Bioinformatic analysis of nucleotide cyclase functional centers and development of ACPred webserver

Nuo Xu
Department of Computer Science
Wenzhou-Kean University
88 Daxue Rd, Wenzhou, Ouhai, Zhejiang, China 325060
xunu@kean.edu

Changjiang Zhang
Department of Computer Science
Wenzhou-Kean University
88 Daxue Rd, Wenzhou, Ouhai, Zhejiang, China 325060
czhang@kean.edu

Leng Leng Lim
Department of Mathematics
Wenzhou-Kean University
88 Daxue Rd, Wenzhou, Ouhai, Zhejiang, China 325060
llim@kean.edu

Aloysius Wong
Department of Biology
Wenzhou-Kean University
88 Daxue Rd, Wenzhou, Ouhai, Zhejiang, China 325060
alwong@kean.edu

ABSTRACT

Cyclic mononucleotides, in particular 3',5'-cyclic guanosine monophosphate (cGMP) and 3',5'-cyclic adenosine monophosphate (cAMP), are molecular signals that mediate a myriad of biological responses in organisms across the tree of life. In plants, they transduce signals such as hormones and peptides perceived at receptors on the cell surface into the cytoplasm to orchestrate a cascade of biochemical reactions that enable them to grow and develop, and adapt to light, hormones, salt and drought stresses as well as pathogens. However, their generating enzymes (guanylyl cyclases, GCs and adenylyl cyclases, ACs) have just been recently discovered and are still poorly understood. Here, we employed a computational approach to probe the physicochemical properties of the catalytic centers of these enzymes and the knowledge of which, was used to create a web-based tool, ACPred (http://gcpred.com/acpred) for the prediction of AC functional centers from amino acid sequence. Understanding the nature of such catalytic centers have enabled the creation of predictive tools such as ACPred which will in turn, facilitate the discovery of novel cellular components across different systems.

CCS CONCEPTS

• **Computing methodologies** → *Data mining; Information extraction; Prediction algorithm;* • **Applied computing** → *Sequence analysis; Bioinformatics*

ACM-BCB'18, August 29–September 1, 2018, Washington, DC, USA

ACM ISBN 978-1-4503-5794-4/18/08… $15.00
DOI: https://doi.org/10.1145/3233547.3233549

KEYWORDS

Search Motif; Functional Centers; Sequence Analysis; ACPred; Adenylyl Cyclase Prediction; Webserver; Computational Biology

1 INTRODUCTION

Cells make use of molecular signals to relay information such as hormones and growth factors perceived at the cell surface, into the cell to trigger a cascade of biochemical reactions that will lead to responses at the physiological level [1, 2]. This transduction of signal from the external environment into the cell is crucial for the growth and development of an organism and, to allow the organism to efficiently response to changes in the environment [3-6]. Universal molecular signals such as small peptides, hormones, organic molecules, calcium ions and cyclic nucleotides (3',5'-cyclic guanosine monophosphate (cGMP) and 3',5'-cyclic adenosine monophosphate (cAMP)) have long been shown to exist in organisms across the tree of life and signaling a myriad of biological responses [7-13]. In animals including humans, cGMP is a vasodilating signal that is perhaps most famous for causing penile erection upon sexual stimulation [14]. In plants, cGMP also signals many biological responses including responses to light, hormones, salt and drought stresses as well as ozone and pathogens [for review, see [10, 13, 15]. Meanwhile, cAMP signals polarized pollen tube growth, stomatal opening, responses to light and temperature, and modulates ion transport [16-20]. These signaling molecules are particularly important for plants because unlike animals, they are sessile organisms and cannot run away from danger. They must therefore rely on a set of molecular signals such as cGMP and cAMP to provide efficient cellular signaling mechanisms in order for them to adapt and survive [1-3, 9, 21, 22]. However, the enzymes (guanylyl cyclases, GCs and adenylyl cyclases, ACs) that generate cGMP and cAMP in plants have just been recently discovered and are still poorly understood although they have been well-characterized in other systems such

as animals and bacteria. One reason for their apparent elusiveness is that plant cells have complex domain architecture of proteins that can perform multiple functions e.g., at the extracellular region, they can perceive and bind to ligands while at the cytosolic region, they can bind to proteins and/or organic compounds at modulatory sites or catalyze certain reactions [12, 13, 15, 23-26]. This is attributed to divergent evolution where plant signal perception and downstream cellular reactions are distinct from those of other eukaryotes. In a relatively crowded plant cell occupied by a large central vacuole, proteins assume multiple roles and GCs and ACs are well-placed in their microenvironments to perform highly localized signaling functions that include for e.g., switching from one signaling network to another [27-32]. Therefore, in plants, receptors and signaling molecules cannot be identified using standard homology-based searches querying with proteins from lower or higher eukaryotes because it is beyond detection limits, hence their apparent elusiveness [24, 33].

Recently, a motif-based approach has led to the discovery of a new class of GCs and ACs some of which, have been studied in greater detail [34]. This new class of enzymes are known as functional centers and are structurally different from canonical GCs and ACs found in other organisms. They are usually found embedded within larger primary domains in complex multi-functional proteins and while they possess the conserved key amino acids for catalysis, they do not resemble the overall structure of stand-alone canonical GCs and ACs [26, 33]. Their discovery has in recent years, prompted intriguing questions regarding their regulatory roles at both the molecular and biological levels. Emerging experimental data have shed light on some of these queries, but many more remain unanswered. Discovering new functional centers will help elucidate unknown functions and contribute to the understanding of the nature of these group of enzymes. As such, the discovery of these functional centers requires automation in the form of a web-server. We have recently created a web-tool for the prediction of GC functional center called GCPred (http://www.gcpred.com) and have tested this tool on both plant and animal proteins [33].

In the same manner, a predictive tool is required for the identification of candidate AC centers. Here, we probe the physicochemical properties of known GCs and ACs to understand the nature of these highly similar catalytic centers and use this knowledge to develop ACPred (http://gcpred.com/acpred).

2 METHODS

2.1 Bioinformatic analysis of AC and GC centers

Experimentally validated GC and AC centers from plant proteins were analyzed in terms of their overall domain organization and structural architecture. The domain organization of proteins were presented as 2D columns with lengths adjusted to approximately reflect their relative amino acid lengths and they are all aligned at their AC centers. Information about the annotated domains, transmembrane regions and other key components of the proteins were obtained from UniProt (http://www.uniprot.org) and presented in different colors accompanied by a legend. Representative 3D structure of each GC and an AC containing docked GTP or ATP substrates were prepared using UCSF Chimera visualization software available at https://www.cgl.ucsf.edu/chimera. The catalytic centers were colored, and substrate orientations clearly defined. The key amino acids in the motif that are involved in substrate interaction functions were colored according to surface charge.

Next, the physicochemical properties of the catalytic centers were analyzed based on known values of 3 categories: hydrophobicity, molecular weight and isoelectric point, of each amino acid. The data for each protein is presented individually as a heatmap with red representing highest and white representing lowest values in that category. Average values of GCs and ACs in each category were also represented in line graphs to show variations between GCs and ACs.

2.2 Development of the ACPred server

The website ACPred was written in HTML, PHP and CSS, running on an Ubuntu server that LNMP package (Nginx, MySQL, PHP) installed. The header of the home page with brown-white linear ingredient background was defined by a CSS library Bootstrap. The image set on the right of the title ACPred, is a representative model of an AC center docked with the ATP substrate, making it clear for the user to understand the purpose of our website, that is to identify candidate ACs. In the home page, we first explain the utility of this tool giving the background knowledge and other important details as well as providing detailed instructions written in large size Arial font to guide first-time users. Next, we include a ratio element following the instruction asking users if metal ion binding feature should be included in the prediction. We also include a FASTA format example that contains the name of the protein in the header and its full-length amino acid sequence. Below the example, there is a large text area that can automatically adjust its width according to the screen size of the web browser. Characters typed inside the text area will be automatically converted to small sized font with courier style and the user is also able to enter multiple sequences at one time. Although the height of text area is fixed at 400px, the user can still input as many as they want because the text area allows for over-flow. At the bottom of the text area, there is a button element that enables user to submit the input data as string to the server. This submit button is disabled until the text area is filled, otherwise a warning message will be shown. To quickly learn the feature of our webserver, the user only need to copy the example into the text area (full sequence required) and then click the submit button to view the result page. At the footer of the home page, we provide the prior works that have led to the development of this server as clickable reference links which will take the user directly to the primary source of the articles.

The main calculation is done in the result page and written in PHP. The header of the result page is almost the same as that in the home page, but we replaced the instruction with an interpretation guide that helps the user make good judgment on their retrieved hits. The AC report for each sequence can be

divided into 3 parts: a sequence panel, a hits table and 3 bar charts. If the users enter more than one sequence, these 3 parts will be shown as a loop, and the report for each sequence is independent. However, none of the 3 parts will be shown if no AC center was identified in any given sequence. The sequence panel contains a bold header with the protein name and a numbered body that shows 100 amino acids in each line. To do this, we created a PHP function "split" that could first split the string that was entered in the home page to several single sequences according to the special symbol ">" which is the symbol at the beginning of a FASTA format sequence and check if the full string contains more than one sequence. Then, position numbers will be generated for each sequence through the PHP function "position" if more than one hits (that contain 14 amino acids and satisfy the AC motif shown in Fig. 1) were identified within each sequence. Each hit with its own position number will be highlighted in the sequence panel and shown in the hits table. Next, to fill in the ACC hydrophobic value, ACC molecular weight, ACC isoelectric point and ACC mean value, we created another PHP function "input" that will call MySQL procedure "input check" with 2 parameters: the full sequence and the unique position number for each ACC hit. Thus, the number of invoke times of this PHP function depends on the number of ACC hits found. The procedure "input check" will first pick out the objective hit according to the 2 parameters that passed from the PHP function, then splits the hit into 14 amino acids. In MySQL database, we have a "hydrophobic table" that stores known amino acids' hydrophobic values, a "molecular weight" table that stores known amino acids' molecular weights, an "isoelectric point" table that stores known amino acids' isoelectric point values. The procedure will calculate those 4 values (termed ACC hydrophobicity, ACC molecular weight, ACC isoelectric point) for the objective AC hit according to the algorithm shown in Fig. 7B. Those 4 values will be scaled 0 to 1 with 3 decimal numbers and shown in 3 possible colors (green, red, black) based on specific cut off points determined from currently available experimental data. Color definition and labels are described under each table. Following each table, there are 3 bar charts created using the open source PHP graphic library "pchart". Because each hit contains 14 amino acids, the horizontal axis of the bar chart is fixed to represent these amino acids beginning from amino acid at position 1 to 14 respectively. Unlike the hits table where each row would only cover the values of one AC domain, the 3 types of the values (ACC hydrophobicity, ACC molecular weight, ACC isoelectric point) of the hits within the same sequence will be combined into the 3 bar charts named as ACC hydrophobicity charts, ACC molecular weight chart and ACC isoelectric point chart. All these 3 charts show the deviation from mean values for each hit. For instance, if the ACC hydrophobic value of one ACC domain is shown in green, then the average height of the bar in the hydrophobic chart will be closer to 0 than the black or red values. However, if there are more than one hit within the same sequence, each hit will be given a unique color, starting from green, then red, blue, etc. All the bar charts have the same height and weight, but they will automatically adjust the unit size of the vertical axis according the maximum and minimum heights of bars. At the bottom of the result page, there is a link that allow the user to download the retrieved hits as a *.csv file.

3 RESULTS AND DISCUSSION

3.1 Domain architecture of GCs and ACs

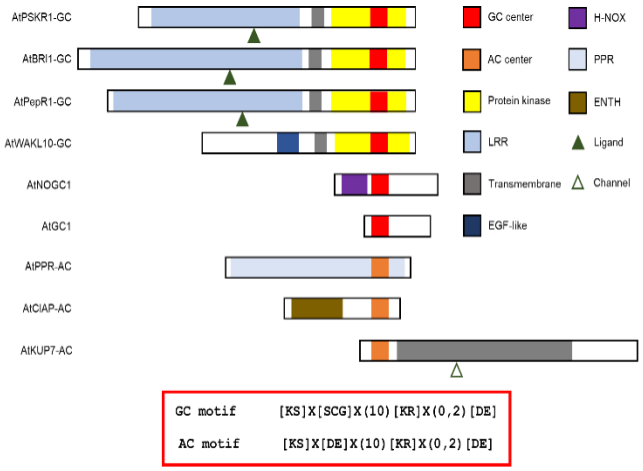

Figure 1: Domain organizations of GC and AC centers.

A motif-based approach has in the past, identified several GC and AC centers [13, 26]. The motif was constructed by including only conserved amino acids within the catalytic centers of canonical GCs and ACs from prokaryotes and eukaryotes. Specifically, the motifs consist of 14-amino acid long amino acids with amino acid in position 1 forming hydrogen bonds with guanine or adenosine of GTP or ATP respectively and amino acid in position 3 determining substrate specificity while amino acid in position 14 of the motif binds to the phosphate acyl group and stabilizes the transition of substrate to its cyclic form [35]. The motifs may also undergo rational modifications to include species-specific and metal-binding filters or amino acids of similar chemical properties and they have been particularly successful when used in tandem with structural modeling and docking simulations [23-25]. In higher plants, known GC centers are AtPSKR1, AtBRI1, AtWAKL10, AtPepR1, AtPNP-R1, AGNOGC1, AtGC1, PnGC1 and HpPepR1 [35-43] while AtClAP, AtPPR-AC, AtKUP7 remain the only experimentally confirmed AC centers to-date [44]. Here, we show the domain architecture of experimentally confirmed proteins with GC and AC activities from the model plant *Arabidopsis thaliana* in Fig. 1. It is obvious that these catalytic centers occupy complex proteins that have different primary functions. For instance, many GC centers are found embedded within a larger kinase domain of hormone/peptide receptor complexes (AtPSKR1, AtBRI1, AtPepR1, AtPNP-R1 and AtWAKl10) thus suggestive of a role for GCs in regulating reactions in the hormone/peptide-dependent pathways [36, 37, 39-41, 45]. Indeed, the GC activities of AtBRI1 and AtPSKR1 have been shown to be intricately linked to their kinase domains which they reside in [27, 32, 37, 45]. Furthermore, binding of the extracellular receptor domains to their natural ligands can elevate

cytosolic cGMP levels and in the case of AtPNP-R1, enables regulation of ion and water homeostasis [41]. Meanwhile, AC centers are found in proteins that have more varied primary functions. For example, AtKUP7 acts primarily as a potassium transporter while AtClAP which assembles clathrin during endocytosis, is implicated in plant defense [44, 46, 47].

In addition to the domain architectures, we also show the 3D structures of typical GC and AC centers using AtPSKR1 and AtKUP7 as representatives (Fig. 2). From a structural perspective, the GC and AC centers share similar secondary folds where amino acids from position 1 to 14 of the motif form an alpha-helix that is followed by a solvent-exposed loop harboring a positively charged [RK] amino acid. At the tertiary level, they form a clear cavity that spatially fits the GTP or ATP substrate albeit accommodating them only at specific substrate orientations. Previous structural works have ascertained that the substrate must assume a binding pose where the nucleotide region of the substrate points towards the residue at position 1 of the motif located deep into the pocket at the catalytic center while the hydrophilic negatively charged phosphate end points towards the positively charged amino acid at position 14 of the motif and protruding outward from the cavity orifice. This binding pose is deemed favorable for catalysis [23-25]. The structural similarity between GC and AC centers is not surprising given that their substrates share considerable physical and chemical properties. But, is there a difference in substrate affinity between such GC and AC centers? And if so, how do they discriminate the GTP and ATP substrates?

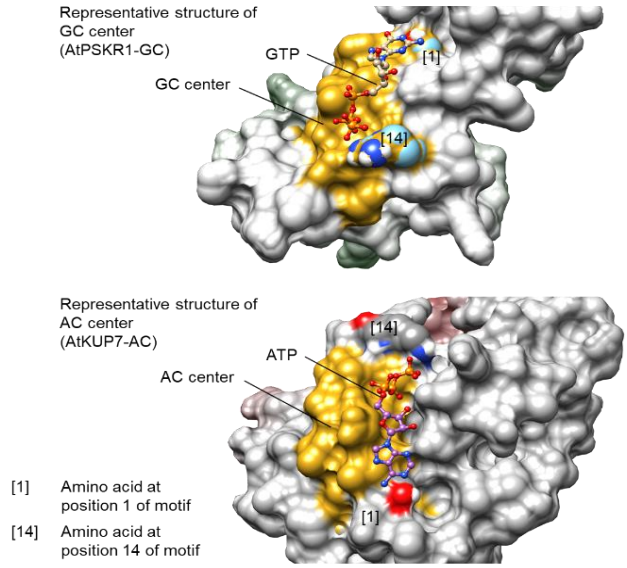

Figure 2: Representative structure of GC and AC centers.

3.2 Physicochemical signatures of GCs and ACs

Previous mutational works and computational simulations have showed hints of substrate preference at the catalytic center of GCs. For instance, in AtBRI1, mutations at position 3 ($G^{989}K$ and $G^{989}I$) of the motif reduced the catalytic activity of the GC center with G-I mutation being the most severe [45] while previous computational simulations on AtPSKR1 also indicate reduced substrate affinity when amino acid at the same position of the motif was mutated [25].

Since mutation of the amino acid at position 3 retains some catalytic activity [45] and it is also unlikely that these catalytic centers can spatially discriminate GTP and ATP, substrate preference must therefore be conferred by surface charges and/or other physical means such as hydrophobicity. To this end, we probe the physicochemical properties of known GC and AC centers. In particular, we analyzed the hydrophobicity, isoelectric point and molecular weight of amino acids in the GC and AC centers and expressed them as heatmap in Fig. 3. Among the GCs, the amino acids at each position of the motif show consistent physicochemical properties but among the ACs, these properties are more varied from one protein to another. For instance, the hydrophobicity of amino acids residing at positions 5 and 8 of the motif show considerable variation among the ACs while the same is also observed for the isoelectric point of ACs at position 10 (red boxes in Fig. 3). In general, the AC centers show greater variation in physicochemical properties across position 1 to 14 of the motif compared to the GC centers and we argue that this is necessary to compensate for the lack of 'stickiness' of the adenine nucleotide that lacked a =O group (red circle in Fig. 4).

We also expressed this analysis as average values for all three physicochemical properties in Fig. 5. When comparing ACs to GCs, there is significant difference in the hydrophobicity and isoelectric point of the amino acids at the positions indicated by red arrows in Fig. 5. We suspect that this, and the greater variation among amino acids within the centers might be an intrinsic nature of ACs that is necessary for optimal binding to the ATP substrate. Unlike GTP, ATP lacks a =O group (see blue and red circles in Fig. 4), thus may require greater difference in amino acid charges and hydrophobicity at the catalytic center for optimal binding and catalysis. It was previously suggested that the negatively charged amino acid [DE] at position 3 of AC motif confers substrate specificity to ATP [48] but it is likely that intermediary or flanking amino acid residues at the catalytic center play a role as well since mutations at position 3 of this motif did not completely abolish enzymatic activity of AtBRI1-GC [45].

Here, we showed that there is indeed considerable difference in the physicochemical properties of intermediary amino acids between GCs and ACs where in ACs, there is high variation in these properties which together with the negatively charged amino acid at position 3 [DE] of the motif, enable ACs to bind ATP. Other factors such as spatial and temporal abundance of GTP and ATP in microenvironments of the cell as well as the dependence on extracellular ligand binding and catalytic activity of primary domains such as kinases, can regulate cyclic mononucleotide generation by GCs and ACs. Further research is required to determine how these functional centers discriminate their substrates as computational methods including the bioinformatics analysis given here as well as docking simulations and biochemical evidence done elsewhere, have previously demonstrated that these catalytic centers can discriminate substrates despite being very similar in nature [15, 25, 33].

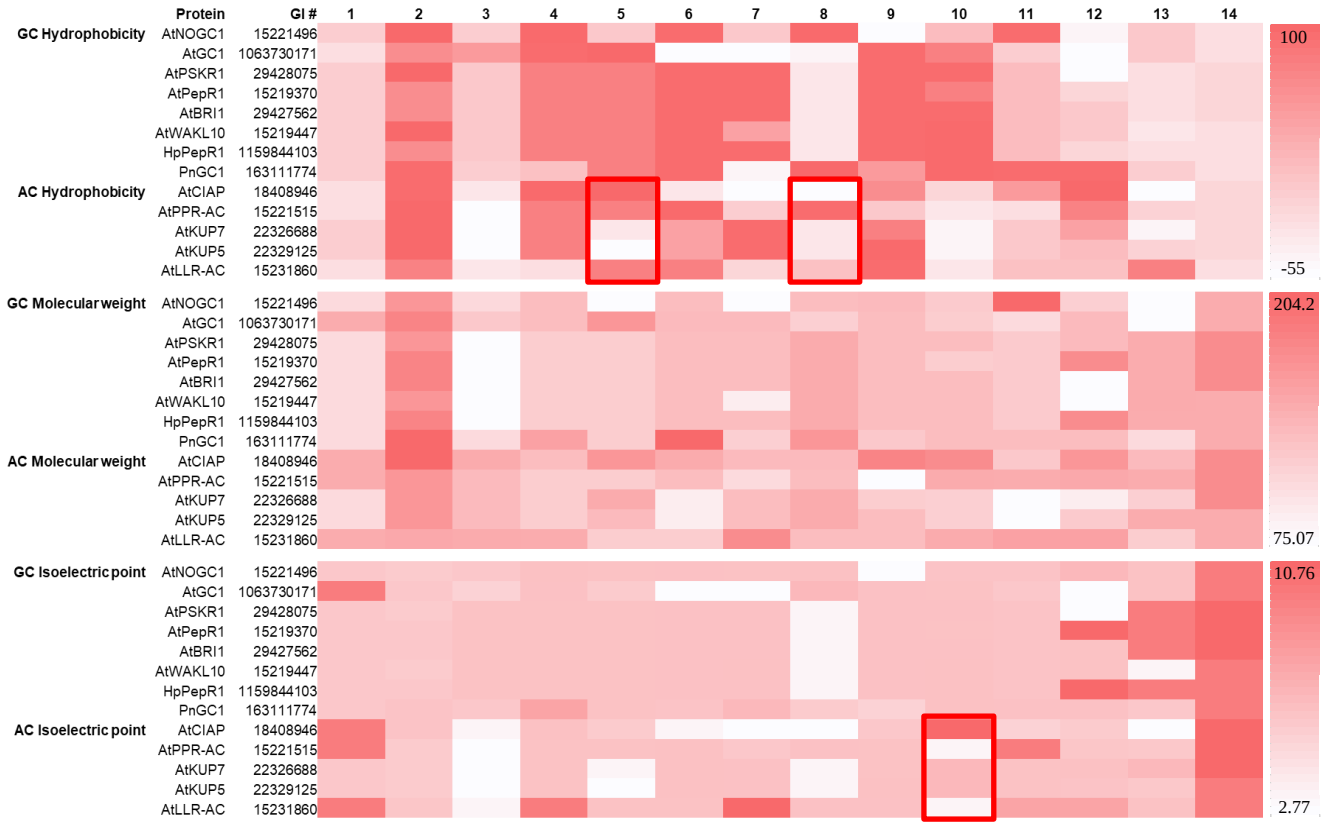

Figure 3: Heatmap illustrating the physicochemical properties of known GC and AC centers.

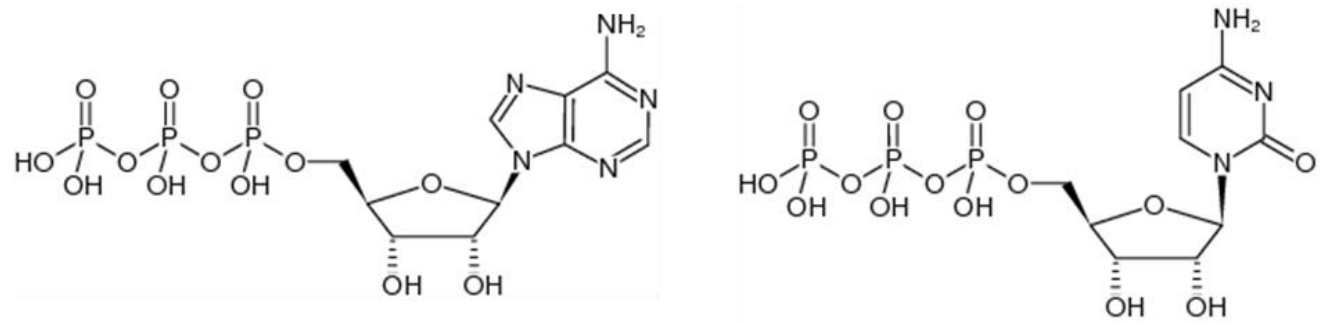

Figure 4: Structure of ATP (left) and GTP (right).

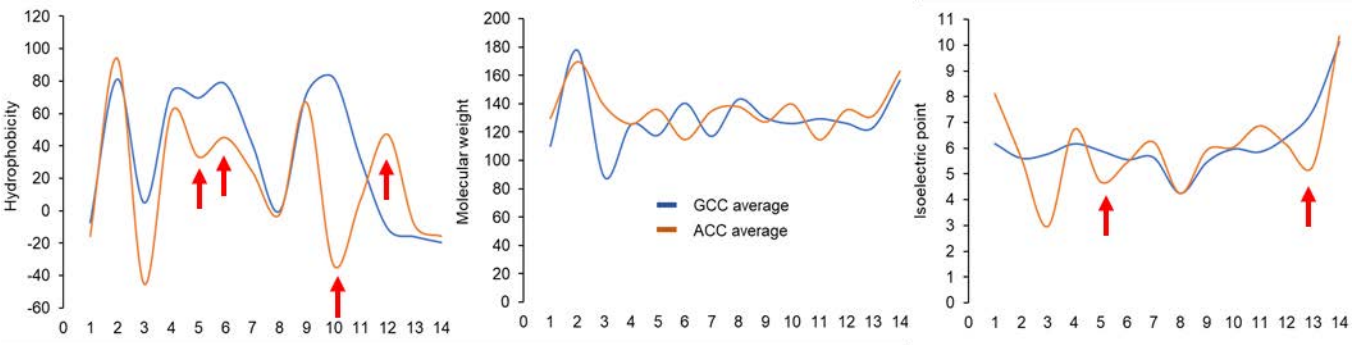

Figure 5: Average physicochemical values of known GC and AC centers.

3.3 The ACPred server

In order to rapidly predict candidate AC centers, we have developed ACPred. ACPred is a webserver built based on the algorithm and function of its parent server GCPred created previously for the prediction of GC centers [33]. The ACPred server allows user to enter single or multiple amino acid sequences in FASTA format and returns predicted hits ranked based on a set of numerical ACC values. The server uses algorithm that calculates predicted ACCs based on a set of physicochemical properties of amino acids in experimentally validated ACCs as presented in Fig. 6, where hits that contain the conserved amino acids at positions 1, 3 and 14 of a continuous string of amino acids are assigned statistical values 0-1, with 1 being the closest to the physicochemical properties of current ACC population. After submitting the queried sequence, the server returns a table of predicted ACCs that is accompanied by a set of ACC values normalized 0-1, which are color-coded to aid interpretations. In addition to the table, the result page also provides visual aids in the form of graphs that depict deviations of amino acids at individual positions of the ACC from mean values calculated from known ACC population. This comparative analysis at the single amino acid level may be useful for those interested in further probing of their candidates by e.g., guiding mutation and structural experiments. Previous works have established that this class of ACC resembles the GCCs where it typically contains 14 amino acids where the amino acids at positions 1, 3 and 14 have direct substrate binding and catalytic functions. We note that ACPred is only able to predict functional AC centers and not canonical AC domains or transmembrane regions.

The workflow and algorithm of ACPred is presented in Fig. 7. In step 1 of the ACPred workflow, the user enters single or multiple amino acid sequence in FASTA format and then click submit. The server screens the user input sequence for the presence of conserved amino acids at positions 1, 3 and 14 in continuous fashion in step 2 and if present, a data filtering process in step 3 removes other amino acids from the input sequence thus retaining only the identified ACC candidates. In step 4, a set of calculations are performed to determine the physicochemical properties of amino acids at each position of the ACC candidates and assigned values of 0-1 based on how close they resemble values of experimentally validated ACC population. The equations are this calculation is presented in the algorithm of Fig. 7. Specifically, if AC Domain "j" exists in sequence A, the AC algorithm calculates ACC hydrophobic (GH), molecular weight (GW), isoelectric point (GP) and ACC mean ($\bar{A}$) values (0-1) using the equations shown in the "perform calculations" box for k^{th} amino acid (where k = intermediary amino acids in the ACC) based on mean values of ACC population in the ACC database. The algorithm then generates a report that includes tables of ACC values and charts depicting variation from population mean and HTTP response sends the result page to the user in step 5 of the workflow. ACC database in Fig. 6 contains mean physicochemical values of amino acids at each position of known ACCs and from which calculations of input sequence were based on. Calculated values of each amino acid property were scaled 0-1 giving rise to "ACC values" where 1 represents closest to ACC population mean thus most probable and 0 is least probable.

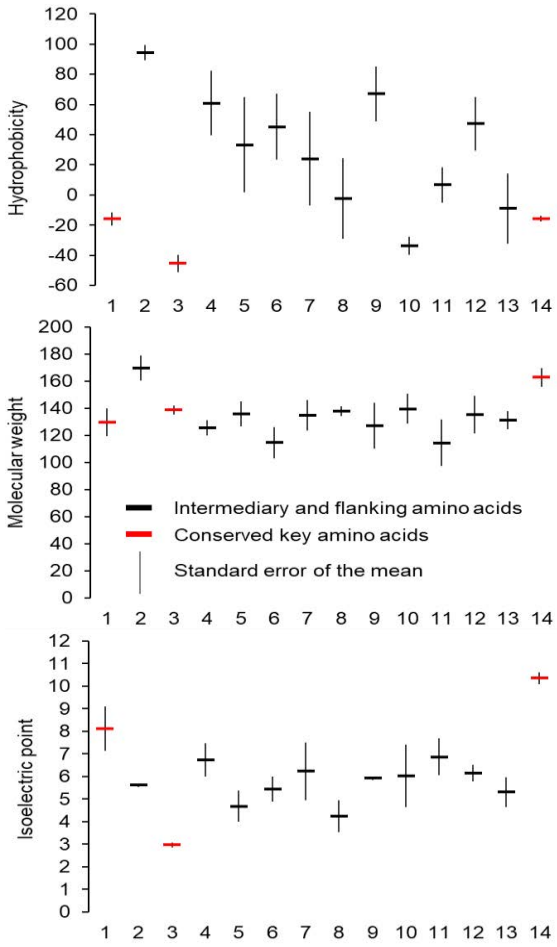

Figure 6: Mean physicochemical values of amino acids at each position of the 14-amino acid long AC motif calculated from experimentally validated ACC population.

An ACC mean value is also generated by averaging the three properties. The ACC values are color-coded where highly recommended hits should contain at least 2 green physicochemical values in addition to green ACC mean and contain no red values. Based on currently available experimental data, cut-off ACC values are determined where the values higher than the upper cut-off limit are colored green (high confidence) and those below the lower cut-off limit are colored red (low confidence). The ACC values are currently set at upper/lower limits of 0.700/0.500 for all physicochemical properties considered (i.e., hydrophobicity, Molecular Weight, Isoelectric point) as well as for the ACC Mean. ACPred also presents an option for metal ion binding which is typically afforded by negatively-charged amino acids at 0-2 positions downstream of the ACC [12]. ACPred is available at http://gcpred.com/acpred without registration or license.

The ACPred webserver is to our knowledge, currently the only tool that can rapidly identify ACs and importantly, it provides statistical values in the form of ACC values which allow user to order the retrieved hits. The latter feature is especially useful in high-throughput applications where ACC values can serve as a reliable indicator of confidence. As such, ACPred provides an added layer of confidence and a way of ranking retrieved hits in the form of scaled color-coded 0-1 ACC values. Previously, the AC motif used for the discovery of AC centers have identified several candidates from the proteome of *Arabidopsis thaliana* but there is no way of ranking them [12]. Using this server, we can now rank selected candidate ACs based on ACC values generated from algorithm that considers the physicochemical properties of intermediary amino acids (Table 1). We demonstrated the utility of this server on the AC candidates reported by [12] and they all contain hits of high confidence (Table 1). However, we note that the predictive strength of ACPred may be weaker in comparison to its parent webserver GCPred (used for the prediction of GC centers). This is due to the fact that experimentally validated AC centers are more varied in terms of their physicochemical properties which we have suggested to be an intrinsic nature of AC centers to enable more optimal binding to the ATP substrate. In addition, AC centers have only been recently identified and currently lack detailed characterization e.g., they are lacking mutational, structural and biochemical analyses that might reveal their substrate binding and inter-domain regulatory mechanisms. As such, we have decided to introduce a relatively low cut-off points to provide less stringent prediction conditions. We will continuously refine the ACC cut-off values to improve the predictive strength and reliability as more experimental data surface and also extend its service to predict other modulatory sites in the near future [49].

Table 1: Testing of Arabidopsis AC centers on ACPred

Name; TAIR ID	Position	ACC Hp	ACC MW	ACC Ip	ACC Mean
*AtClAP; At1g68110	329-342	0.646	0.780	0.724	0.716
*AtPPR; At1g62590	485-498	0.706	0.731	0.765	0.734
*AtKUP7; At5g09400	80-93	0.801	0.810	0.904	0.839
At1g25240	324-337	0.562	0.782	0.821	0.722
At2g34780	271-284	0.743	0.780	0.650	0.725
At3g02930	149-162	0.815	0.762	0.851	0.809
At3g04220	62-75	0.537	0.735	0.743	0.672
At3g18035	382-395	0.684	0.853	0.735	0.757
At3g28223	276-289	0.637	0.839	0.847	0.774
At4g39756	250-263	0.718	0.747	0.793	0.753

Note: These are candidate AC centers reported by [12] and * indicates experimentally confirmed ACCs. Hp: Hydrophobicity; MW: Molecular weight; Ip: Isoelectric point.

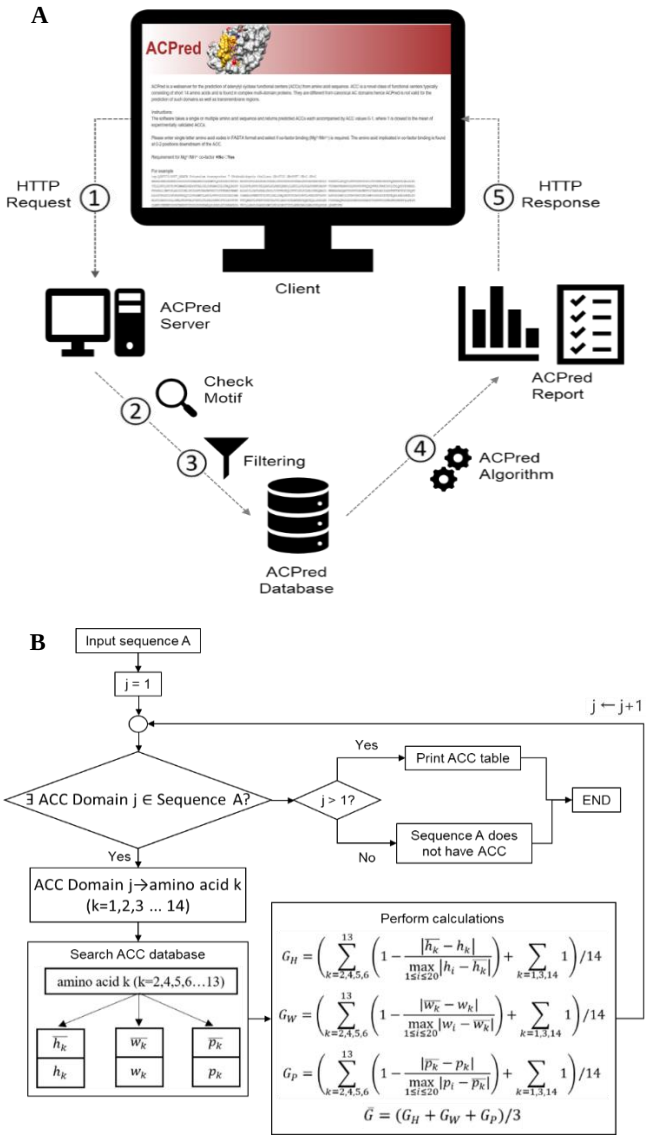

Figure 7: ACPred webserver workflow and algorithm.

4 CONCLUSION

In summary, we have conducted bioinformatic analysis on both GC and AC centers specifically probing their physicochemical properties to determine if there is any molecular basis for GTP and ATP substrate discrimination. Based on our analysis, we have linked the higher variation in charge and hydrophobicity within AC centers in addition to previously assigned [DE] amino acid at position 3 of the motif, for preferential ATP binding. We have also presented the development of a new webtool ACPred, that can rapidly predict candidate AC centers with novel ACC values that enable ranking of retrieved hits. We believe that the understanding of the nature of these new class of catalytic centers (GCs and ACs) have enabled the creation of predictive tools such as ACPred which will in turn, facilitate the discovery of novel cellular components across different biological systems.

ACKNOWLEDGMENT

This research is supported by the Student-partnering-with-Faculty (SpF) program (Grant Number: WKU201718009 awarded to AW) of Wenzhou-Kean University, Zhejiang China. The authors would like to thank Prof. Chris Gehring (University of Perugia, Italy) for testing the server and providing valuable feedback.

REFERENCES

[1] S. Gilroy, and A. Trewavas, 2001. Signal processing and transduction in plant cells: the end of the beginning? *Nature Reviews Molecular Cell Biology* 2, 4, 307-314.

[2] A. Trewavas, 2002. Plant intelligence: Mindless mastery. *Nature* 415, 6874, 841-841.

[3] C. M. Prasch, 2015. Signaling events in plants: Stress factors in combination change the picture. *Environmental and Experimental Botany* 114, 4-14.

[4] J. D. Jones, and J. L. Dangl, 2006. The plant immune system. *Nature* 444, 323-329.

[5] S. E. Clark, 2001. Cell signalling at the shoot meristem. *Nature Reviews Molecular Cell Biology* 2, 4, 276-284.

[6] A. Santner, and M. Estelle, 2009. Recent advances and emerging trends in plant hormone signalling. *Nature* 459, 7250, 1071-1078.

[7] D. R. McCarty, and J. Chory, 2000. Conservation and Innovation in Plant Signaling Pathways. *Cell* 103, 2, 201-209.

[8] N. Pauly, M. R. Knight, P. Thuleau, A. H. van der Luit, and others. 2000. Cell signalling: Control of free calcium in plant cell nuclei. *Nature* 405, 6788, 754-755.

[9] S. Meier, L. Madeo, L. Ederli, L. Donaldson, and others. 2009. Deciphering cGMP signatures and cGMP-dependent pathways in plant defence. *Plant Signaling & Behavior* 4, 4, 307-309.

[10] F. Lemtiri-Chlieh, L. Thomas, C. Marondedze, H. Irving, and others. 2011. *Cyclic Nucleotides and Nucleotide Cyclases in Plant Stress Responses*. InTech, Rijeka, Croatia.

[11] L. Ederli, S. Meier, A. Borgogni, L. Reale, and others. 2008. cGMP in ozone and NO dependent responses. *Plant Signaling & Behavior* 3, 1, 36-37.

[12] C. Gehring, 2010. Adenyl cyclases and cAMP in plant signaling - past and present. *Cell Communication and Signaling* 8, 15.

[13] C. Gehring, and I. S. Turek, 2017. Cyclic nucleotide monophosphates and their cyclases in plant signaling. *Frontiers in Plant Science* 8, 1074.

[14] D. I. Lee, G. Zhu, T. Sasaki, G.-S. Cho, and others. 2015. Phosphodiesterase 9A controls nitric-oxide-independent cGMP and hypertrophic heart disease. *Nature* 519, 472.

[15] C. Marondedze, A. Wong, L. Thomas, H. Irving, and others. 2017. *Cyclic Nucleotide Monophosphates in Plants and Plant Signaling.* Springer International Publishing, Cham (ZG), Switzerland.

[16] A. Moutinho, P. J. Hussey, A. J. Trewavas, and R. Malho. 2001. cAMP acts as a second messenger in pollen tube growth and reorientation. *Proceedings of the National Academy of Sciences of the United States of America* 98, 18, 10481-10486.

[17] L. Thomas, C. Marondedze, L. Ederli, S. Pasqualini, and others. 2013 Proteomic signatures implicate cAMP in light and temperature responses in Arabidopsis thaliana. *Journal of Proteomics* 83, 47-59.

[18] N. Curvetto. 1994. Effect of two cAMP analogs on stomatal opening in Vicia faba : possible relationship with cytosolic calcium concentration. *Plant Physiology and Biochemistry* 32, 365-372.

[19] F. Lemtiri-Chlieh, and G. A. Berkowitz. 2004. Cyclic adenosine monophosphate regulates calcium channels in the plasma membrane of Arabidopsis leaf guard and mesophyll cells. *Journal of Biological Chemistry* 279, 34, 35306-35312.

[20] A. J. Trewavas. 1997. Plant cyclic AMP comes in from the cold. *Nature* 390, 6661, 657-658.

[21] A. Trewavas. 2002. Plant cell signal transduction: the emerging phenotype. *Plant Cell* 14, Suppl, S3-S4.

[22] S. H. Spoel, and X. Dong. 2012. How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews Immunology* 12, 2, 89-100.

[23] A. Wong, and C. Gehring. 2013. The Arabidopsis thaliana proteome harbors undiscovered multi-domain molecules with functional guanylyl cyclase catalytic centers. *Cell Communication and Signaling* 11, 48.

[24] A. Wong, and C. Gehring. 2013. *Computational identification of candidate nucleotide cyclases in higher plants.* Methods in Molecular Biology 1016, 195-205.

[25] A. Wong, C. Gehring, and H. R. Irving. 2015. Conserved functional motifs and homology modelling to predict hidden moonlighting functional sites. *Frontiers in Bioengineering and Biotechnology* 3, 82.

[26] A. Wong, X. Tian, C. Gehring, and C. Marondedze. 2018. Discovery of novel functional centers with rationally designed amino acid motifs. *Computational and Structural Biotechnology Journal* 16, 70-76.

[27] L. Freihat, V. Muleya, D. T. Manallack, J. I. Wheeler, and others. 2014. Comparison of moonlighting guanylate cyclases: roles in signal direction? *Biochemical Society Transactions* 42, 6, 1773-1779.

[28] C. J. Jeffery. 1999. Moonlighting proteins. *Trends in Biochemical Sciences* 24, 1, 8-11.

[29] C. J. Jeffery. 2015. Why study moonlighting proteins? *Frontiers in Genetics* 6, 211.

[30] V. Muleya, J. I. Wheeler, O. Ruzvidzo, L. Freihat, and others. 2014. Calcium is the switch in the moonlighting dual function of the ligand-activated receptor kinase phytosulfokine receptor 1. *Cell Communication and Signaling* 12, 60.

[31] P. Tompa, C. Szász, and L. Buday. 2005. Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences* 30, 9, 484-489.

[32] H. R. Irving, D. M. Cahill, and C. Gehring. 2018. Moonlighting proteins and their role in the control of signaling microenvironments, as exemplified by cGMP and phytosulfokine receptor 1 (PSKR1). *Frontiers in Plant Science* 9, 415.

[33] N. Xu, D. Fu, S. Li, Y. Wang, and others. 2018. GCPred: a web tool for guanylyl cyclase functional centre prediction from amino acid sequence. *Bioinformatics* 34, 12, 2134-2135.

[34] S. Meier, C. Seoighe, L. Kwezi, H. Irving, and others. 2007. Plant nucleotide cyclases: an increasingly complex and growing family. *Plant Signaling & Behavior* 2, 6, 536-539.

[35] N. Ludidi, and C. Gehring. 2003. Identification of a novel protein with guanylyl cyclase activity in Arabidopsis thaliana. *Journal of Biological Chemistry* 278, 8, 6490-6494.

[36] L. Kwezi, S. Meier, L. Mungur, O. Ruzvidzo, and others. 2007. The Arabidopsis thaliana brassinosteroid receptor (AtBRI1) contains a domain that functions as a guanylyl cyclase in vitro. *PloS ONE* 2, 5.

[37] L. Kwezi, O. Ruzvidzo, J. I. Wheeler, K. Govender, and others. 2011. The phytosulfokine (PSK) receptor is capable of guanylate cyclase activity and enabling cyclic GMP-dependent signaling in plants. *Journal of Biological Chemistry* 286, 25, 22580-22588.

[38] T. Mulaudzi, N. Ludidi, O. Ruzvidzo, M. Morse, and others. 2011. Identification of a novel Arabidopsis thaliana nitric oxide-binding molecule with guanylate cyclase activity in vitro. *FEBS Letters* 585, 17, 2693-2697.

[39] S. Meier, O. Ruzvidzo, M. Morse, L. Donaldson, and others. 2010. The Arabidopsis wall associated kinase-like 10 gene encodes a functional guanylyl cyclase and is co-expressed with pathogen defense related genes. *PloS ONE* 5, 1, e8904.

[40] Z. Qi, R. Verma, C. Gehring, Y. Yamaguchi, and others. 2010. Ca2+ signaling by plant Arabidopsis thaliana Pep peptides depends on AtPepR1, a receptor with guanylyl cyclase activity, and cGMP-activated Ca2+ channels. *Proceedings of the National Academy of Sciences of the United States of America* 107, 49, 21193-21198.

[41] I. Turek, and C. Gehring. 2016. The plant natriuretic peptide receptor is a guanylyl cyclase and enables cGMP-dependent signaling. *Plant Molecular Biology* 91, 3, 275-286.

[42] A. Szmidt-Jaworska, K. Jaworski, A. Pawelek, and J. Kocewicz. 2009. Molecular cloning and characterization of a guanylyl cyclase, PNGC-1, involved in light signaling in Pharbitis nil. *Journal of Plant Growth Regulation* 28, 367-380.

[43] B. Swiezawska, K. Jaworski, M. Duszyn, A. Pawelek, and others. 2017. The Hippeastrum hybridum PepR1 gene (HpPepR1) encodes a functional guanylyl cyclase and is involved in early response to fungal infection. *Journal of Plant Physiology* 216, 100-107.

[44] I. Al-Younis, A. Wong, and C. Gehring. 2015. The Arabidopsis thaliana K(+)-uptake permease 7 (AtKUP7) contains a functional cytosolic adenylate cyclase catalytic centre. *FEBS Letters* 589, 24 Pt B, 3848-3852.

[45] J. I., Wheeler, A. Wong, C. Marondedze, A. J. Groen, and others. 2017. The brassinosteroid receptor BRI1 can generate cGMP enabling cGMP-dependent downstream signaling. *Plant Journal* 91, 4, 590-600.

[46] P. Chatukuta, T. Dikobe, D. Kawadza, K. Sehlabane, and others. 2018. An Arabidopsis clathrin assembly protein with a predicted role in plant defense can function as an adenylate cyclase. *Biomolecules* 8, 2, E15.

[47] O. Ruzvidzo, B. T. Dikobe, D. T. Kawadza, G. H. Mabadahanye, and others. 2013. *Recombinant Expression and Functional Testing of Candidate Adenylate Cyclase Domains.* Humana Press, New York City, USA.

[48] C. L. Tucker, J. H. Hurley, T. R. Miller, and Hurley, J. B. 1998. Two amino acid substitutions convert a guanylyl cyclase, RetGC-1, into an adenylyl cyclase. *Proceedings of the National Academy of Sciences of the United States of America* 95, 11, 5993-5997.

[49] A. Ooi, F. Lemtiri-Chlieh, A. Wong, and C. Gehring. 2017. Direct modulation of the guard cell outward-rectifying potassium channel (GORK) by abscisic acid. *Molecular Plant* 10, 11, 1469-1472.