OXFORD

Sequence analysis

GCPred: a web tool for guanylyl cyclase functional centre prediction from amino acid sequence

Nuo Xu[1], Dongfang Fu[1], Shiang Li[1], Yuxuan Wang[2] and Aloysius Wong[3],*

[1]Department of Computer Science and [2]Department of Psychology and [3]Department of Biology, Wenzhou-Kean University, Ouhai, Wenzhou, Zhejiang Province 325060, China

*To whom correspondence should be addressed.
Associate Editor: John Hancock

Abstract

Summary: GCPred is a webserver for the prediction of guanylyl cyclase (GC) functional centres from amino acid sequence. GCs are enzymes that generate the signalling molecule cyclic guanosine 3′, 5′-monophosphate from guanosine-5′-triphosphate. A novel class of GC centres (GCCs) has been identified in complex plant proteins. Using currently available experimental data, GCPred is created to automate and facilitate the identification of similar GCCs. The server features GCC values that consider in its calculation, the physicochemical properties of amino acids constituting the GCC and the conserved amino acids within the centre. From user input amino acid sequence, the server returns a table of GCC values and graphs depicting deviations from mean values. The utility of this server is demonstrated using plant proteins and the human interleukin-1 receptor-associated kinase family of proteins as example.

Availability and implementation: The GCPred server is available at http://gcpred.com.

Contact: alwong@kean.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Guanylyl cyclases (GCs) are enzymes that generate the signalling molecule cyclic guanosine 3′, 5′-monophosphate (cGMP) from guanosine-5′-triphosphate (GTP). It is well-documented that cGMP has key signalling roles in bacteria, plants and animals mediating cellular and physiological responses including ion homeostasis, hormones and peptides signalling as well as responses to abiotic and biotic stresses (Gehring and Turek, 2017; Marondedze *et al.*, 2017). However, canonical GC domains were undetected in higher plants despite overwhelming evidence for cGMP detection in plant tissues (Isner and Maathuis, 2011) and cGMP-mediated plant responses (Hussain *et al.*, 2016; Joudoi *et al.*, 2013). A novel class of GC centres (GCCs) have been identified in complex plant proteins using a motif-based approach described elsewhere (Ludidi and Gehring, 2003; Meier *et al.*, 2007; Wong *et al.*, 2015). Supported by homology modelling and ligand docking simulations, this approach

identified several GCCs in plants including those embedded within larger domains of well-characterized hormone receptor complexes. They are therefore classified as functional centres and have been shown to intricately regulate signalling networks (Meier *et al.*, 2007; Muleya *et al.*, 2014; Wheeler *et al.*, 2017) with roles observed at the molecular and biological levels (Gehring and Turek, 2017). Given the biological significance of cGMP in different systems (Hartwig *et al.*, 2014), the identification of GCCs requires automation in the form of a webserver. GCPred is created to enable rapid prediction of GCCs from amino acid sequences and importantly, provides a way to rank hits based on algorithm that considers the conserved amino acids with assigned catalytic functions and the physicochemical properties of amino acids at the GCCs. The latter consideration is made due to structural reasons since GCCs contain 'helical-loop' signatures with suitably charged amino acids solvent-exposed and orientated for interactions with GTP (Wong *et al.*, 2015).

2 The GCPred server

GCPred is available at http://gcpred.com without registration or licence. The server allows user to input single or multiple amino acid sequences in FASTA format and calculates predicted GCCs based on a set of physicochemical properties of amino acids in experimentally validated GCCs. The server returns a result page containing a table of predicted GCCs accompanied by their respective GCC values colour-coded to aid interpretations, and graphs depicting deviations from mean values (Supplementary Fig. S1). Previous works have ascertained that this unique class of GCC typically contains 14 amino acids where the amino acids at positions 1, 3 and 14 have direct substrate binding and catalytic functions (Ludidi and Gehring, 2003). They are different from the canonical GCs found in bacteria and animals (Schaap, 2005), and more recently also in plants (Świeżawska et al., 2015; Yuan et al., 2008), thus GCPred is not valid for the prediction of this class of GCs. The server will first screen sequences for conserved key amino acids at their respective positions 1, 3 and 14. If present, the server will calculate three properties of each amino acid at positions 2 and 4–13 namely hydrophobicity, molecular weight and isoelectric point as well as an overall mean value. These scores are known as GCC values and scaled 0 to 1 for each hit; where 1 is closest to values of experimentally validated GCCs (see GCPred algorithm in Supplementary Fig. S2). Based on currently available experimental data, cut-off GCC values are determined where the values higher than the upper cut-off limit are coloured green (high confidence) and those below the lower cut-off limit are coloured red (low confidence). Based on the testing of experimentally validated GCCs (Supplementary Table S1), we recommend user to select hits that contain two or more green physicochemical values in addition to green GCC mean values and no red GCC values. GCPred also presents an option for metal ion binding which is typically afforded by negatively-charged amino acids at 0–2 positions downstream of the GCC (Ludidi and Gehring, 2003).

3 Discussion

The strength of this server lies in its ability to rapidly identify GCC candidates based not only on conserved amino acids, but also the properties of intermediary and flanking amino acids. This provides another layer of confidence and a way of ranking retrieved hits in the form of scaled colour-coded 0–1 GCC values. The user can also take into consideration the deviations of amino acids at each position from the population average by analyzing the accompanying graphs. Previously, this GC motif identified >40 candidate GCCs in the proteome of Arabidopsis thaliana but there is no way of ranking them (Wong and Gehring, 2013). Using this server, we can now rank selected candidate GCs based on GCC values generated from algorithm that considers the physicochemical properties of intermediary amino acids (Supplementary Table S2). We demonstrated the utility of this server on functional plant proteins PnGC1 and HpPepR1 (Świeżawska et al., 2017; Szmidt-Jaworska et al., 2009) and in both instances, GCPred returned hits of high confidence (Supplementary Table S3). We also demonstrated the utility of GCPred with the human interleukin-1 receptor-associated kinase (IRAK) family of proteins. The GCPred server returned hits in all four IRAK proteins but only IRAK1 and IRAK3 harbour GCCs of high confidence (Supplementary Fig. S3). The GCC values of IRAK3 are above the population average and this is consistent with previous report of IRAK3 being capable of cGMP generation in vitro (Freihat et al., 2014). We note that this GCC is also present in orthologues thus suggesting a common role in mammals (Supplementary Fig. S3). In addition to continuously refining the GCC cut-offs based on new experimental data, the server will also incorporate the option for additional filters and extend its service to predict other functional centres such as adenylyl cyclases and modulatory sites (Al-Younis et al., 2015; Ooi et al., 2017).

References

Al-Younis,I. et al. (2015) The Arabidopsis thaliana K(+)-uptake permease 7 (AtKUP7) contains a functional cytosolic adenylate cyclase catalytic centre. FEBS Lett., 589, 3848–3852.

Freihat,L. et al. (2014) Comparison of moonlighting guanylate cyclases: roles in signal direction? Biochem. Soc. Trans., 42, 1773–1779.

Gehring,C. and Turek,I.S. (2017) Cyclic nucleotide monophosphates and their cyclases in plant signaling. Front. Plant Sci., 8, 1704.

Hartwig,C. et al. (2014) cAMP, cGMP, cCMP and cUMP concentrations across the tree of life: high cCMP and cUMP levels in astrocytes. Neurosci. Lett., 579, 183–187.

Hussain,J. et al. (2016) Constitutive cyclic GMP accumulation in Arabidopsis thaliana compromises systemic acquired resistance induced by an avirulent pathogen by modulating local signals. Sci. Rep., 6, 36423.

Isner,J.C. and Maathuis,F.J. (2011) Measurement of cellular cGMP in plant cells and tissues using the endogenous fluorescent reporter FlincG. Plant J., 65, 329–334.

Joudoi,T. et al. (2013) Nitrated cyclic GMP modulates guard cell signaling in Arabidopsis. Plant Cell, 25, 558–571.

Ludidi,N. and Gehring,C. (2003) Identification of a novel protein with guanylyl cyclase activity in Arabidopsis thaliana. J. Biol. Chem., 278, 6490–6494.

Marondedze,C. et al. (2017) Cyclic nucleotide monophosphates in plants and plant signaling. Handb. Exp. Pharmacol., 238, 87–103.

Meier,S. et al. (2007) Plant nucleotide cyclases: an increasingly complex and growing family. Plant. Signal Behav., 2, 536–539.

Muleya,V. et al. (2014) Calcium is the switch in the moonlighting dual function of the ligand-activated receptor kinase phytosulfokine receptor 1. Cell Commun. Signal, 12, 60.

Ooi,A. et al. (2017) Direct modulation of the guard cell outward-rectifying potassium channel (GORK) by abscisic acid. Mol. Plant, 10, 1469–1472.

Schaap,P. (2005) Guanylyl cyclases across the tree of life. Front. Biosci., 10, 1485–1498.

Szmidt-Jaworska,A. et al. (2009) Molecular cloning and characterization of a guanylyl cyclase, PnGC-1, involved in light signaling in Pharbitis nil. J. Plant Growth Regul., 28, 367–380.

Świeżawska,B. et al. (2015) Identification of a Hippeastrum hybridum guanylyl cyclase responsive to wounding and pathogen infection. J. Plant Physiol., 189, 77–86.

Świeżawska,B. et al. (2017) The Hippeastrum hybridum PepR1 gene (HpPepR1) encodes a functional guanylyl cyclase and is involved in early response to fungal infection. J. Plant. Physiol., 216, 100–107.

Wheeler,J.I. et al. (2017) The brassinosteroid receptor BRI1 can generate cGMP enabling cGMP-dependent downstream signaling. Plant J., 91, 590–600.

Wong,A. and Gehring,C. (2013) The Arabidopsis thaliana proteome harbors undiscovered multi-domain molecules with functional guanylyl cyclase catalytic centers. Cell Commun. Signal, 11, 48.

Wong,A. et al. (2015) Conserved functional motifs and homology modeling to predict hidden moonlighting functional sites. Front. Bioeng. Biotechnol., 3, 82.

Yuan,J. et al. (2008) A guanylyl cyclase-like gene is associated with Gibberella ear rot resistance in maize (Zea mays L.). Theor. Appl. Genet., 116, 465–479.