG released an anti-inflammatory factor(s) and that this process was controlled by cervical sympathetic nerves.

Salivary glands have both exocrine and endocrine functions, with both local and systemic effects, such as in the liver and mammary glands. To identify the anti-inflammatory factor(s) released by SMG, we obtained extracts from rat SMG and purified a seven-amino acid peptide (TDIFEGG) using reverse phase high-performance liquid chromatography that had anti-inflammatory activity. This heptapeptide was identified within the prohormone submandibular rat 1 (SMR1). SMR1 had originally been identified in rat SMG and the messenger RNA (mRNA) was localized to acinar cells. Other peptides have been identified from different regions of SMR1 prohormone, including the undecapeptide VRGPRQRHNPR, hexapeptide RQHNPR, and a pentapeptide QHNPR. The authors established that, while the undecapeptide and hexapeptide were found within the SMG, the pentapeptide was only found in the submandibular secretions from the primary excretory ducts, suggesting that some SMR1 proteolytic processing took place outside of the gland. Treatment of rats with the sympathomimetic hormone, epinephrine, induced the presence of the hexapeptide in blood, and this was blocked by ligature of SMG ducts and blood vessels, indicating that the origin of the hexapeptide was the SMG. Over the following decade, numerous reports identified several anti-inflammatory and antiallergenic roles of SMR1-derived TDIFEGG and its derivatives in mice, rats, and sheep.

SMR1 is not expressed in the human genome, but a similar sequence to the SMR1 heptapeptide, TDIFEGG, is present in human calcium-binding protein, spermadial-associate 1 (hCABS1), namely, TDIFELL. This sequence and its derivatives have anti-inflammatory activity in mouse and rat models. Interestingly, several molecular forms of hCABS1 have been found in human testes, lungs, salivary gland extracts, and saliva. To determine if hCABS1 is under neural control as SMR1 is, we investigated the relationship between levels of molecular forms of hCABS1 in saliva during acute and prolonged stress. We detected a 27 kDa form of hCABS1 in saliva that correlated with increased self-perceived stress, and smaller molecular forms (<20 kDa) associated with resilience to stress, data suggesting that hCABS1 is a biomarker of stress.

Although there has been significant progress in understanding the expression of hCABS1 and its relationship with stress, numerous unanswered questions surround the biology of this novel biomarker. This manuscript focuses on in silico analyses of hCABS1, including its highly disordered nature, location of Ca^{2+}, Mg^{2+}, Zn^{2+}, leucine, and TPP (thiamine pyrophosphate, or vitamin B1) binding domains, posttranslational modifications like phosphorylation, catalytic processing by neutrophil elastase and components of the phylogeny of the putative anti-inflammatory peptide sequence. In addition, there is an in-depth analysis of how factors that may bind hCABS1 could participate in its putative extracellular and intracellular roles, including in undiscovered enzymatic roles of hCABS1 intracellularly, perhaps linked to mitochondrial function.

2 | MATERIALS AND METHODS

The amino acid sequence of human CABS1 was retrieved from the annotation recorded in the Universal Protein Resource Knowledgebase, Q96KC9, and inputted into each of the predictive software described below.

2.1 | CABS1 structural disordered domain prediction

The FoldIndex algorithm was downloaded as a script from the developer’s website and run using a Python interface. Output was a set of unfoldability predictor values of each individual amino acid based on the equation described by Prilusky et al which aims to discriminate between folded and intrinsically unfolded proteins. We also used the Predictor of Natural Disordered Regions (PONDR) online web server. In this interface, we selected five algorithms, which generate scores aimed to elucidate disordered domains in a given amino acid sequence. These algorithms are VSL2, VL3, VL-XT, XL1-XT, and CaN-XT. Output of each algorithm was given as a set of scores of each individual amino acid. Data obtained from
FoldIndex and PONDR were used to generate graphs showing each algorithm’s predictor values/scores (y-axis) for every sequence of 10 amino acids (x-axis).

2.2 Three-dimensional structural prediction approaches

We used I-TASSER, IonCom, and RaptorX online web servers to predict the three-dimensional (3D) structures of hCABS1. I-TASSER provided additional information on putative cofactor-binding domains. IonCom provided additional information on presumed ion-binding domains. 3D files were downloaded from these servers in pdb format. We used PyMol v.1.7.6.0 to generate visual representations of the hCABS1 molecule based on the information provided by NetPhos and PeptideCutter, as well as to visualize the output from I-TASSER, IonCom, and RaptorX. All 3D models were validated with ProSA and ERRAT. SwissDock was used to validate I-TASSER ligand binding predictions for TPP.

2.3 Posttranslational modification analysis

We used NetPhos online web server to predict serine, threonine, or tyrosine phosphorylation sites in hCABS1. To predict protein kinase CK2 targets, we used Group-based Prediction System ver 5.0. To predict neutrophil elastase digestion, we utilized PeptideCutter from ExPASy.

2.4 Extracellular secretion signal peptide analysis

We used SignalP 5.0 web server to assess whether hCABS1 contained signal peptides.

2.5 Amino acid sequence conservation analysis

We obtained all 72 animal calcium-binding protein, spermatid-associated (or specific) 1 annotated sequences from the protein database of the National Center for Biotechnology Information (NCBI). We excluded sequences that were listed as “predicted” or “low quality.” All sequences were inputted in FASTA format into the EMBL-EBI Clustal Omega multiple sequence alignment online tool. After online analysis, we downloaded the alignment file (Supplemental File 2) to visualize it on Jalview v.2.11.0.

We performed a phylogenetic analysis using the online platform www.phylogeny.fr, which has a pipeline that uses the MUSCLE algorithm to align sequences, followed by the GBlocks method to eliminate poorly aligned positions and divergent regions. Refined sequences were inputted into PhyML for tree construction using a Maximum-likelihood method and finally sent to TreeDyn, a software that renders the phylogenetic tree.

We used the Protein Variability Server to calculate Shannon entropy values, a widely used measurement of randomness of a given data set that evaluates the conservation of amino acid sequences. For this analysis, we excluded one species (Tupaia chinensis) that had significant insertions in CABS1 and three species with significant deletions (Heterocephalus glaber, Dasypus novemcintus, and Fukomys damarenensis). All other species (n = 68) were aligned in Clustal Omega and a Jalview file was downloaded. Once aligned in Jalview in a way that flanking regions were shown, we selected from the most prevalent beginning of the protein (residue 20) to the most prevalent end of the protein (residue 438) and generated a FASTA file. The file was then uploaded to the online PVS server and we obtained Shannon entropy graphs.

3 RESULTS

3.1 Assessment of hCABS1 structural unfoldability

Early postulates on hCABS1 biology have been developed from work done on other species. Rat CABS1, for example, is intrinsically disordered, explaining its hypersensitivity to proteolytic degradation. Rat and hCABS1 contain similar high proportions of acidic amino acids, also a feature of intrinsically disordered proteins. Thus, we postulated that hCABS1 is an intrinsically disordered protein. To further test this postulate, we performed structural unfoldability analysis of hCABS1 using two independent approaches (Figure 1). The predictor of naturally disordered regions (PONDR) software utilizes several neural networks that were trained by using well-documented disordered regions from several proteins. We used five of these systems independently (VLXT, XL1_XT, CAN_XT, VL3, and VSL2) to increase the rigor of our analysis. The predicted disorder probability was plotted as PONDR Score (Figure 1A). Areas where all five neural network predictions agreed on positive values are highlighted under the curves. A hCABS1 model was created to depict the relative location of disordered regions within residues 35 to 42, 210 to 247, 286 to 302, and 332 to 360 (Figure 1A).
A second method used to predict whether hCABS1 sequences are intrinsically unfolded was FoldIndex. This approach, by contrast, takes into account residue hydrophobicity and net charge of the sequences. 

Interestingly, this methodologically unrelated approach also detected four main regions with highly unfolded probabilities within residues 41 to 90, 180 to 247, 250 to 308, and 321 to 375 (Figure 1B). Regions where both approaches agreed in their predictions were residues 210 to 247, 286 to 302, and 332 to 360 (Figure 1C).

### 3.2 3-D predictions of hCABS1 structure

To validate the existence of multiple regions with highly disorder structures, we used two threading applications, RaptorX\(^1^8\) and Ion Ligand Binding Site Prediction (Ion-Com),\(^1^7\) as well as the Protein Structure and Function Predictions (I-TASSER) software.\(^1^6\) The latter combines homology modeling and de novo (ab initio) prediction strategies. We used these approaches because of their reported success, and to contrast results generated from distinct strategies. While homology modeling looks at evolutionarily related structures, threading predicts likely secondary structures, and then predicts hinge points that develop into motifs, domains, and later, tertiary and quaternary structures. De novo predictions utilize physicochemical principles only, demanding extensive computer resources.

We validated all model predictions using ProSA\(^2^0\) and ERRAT.\(^2^1\) ProSA analysis generated z-scores of \(-2.41\), \(-2.39\), and 0.19, for models generated with I-TASSER, IonCom, and RaptorX, respectively. ProSA residue scores using average energy over 40-residues generated plots with high fluctuations in residue quality for all three models (data not shown). ERRAT analysis generated results that agreed with those by ProSA. Overall quality factors for our models were of 64.533, 8.421, and 16.887 for I-TASSER, IonCom, and RaptorX, respectively. Therefore, we used the model generated by I-TASSER for all subsequent analyses.

I-TASSER identified several homology templates; however, none were of animal origin. The highest homology was identified as pleuralin-1 (PDB ID 2NBI) with a z-score of 4.90, a protein present in the marine diatom *Cylindrotheca fusiformis*. Other templates included yeast fatty acid synthase (PDB ID 2PFF) with a z-score of 1.62, and type II-C Cas9 enzyme (PDB ID 4OGC) with a z-score of 0.57 found in the bacterium *Actinomyces naeslundii*.

The hCABS1 model predicted by I-TASSER showed a molecule with minimal secondary structure (Figure 2A). Residues 210 to 247, 286 to 302, and 332 to 360, previously predicted to be in regions highly disordered (Figure 1C), do indeed appear as random coils in the I-TASSER model (sequences highlighted in red). The hCABS1 model predicted by IonCom also shows these three regions as random coils but predicted additional secondary structures at the N- and C-terminal ends of hCABS1 (Figure 2B). Interestingly, both the RaptorX software and IonCom predict the existence of an α-helix motif at the carboxyl terminal of hCABS1 (Figures 2A and 2C), which coincides with the anti-inflammatory motif TDIFELL, highlighted in green in all three models (Figure 2).
At present, 72 CABS1 sequences have been annotated in the NCBI protein database (Supplemental File 1). After analyzing them in Clustal Omega, we noted that the longest CABS1 protein sequence is 512 amino acids long, while the shortest is 156. The average length was 389 amino acids (Supplemental File 2).

Species with flanking amino acid domain insertions at the amino terminus were Acinonyx jubatus, Bubalus bubalis, Callorhinus ursinus, Camelus dromedarius, Camelus ferus, Enhydra lutris kenyoni, Eptesicus fuscus, Eumetopias jubatus, Felis catus, Lynx canadensis, Myotis davidii, Neomonachus schauinslandi, Odocoileus virginianus texanus, Phyllostomus discolor, Puma concolor, Sus scrofa, Ursus arctos horribilis, Vicugna pacos, and Zalophus californianus. The latter also had a domain insertion in the carboxyl terminus. Tupaia chinensis had an insertion in the middle region at residue 115 that no other species had. Species with significant amino acid deletions were Desypus novemcinctus, Fukomys damarensis, and Heterocephalus glaber (Supplemental File 2). Because of their significant insertions or deletions, we excluded these species for a subsequent Shannon entropy analysis of protein sequence conservation.

3.4 | An anti-inflammatory peptide sequence of CABS1, TxIFELL, is only found in primates of the infraorder Simiiformes

The anti-inflammatory sequence TDIFELL was first identified in humans and experiments showed that FELL also had anti-inflammatory activity. Thus, we categorized species containing FELL within the TxIFELL peptide as ones with the anti-inflammatory domain. Out of the 72 analyzed species, only 14 contain the putative anti-inflammatory domain “TxIFELL,” all of which were primates of the infraorder Simiiformes (Figure 3). Of note, Nancy Ma’s Night monkey (Aotus nanymaae) and Northern white-cheeked Gibbon (Nomascus leucogenys) show a point mutation, T$\rightarrow$GIFELL, whereas all the others have TDIFELL (Table 1; Figure 3). Interestingly, a model of intestinal anaphylaxis showed that “TxI” was not necessary to elicit anti-inflammatory activity. We observed that “FELL” elicited this activity and postulated that this tetrapeptide contains the core anti-inflammatory component. Moreover, a Shannon entropy analysis indicates that the anti-inflammatory sequence displays noticeably lower level of entropy in primates (Figure S1).

Current data indicates that annotated primates outside infraorder Simiiformes do not contain FELL, but FKLL. CABS1 in the testes has been characterized in the rat and house mouse. The peptide sequences in these rodent species in the region of the anti-inflammatory sequence in primates are SGIFKLL and SGLFKLL, respectively.

3.5 | Posttranslational modifications of hCABS1

NetPhos software was used to predict phosphorylation sites on hCABS1. A total of 14 residues showed a score of more than 95% probability, including 13 serine (residues S43, S77, S110, S170, S245, S273, S319, S337, S338, S344, S360, S364, and S376), and one threonine residue at position T303. Because protein kinase CK2 was previously reported to phosphorylate rat CABS1, we used Group-based Prediction System (GPS) to assess the specific contribution of CK2 on hCABS1. GPS determined 24 residues as likely targets for CK2-mediated phosphorylation. Seven of these residues were in agreement with NetPhos (Figure 4A). These seven residues and their
scores were S43/2.0802, S77/4.682, S110/1.801, S337/4.657, S338/4.329, S444/3.127, and S376/5.097. The sites were plotted on a cartoon representation of hCABS1 and on a stick representation model of hCABS1 using the I-TASSER model developed in Figure 2A.

Using SignalP-5.0 Server for signal peptide sequence prediction, we uploaded the FASTA sequence for hCABS1 and determined that the likelihood of a N-terminus hydrophobic signal peptide sequence was negligible (0.001). Signal peptide sequences specifically

![FIGURE 3](image-url)  
Phylogenetic tree of CABS1 using a maximum likelihood tree based on CABS1 amino acid sequences of annotated species in the NCBI protein database. Numbers in blue indicate branch support values (%). The anti-inflammatory sequence TDIFELL is restricted to members of the infraorder Simiiformes. CABS1, calcium-binding protein, spermatid-associated 1; NCBI, National Center for Biotechnology Information.

<table>
<thead>
<tr>
<th>Species</th>
<th>CABS1 anti-inflammatory domain</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Pan paniscus</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Gorilla gorilla gorilla</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Pongo abelii</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Nomascus leucogenys</td>
<td>381 T</td>
<td>G</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Rhinopithecus roxellana</td>
<td>381 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Piliocolobus tephrosceles</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Macaca fascicularis</td>
<td>381 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Macaca nemestrina</td>
<td>381 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Macaca mulatta</td>
<td>381 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Theropithecus gelada</td>
<td>380 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Papio anubis</td>
<td>381 T</td>
<td>D</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td>Aotus nancymaee</td>
<td>371 T</td>
<td>G</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>L</td>
</tr>
</tbody>
</table>

**TABLE 1**  
CABS1 anti-inflammatory domain is conserved in primate species

Abbreviation: CABS1, calcium-binding protein, spermatid-associated 1.
bind to the signal recognition particle receptor on the endoplasmic reticulum (ER) membrane to allow insertion of proteins through the ER membrane and enter the ER lumen space. Because hCABS1 lacks this sequence, it is not possible for hCABS1 to enter the ER and therefore must be translated as an uncleaved protein by ribosomes in the cell cytoplasm. This suggests that hCABS1 is secreted from salivary glands through nonclassical secretion mechanism(s) that is independent of the canonical ER-Golgi apparatus-secretory vesicle pathway. Several nonclassical secretion pathways include membrane transporters, multivesicular bodies, exosome/microvesicle shedding, lysosome secretion, and lytic release.

Because several molecular forms of hCABS1 have been identified, the PeptideCutter software, available through ExPASy, was used to predict potential cleavage sites. The software offers in silico digestions for numerous proteases. We found that 15 out of 34 available enzymes did not cleave hCABS1. From the remaining 19 enzymes, only eight are expressed in humans. In descending order, according to cleavage frequency, these enzymes are pepsin (137 cleavage sites), neutrophil elastase (48 cleavage sites), trypsin (28 cleavage sites), chymotrypsin (18 cleavage sites), Arg-C proteinase (5 cleavage sites), caspase 1 (2 cleavage sites), enterokinase (1 cleavage site), and thrombin (1 cleavage site). It is not surprising that digestive enzymes like pepsin, trypsin, and chymotrypsin could cleave hCABS1 at numerous sites. Also, not surprisingly, neutrophil elastase appears to cleave hCABS1 at a high number of sites (Figure 4B). Neutrophil elastase is an aggressive protease known to digest virtually all extracellular matrix proteins, in addition to its role in proteolytic activation of several matrix metalloproteinases during inflammation. The presence of neutrophil elastase in the oral cavity makes it a likely candidate for hCABS1 processing. We previously reported the presence of several hCABS1 forms in saliva using immunodetection by Western blot and capillary nano-immunoassay. To better understand the complex pattern of likely P1 residues for neutrophil elastase shown in Figure 4B, we calculated the “20 amino acid moving average” of P1 sites along the hCABS1 sequence (Figure 4C). Interestingly, cleavage site density correlated with hCABS1 regions that display highly unfolded secondary structure. This supports the hypothesis that hCABS1 is a proteolytic target of neutrophil elastase and that this enzyme may play an important role in the genesis of functionally distinct forms of hCABS1.

3.6 Cofactors and their role in hCABS1 biological function

In addition to tertiary structure, I-TASSER also predicts protein-cofactor interactions and likely polypeptide biological function. This software predicted interactions of hCABS1 with Ca$^{2+}$ (residues 42 and 44), Mg$^{2+}$ (residues 219 and 223), Zn$^{2+}$ (residues 214 and 216), leucine (residues 194, 195, 257, 263, and 267), and thiamine diphosphate (residue 360) (Figure 5A). To validate hCABS1 and TPP interaction, the software SwissDock was used. Swissdock reported a Full Fitness interaction of −1043.70 kcal/mol with an estimated ΔG of −9.06 kcal/mol. Space-filling and stick models showing amino acid residues involved in the interaction are shown in Figure 5B,C. Residues predicted to make H-bonds with

**DISCUSSION**

Acute and chronic stress play significant roles in the composition of soluble factors in saliva. While parasympathetic signals induce a water-rich salivary discharge, sympathetic stimulation induces a high protein salivary output. Accessibility to saliva and the presence of cell-free DNA, several RNA forms (mRNA, microRNA, and piwi-interacting RNA) and soluble or exosomic granules (exosomes, microvesicles, and apoptotic bodies), make this fluid an attractive diagnostic sample source.

CABS1 is an exciting, novel salivary biomarker with the potential to predict stress and resilience to stress. This duality is so far unique within the literature of stress biomarkers. Our current understanding of hCABS1 presents an interesting challenge. While CABS1 has been detected in sperm and mitochondria in nonhuman samples, our group has identified human forms of CABS1 secreted from salivary glands. We postulate that intracellular and extracellular localization for hCABS1 may represent two very distinct functional modes. For example, secreted hCABS1 may be involved in stress and immunoinflammatory processes, whereas intracellular hCABS1 may regulate energy metabolism and intracellular stress pathways (Figure 6). Our in silico evidence illustrates these two potential modes of action for hCABS1 by shedding light on interactions with cofactors like TPP, Ca²⁺, and Leu. Secretion through nonclassical mechanisms and tissue-specific processing are likely to be critical components of hCABS1 biology.

The stability of proteins in the extracellular environment impacts their function and mechanism of regulation. CABS1 lacks a stable tertiary structure and is therefore intrinsically disordered by nature. There is an "order-structure continuum" for tertiary structures, with rigid and well-structured protein families at one end of the spectrum, and extremely flexible and dynamic...
proteins at the other. It has been estimated that more than 50% of all proteins in eukaryotes contain long disordered regions that may play important roles in their function. In this study, authors reviewed the role of several disordered domain proteins in the biomineralization of teeth including dentin sialophosphoprotein, osteopontin, and dentin matrix protein-1. Others have reviewed the role of disordered domains and how this entropic ability allows, for example, spring-like function of titin and the bristle-like role of MAP2, to provide spacing in the cytoskeleton. Disordered proteins may also function as inhibitors as is the case of calpastatin and its ability to inhibit calpain due to long disordered domains that facilitate positioning of the inhibitory domains. The disordered nature of hCABS1 may enable its increased proteolytic processing by enzymes such as neutrophil elastase (Figure 4B,C). It is not surprising that there is an obvious overlap between predicted intrinsically disorder regions (Figure 1C) and neutrophil elastase cleavage sites (Figure 4C).

Neutrophil elastase activity may produce bioactive fragments of hCABS1 and contribute toward tissue-specific processing events. Indeed, the location of hCABS1 anti-inflammatory domain (380-TDIFELL-386) appears to be strategically isolated near the C-terminal, adjacent to a region with the highest density of neutrophil elastase putative P1 residues (Figure 4C). We predict that neutrophil elastase cleaves salivary hCABS1 to produce active components such as TDIFELL-containing peptides, which may act locally or possibly disseminate systemically to exert physiological effects in other tissues. Interestingly, the anti-inflammatory peptide sequence in rat SMR1, TDIFEGG, has an adjacent elastase cleavage site (unpublished and observation) and we have previously identified TDIFEGG-containing immunoreactivity in a less than 3000 Da fraction of rat saliva.

We aim to identify the role of neutrophil elastase and other proteases in the processing and function of molecular forms of hCABS1. Due to the high abundance of acidic residues in hCABS1, the migration of hCABS1 is slowed during electrophoresis, resulting in a significant discrepancy between its predicted molecular mass (M₀) and its observed Mₓ in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). This observation has been reported in human (43 kDa calculated and 75 kDa observed), Sus scrofa (42.8 kDa calculated and 75 kDa observed), mouse (42.3 kDa calculated and 66 kDa observed), and rat (42.3 kDa calculated and 80.0 kDa observed). Neutralization of acidic residues in CABS1 with carbodiimide/ethanolamine altered the migration profile of rat CABS1 close to its predicted M₀, supporting acidic residues as the reason for the Mₓ discrepancy. In agreement with this observation, we have shown that N- or O-glycanase treatment does not alter CABS1 Mₓ in SDS-PAGE, suggesting that glycosylation events do not explain the discrepancies in Mₓ of the protein.

Other posttranslational modifications like phosphorylation have been observed for CABS1 in vivo. However, a single phosphorylation event only contributes 80 Da in Mₓ. Thus, this posttranslational modification

**FIGURE 6** Conceptual model of hCABS1 biology. The human CABS1 gene (hCABS1 gene, blue label) contains a single exon and generates a single gene product. The polypeptide translated from this transcript (hCABS1, black label) undergoes posttranslational intracellular modifications such as proteolytic processing and phosphorylation (red circles). We postulate that in a tissue-specific manner, various forms of hCABS1 can either localize to the mitochondria or be secreted via multivesicular bodies and exosomes. Mitochondrial hCABS1 is associated to the inner mitochondrial membrane and interacts with various cofactors including leucine (Leu), Mg²⁺ (yellow ion), Ca²⁺ (blue ion), Zn²⁺ (orange ion), and thiamine diphosphate (TPP). Interaction of cofactors with hCABS1 enhances its function in regulating energy metabolism and intracellular stress pathways. Secreted hCABS1 undergoes further tissue-specific processing. hCABS1 processing would vary according to the proteolytic enzyme environment. Various hCABS1 molecular forms (hCABS1 forms) are likely to have specific roles in stress, immunoinflammation and physiology. hCABS1, human calcium-binding protein, spermatid-associated 1
does not account for the unexpected increase of ~30 kDa in observed Mr.

Though phosphorylation minimally adds to molecular weight, phosphorylation events typically have enhancing effects on protein function. Protein kinase CK2 is critical in rat spermiogenesis, and CK2 was shown to phosphorylate rat CABS1 in vitro and confirmed in vivo.\(^{36}\) The phosphorylation of CABS1 by CK2 suggests that their interaction plays an important role in spermiogenesis. We suspect that hCABS1 phosphorylation may play various roles in regulating mitochondrial pathways in energy metabolism and stress (Figure 6). Conversely, CABS1 phosphorylation may be necessary for its secretion. Recently, Klement and Medzihradszky\(^ {26}\) provided an in-depth review on the importance of extracellular protein phosphorylation. They reported that in some cases, as high as 50% of all secreted proteins may be released via nonclassical ER/Golgi pathways. Phosphorylation of this group of secreted proteins may be a required modification for secretion. In this case, hCABS1 phosphorylation may be a prerequisite for its secretion from SMG. Extracellular phosphorylation of hCABS1 could be critical in regulating its functions in stress, immunoinflammation, and physiology, possibly by controlling proteolytic processing of hCABS1.

Interestingly, calcium plays a role in regulating proteolytic cleavage and is involved in cell signaling within the salivary glands.\(^ {49}\) Reinhardt et al.\(^ {50}\) showed that proteolytic degradation of fibrillin-1 in the presence of CaCl\(_2\) was significantly slower than in the presence of EDTA, demonstrating that calcium confers protection against proteolysis. This calcium-dependent protection from proteolysis has been reported in other studies.\(^ {51,52}\)

The possibility that hCABS1 binds calcium is inferred from studies in mice and pigs using ruthenium red and Stains-all\(^ {25,37}\) and in rats using \(^ {45}\)Ca\(^ {2+}\)-binding autoradiography assays.\(^ {36}\) Our in silico analysis confirms this hypothesis and predicts the location of these sites at positions 42 to 43 and 248 to 252 (Figure 5A). Future studies can be conducted to confirm these Ca\(^ {2+}\)-binding sites by scanning mutagenesis and other labeling methods.

Calcium maintains structural stability of proteins, including a highly abundant salivary protein, \(\alpha\)-amylase.\(^ {53}\) In the presence of sympathetic stimulus, calcium may be involved in signaling pathways for hCABS1, by inducing structural modifications that regulate proteolytic cleavage of a hCABS1 precursor, forming smaller molecular forms that may have varying properties. Proteolytic degradation assays should, therefore, show increased degradation of hCABS1 in the absence of calcium, and will be reversed in the presence of CaCl\(_2\), providing evidence that Ca\(^ {2+}\)-binding promotes a more ordered secondary structure, regulating hCABS1 processing.

We were intrigued by the I-TASSER predictions for hCABS1 binding sites for cofactors thiamine diphosphate (TPP), Ca\(^ {2+}\), and Mg\(^ {2+}\) (Figure 5A). In sperm, CABS1 localizes to the mitochondria, flagella, and the acrosome suggesting a role in energy supply.\(^ {6,23}\) In humans, there are 12 enzymes that use TPP as a cofactor as reported in the Kyoto Encyclopedia of Genes and Genomes, all of which play crucial roles in energy metabolism. Interestingly, all TPP-binding enzymes are intracellular proteins. If hCABS1 binds TPP, this would be indicative of novel intracellular roles likely in the regulation of energy metabolism. Based on metabolic pathways, TPP-binding enzymes can be divided into six groups, three groups that function as cytosolic enzymes, and three other groups that function within mitochondria. The first group of cytosolic enzymes converts thiamine into TPP or into thiamine triphosphate; these enzymes are thiamine phosphate kinase (2.7.4.16), thiamine diphosphokinase (2.7.6.2), thiamine triphosphatase, (3.6.1.28), and thiamine diphosphate kinase (2.7.4.15). The second group of cytosolic enzymes are involved in ATP hydrolysis; these enzymes are adenylate kinase (2.7.4.3) and nucleoside triphosphate phosphatase (3.6.1.15). The third group of cytosolic enzymes is transketolase (2.2.1.1), involved in the pentose phosphate pathway.

The three groups that are inside mitochondria are involved in the tricarboxylic acid (TCA) cycle, pyruvate metabolism, and branched-chain amino acid catabolism, respectively. These three groups of enzymes form closely related multienzyme complexes. Within the oxoglutarate dehydrogenase complex in the mitochondrial matrix, the enzyme oxoglutarate decarboxylase (1.2.4.2) of the TCA cycle requires TPP. Another mitochondrial matrix complex is the pyruvate dehydrogenase complex involved in the conversion of pyruvate to acetyl-CoA; within this complex, pyruvate dehydrogenase (1.2.4.1) also requires TPP. In the mitochondrial inner membrane, the branched-chain alpha-ketoacid dehydrogenase complex plays a role in the metabolism of branched-chain amino acids (leucine, isoleucine, and valine); within this complex, the enzyme branched-chain alpha-ketoacid dehydrogenase (1.2.4.4) mediates the TPP-requiring step. For a more comprehensive list of TPP enzymes present in other organisms including bacteria, see review.\(^ {54}\)

For TPP-binding enzymes, the combination of TPP with either Mg\(^ {2+}\) or Ca\(^ {2+}\) is necessary for enzyme activation.\(^ {55,56}\) The divalent cation location observed in X-ray crystallography studies for pyruvate dehydrogenase\(^ {57}\) and branched-chain alpha-ketoacid dehydrogenase\(^ {58}\) show the ion in close proximity to TPP. In contrast, the hCABS1 structure shows Ca\(^ {2+}\) and Mg\(^ {2+}\) binding sites relatively far from the TPP site (Figure 5A). It is possible that the long random coil (res 268-315) of hCABS1 acts as a molecular...
switch to bring TPP close to either Ca\(^{2+}\) or Mg\(^{2+}\) ions. In this model, hCABS1 would have enzymatic activity when the TPP and divalent ion regions are in close proximity. Putative sites for neutrophil elastase cleavage exist between these cofactor domains of hCABS1, thus elastase-mediated cleavage could result in deactivation of hCABS1 enzymatic activity extracellularly.

The branched-chain alpha-ketoacid dehydrogenase complex is located in the inner mitochondrial membrane where it utilizes TPP in the first irreversible step in the catabolism of leucine (Leu).\(^3\) Leu is an essential amino acid that plays a role in protein structure and function. Leu also plays roles in numerous other protective pathways including wound healing, muscle repair, and protects against stress-induced breakdown of muscle proteins. Given hCABS1 association with stress responses, it is possible that binding of Leu to hCABS1 could play a regulatory function by acting as an allosteric regulator. Indeed, Leu functions as an allosteric activator of the enzyme glutamate dehydrogenase.\(^60\)

To enable the analysis of genetic data, the Gene Ontology project developed the Phylogenetic Annotation and Inference Tool (PAINT). One of the PAINT objectives was to assist annotators in assignment of protein function within a given protein family. This tool predicts that hCABS1 should have a subcellular location to the cytosol and the inner mitochondrial membrane.\(^61\) Indeed, Calvel et al\(^36\) reported that rat CABS1 is associated to mitochondria inner membranes of spermatids. The mitochondria inner membrane is home to numerous metabolic pathways, key in energy metabolism. As described above, the branched-chain alpha-ketoacid dehydrogenase complex metabolizes Leu catabolism, requires TPP and Mg\(^{2+}\) and it is found in the inner mitochondrial membrane, all features we have identified for hCABS1 in this manuscript. At this point, it is difficult to predict weather Leu acts as a substrate or as a regulator of hCABS1 activity.

We are also interested in the occurrence of CABS1 across species. Annotations of the protein only appear to occur in placental mammals. An additional segment at the amino terminus of at least 7 amino acids was observed in annotated felines, most pinnipeds, most bats, bovids, camelids, the grizzly bears, wild boars, and sea otters. Since we have shown that hCABS1 contains an anti-inflammatory domain, TDIFELL, near the C-terminus, we investigated the species distribution of this or similar motifs. Interestingly, the 14 species with TDIFELL were all primates (Figure 3). A domain in CABS1 with anti-inflammatory properties, is conserved only in parvorders Catarrhini (Old World monkeys, humans) and Platyrrhini (New World monkeys), which belong to the order Primates (Figure 3, highlighted red square).\(^62\) This domain is not present in Philippine tarsier (Carlito syrichta), which is also part of this order. Two primate species showed a point mutation in the anti-inflammatory motif, D→G. However, aspartic acid and glycine share the physicochemical property of being small.\(^63\) Moreover, we have showed that FELL is sufficient to elicit an anti-inflammatory effect in rodents,\(^12\) so we speculate that this CABS1 domain has anti-inflammatory activity across primates.

NCBI’s Conserved Protein Domain Family Database, reports hCABS1 as the only member of its superfamily. This analysis reflects the unique nature of the protein, a limitation that hampers our ability to make functional predictions using current repositories. Future research on hCABS1 regulation and function will garner insight on how hCABS1 contributes to the stress response and the relationship between its localization and its function. Beyond a biomarker of stress, this study may reveal other unrecognized roles of hCABS1, including possible enzymatic activities. This manuscript contributes to the growing body of knowledge of the regulation of this emerging protein with wide-ranging implications on our understanding of its biology.

ACKNOWLEDGMENTS

We thank Stewart Cook (Dean of the School of Applied Sciences and Technology from NAIT) and Trevor April (Department Head of the Department of Natural Sciences and Academic Studies from NAIT) for providingdownload support. The authors gratefully acknowledge the assistance of Meghan Dueck, from the University of Alberta who advised on phylogenetic analysis. We thank AllerGen NCE Inc, The University of Alberta’s Faculty of Medicine and Dentistry/Alberta Health Services Graduate Student Recruitment Studentship, and the 75th Anniversary Award for financial support.

CONFLICT OF INTERESTS

There are no conflicts of interests in the publication or work associated with this manuscript.

AUTHOR CONTRIBUTIONS

All six authors contributed conceptually and in the design of the study. The document was written primarily by MM-P with contributions from ER-S and PL. All figures were created by MM-P and ER-S. All authors provided editorial contributions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions. The data supporting the findings in this manuscript are available as Supplemental Files 1 and 2.
REFERENCES


SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

https://doi.org/10.1002/jcb.29824