Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Shawn Walsh

April 15, 2015

Investigating Nucleotide Interactions of BdcA, a Rossmann-fold containing protein capable of biofilm dispersal

By

Shawn Walsh

Dr. Emily Weinert Adviser

Department of Chemistry

Dr. Emily Weinert

Adviser

Dr. Stefan Lutz

Committee Member

Dr. Jose Soria

Committee Member

Dr. Leah Roesch

Committee Member

2015

Investigating Nucleotide Interactions of BdcA, a Rossmann-fold containing protein capable of biofilm dispersal

By

Shawn Walsh

Dr. Emily Weinert Adviser

An abstract of a thesis submitted to the Faculty of Emory College of Arts and Sciences of Emory University in partial fulfillment of the requirements of the degree of Bachelor of Sciences with Honors

Department of Chemistry

2015

Abstract

Investigating Nucleotide Interactions of BdcA, a Rossmann-fold containing protein capable of biofilm dispersal

By Shawn Walsh

Biofilms are a major problem in human health because they confer increased resistance to antimicrobial agents and adhere to medical devices. As a potential therapeutic, the protein BdcA was previously engineered to disperse biofilms, and proposed to function via the sequestering of cyclic dimeric GMP (c-di-GMP).¹ This ubiquitous bacterial second messenger is known to control a variety of cellular processes including biofilm formation and dispersal. Unlike previously characterized c-di-GMP receptors, BdcA contains a Rossmann fold, which typically binds NAD(P)(H). We set out to characterize BdcA as the first of a potentially novel class of c-di-GMP receptors. However, we show that BdcA has no affinity for c-di-GMP. Rather, it binds nicotinamide adenine dinucleotide phosphate (NADPH). We have also created seven rational mutants of BdcA, and initiated isothermal titration calorimetry (ITC) and biofilm dispersal assays in order to correlate NADPH interactions with biofilm dispersal, thereby shedding light on the now enigmatic mechanism behind BdcA-mediated biofilm dispersal.

Investigating Nucleotide Interactions of BdcA, a Rossmann-fold containing protein capable of biofilm dispersal

By

Shawn Walsh

Dr. Emily Weinert Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences of Emory University in partial fulfillment of the requirements of the degree of Bachelor of Sciences with Honors

Department of Chemistry

2015

Table of Contents

Intro	luction	1
Resul	ts and Discussion	4
1.	Work based on c-di-GMP as the substrate for BdcA	4
2.	Work based on NADPH-specific binding	11
Concl	usion	18
Metho	ods	19
1.	Genomic Isolation and Cloning	19
2.	Site-directed Mutagenesis	19
3.	Protein Expression and Purification	20
4.	Phosphodiesterase Activity Assay	20
5.	HPLC Analysis of nucleotides	21
6.	Sequence and Structure Alignments	22
7.	Rapid Equilibrium Dialysis (RED)	22
8.	Isothermal Titration Calorimetry	24
9.	Biofilm Dispersal Assay	25
Work	s Cited	26
Suppl	emental Figures and Tables	28

List of Figures and Tables

Figure 1	Nucleotide Structures	2
Figure 2	3-D BdcA Model	3
Figure 3	BdcA Overlaid with 30P4	6
Figure 4	Phosphodiesterase Assay Chromatograms	8
Figure 5	C-di-GMP Chromatograms in Various Mobile Phases	9
Figure 6	ITC Results Using c-di-GMP	10
Figure 7	BdcA Overlaid with FabG	12
Figure 8	Mutant Summary	13
Figure 9	ITC Results Using NADPH	14
Figure 10	Biofilm Dispersal Assay	17
Figure S1	Sequence Alignment with DALI Hits	28
Figure S2	Sequence Alignment with 3SJ7 and 30P4	29
Figure S3	Protein Purification Gel	30
Table S1	Primers for Cloning and Mutagenesis	31
Figure S4	Gel Verification of Cloning	32

Introduction

The vast majority of bacteria in nature live and propagate in complex, multi-species communities attached to a solid surface²; these bacterial communities are termed biofilms. Characteristic features of biofilms are radically altered gene expression in comparison to planktonic cells and the production of extracellular polymers that encase the bacteria within a protective barrier. Biofilms are notoriously hard to eliminate, and cause problems in a wide range of settings from the fouling of ship hulls to computer-chip malfunction.^{3,4} However, the most alarming consequences of biofilms are those related to human health. Biofilms confer resistance to physical and chemical forces including disinfectants, immune system response, and antibiotics.^{5,6} Consequentially, they are involved in 80% of all human infections.⁷ They are especially problematic in hospital settings, in which they can contaminate vents and medical devices⁸, and even increase the chance of horizontal gene transfer and the resulting spread of antibiotic resistance.⁹

The protein BdcA (biofilm dispersal via c-di-GMP) from *E. coli* was recently discovered as a tool for eliminating biofilms.¹ Subsequent engineering of this protein through site-saturated mutagenesis produced a mutant that showed improved biofilm dispersal capabilities. Expression of BdcA with this single mutation, E50Q, almost completely disperses biofilms in *E coli*. In addition, *bdcA* also shows high conservation among genes in a number of species. When expressed in other gram-negative bacteria, such as *Psuedomonas aeruginosa, P. fluorescens,* and *Rhizobium meliloti*, BdcA retains its ability to disperse biofilms.¹⁰ These findings highlight the potential of BdcA alone to effectively eliminate biofilms formed by a variety of gram-negative bacteria.

1

BdcA was initially proposed to eliminate biofilms through the sequestering and consequent reduction of the effective concentration of cyclic diguanylate (c-di-GMP) (Figure 1A), which is a ubiquitous small-molecule regulator of biofilms in bacteria. Binding affinity was measured by incubating c-di-GMP with BdcA, separating free c-di-GMP and BdcA-bound c-di-GMP via spin filtering, recovering bound c-di-GMP using a trypsin digest, and finally quantifying c-di-GMP with HPLC. These binding experiments established the binding affinity for BdcA for c-di-GMP as 11.7 μ M and showed that the E50Q mutation, which disperses biofilms better than wild-type BdcA, had a higher affinity for c-di-GMP, further supporting the correlation between c-di-GMP sequestering and biofilm dispersal.¹





Several residues that could be responsible for c-di-GMP binding were identified (Figure 2). Specifically, BdcA contains an EAL sequence, which is necessary within phosphodiesterases that break down c-di-GMP into 5'-pGpG. In addition, some proteins with a catalytically inactive EAL domain function as c-di-GMP receptors. However, it seemed unlikely that this EAL sequence was part of a biologically relevant EAL domain. The motif was found on the exterior of the protein rather than the core, which is inconsistent with other EAL phosphodiesterase structures.¹¹ Other potentially relevant residues identified in previous work were based on conserved motifs of c-di-GMP binding.^{1,12}



Figure 2. 3-D model of BdcA. The EAL sequence, E50, and other potentially important residues identified are highlighted in orange, blue, and yellow respectively.

Surprisingly, subsequent work from the same group identified BdcA as a short-chain dehydrogenase/reductase (SDR) that binds nicotinamide adenine dinucleotide phosphate (NADPH) (Figure 1B) specifically, rather than c-di-GMP (the authors could not detect c-di-GMP binding using isothermal titration calorimetry (ITC)), which discredited the established hypothesis.¹³ As BdcA binding to c-di-GMP has now been refuted, the mechanism of BdcA-dependent biofilm dispersal has become an enigma. There are surprisingly few examples of SDRs that have been shown to influence any kind of quorum sensing phenotypes.¹⁴ Dissecting correlations between NADPH dependence of BdcA on biofilm dispersal would provide valuable insights into its mechanism and potentially validate a link between these two disparate fields.

In this work, we initially created seven mutants to investigate c-di-GMP binding to BdcA. This protein contain a Rossmann fold, which was entirely new in a c-di-GMP receptor and suggested that BdcA might employ a previously uncharacterized mechanism of c-diGMP binding. The Rossmann fold is the common motif of SDRs that typically binds to NAD(P)(H) and, although the nearly 47,000 distinct SDRs known show low sequence similarity¹⁴, the Rossmann fold can be identified by its characteristic structure; a central beta sheet with 6-7 strands flanked on each side by 3-4 alpha-helices.¹⁵ The Rossmann fold contains an integral glycine-rich region for binding the pyrophosphate moiety of NADPH. A conserved Asn-Ser-Tyr-Lys tetrad is responsible for oxidoreductase activity. Following the publication of the second paper refuting c-di-GMP binding, the panel of mutants has been used to investigate determinants of NADPH binding and anti-biofilm activity.

Our results validate NADPH binding to BdcA and the lack of c-di-GMP affinity. Furthermore, we hope to shed light on the mechanism through which BdcA exerts its remarkable ability to disperse biofilms, which until recently had a seemingly simple explanation. The rational mutations that were incorporated within BdcA are in prime positions to establish the connection between NADPH binding affinity, potential oxidoreductase ability, and biofilm dispersal.

Results and Discussion

1. Work based on c-di-GMP as the substrate for BdcA

Initially, we focused on characterizing BdcA as a novel receptor for c-di-GMP due to the report of BdcA as a novel c-di-GMP binding protein.¹ Although several conserved binding domains of c-di-GMP are well known, the search for novel effectors that may help explain the complexity of c-di-GMP signaling is ongoing. Prior to work showing that BdcA bound c-di-GMP,¹ there were no examples of c-di-GMP receptors that contained a Rossmann fold. We reasoned that BdcA might be the first identified protein within a novel class of c-di-GMP receptors, and that by understanding how c-di-GMP bound to BdcA we could identify other new receptors with a similar binding domain.

Towards this end, we performed a homology search to identify potential homologs based on structure. We choose to focus on structural homologs, as opposed to identifying proteins based on sequence similarity, for two reasons. First, proteins containing a Rossmann fold are notorious for exhibiting low sequence similarity (typically 20-30%)¹⁶, whereas they can be quite structurally similar.¹⁴ Second, validating c-di-GMP binding in several proteins with structural similarity to BdcA but a low sequence homology would provide valuable insight into the essential residues of the c-di-GMP binding domain involved.

Lacking a crystal structure of BdcA, we created a 3-D model using an online structure homology-modeling server, SWISS-MODEL, which automatically identified a dehydrogenase/reductase from *Sinorhizobium meliloti* 1021 (PBD ID: 3V2G) as a template.^{17–20} We then used this model to search for structural homologs on the Protein Data Bank (PDB) using the Dali Network Service.²¹ A sequence alignment with BdcA for the top results of this search is shown in Figure S1. Based on these results, it seemed unlikely that we would be able to identify promising candidates to test for c-di-GMP affinity because none of the hits contained the EAL sequence motif or the small collection of residues that the previous group to work with BdcA had identified as possibly being involved in c-di-GMP binding. Specifically, these residues were D136, D180, Q49, and E220.¹

Without clear c-di-GMP binding homologs that could be expressed for direct comparison to BdcA, we decided to begin probing c-di-GMP binding through the use of mutagenesis. We created the mutations E90A, D180A, and D136T to investigate whether

5

these residues were really involved in c-di-GMP binding. In addition, we recreated the E50Q mutant, which had been previously shown to affect c-di-GMP binding and biofilm dispersal.¹ Departing from previous work, it seemed that the residues within the Rossmann fold that had known roles in NAD(P)(H) binding were also attractive candidates for mutagenesis. NADH has a structure similar to c-di-GMP (Figure 1), especially in that they both contain a purine-sugar-phosphate moiety. This suggested that c-di-GMP might bind within the Rossmann fold itself. With this possibility in mind, we aligned BdcA with the crystal structure of a 3-ketoacyl reductase (PBD ID: 30P4) in complex with NADP+, which DALI had identified as a potential homolog (Figure 3).

B

А

Figure 3. (A) BdcA (green) overlaid with 3OP4 (cyan). (B) Close-up of N63 (yellow) on 3OP4 and D59 (magenta) on BdcA next to the purine of NADP+

In 3OP4, the Asparagine at position 63 (Figure 3B, magenta) is involved in hydrogen bonding to the amine group of the purine ring on NADP+. When aligned with 3OP4, BdcA shows an aspartic acid (Figure 3B, yellow) in roughly the same location, a residue that is also capable of hydrogen bonding. Therefore, we created the mutation D59A based on the rationale that it would disrupt c-di-GMP binding if c-di-GMP bound to BdcA with a similar orientation to NADPH binding within the Rossmann fold. Finally, we created the mutations G14V and K150I. The former is within a well-established glycine-rich region of Rossmann fold proteins that is involved in NADPH binding, and the mutation was designed to disrupt c-di-GMP binding if binding occurred in a manner similar to NADP+ binding to 30P4. The K150I mutation was incorporated because the residue is essential within the catalytic tetrad required for the redox chemistry of Rossmann fold proteins. A representative gel of the protein purification can be found in S3.

Before conducting binding studies of the BdcA mutants, we wanted to verify that BdcA showed no catalytic activity towards c-di-GMP. The protein carries an EAL sequence motif, which is characteristic of phosphodiesterases that cleave c-di-GMP specifically into 5' pGpG. Catalytic activity of EAL proteins depends on coordination of Mg²⁺ or Mn²⁺ ions.²² EAL domains that lack metal coordination sites are catalytically inactive and can function as c-di-GMP effectors.²³

Although previous work reported that BdcA lacked phosphodiesterase activity, these experiments were carried out in solution with EDTA.¹ Therefore, lack of c-di-GMP turnover in these experiments could plausibly have been a result of catalytically inactive protein or simply the sequestering of metal ions required for catalysis. Therefore, the first step in characterization of BdcA was testing the phosphodiesterase activity.

The BdcA gene was cloned from the genomic DNA of *E. coli* K12 MG1655 strain. Cloning of the gene into the pET28a vector was verified by agarose gel electrophoresis (Fig. S3) and sequencing. Protein expression and purification were performed as described in the Methods section; a representative gel from the purification can be found in Fig. S4.

We observed no catalytic activity of BdcA towards c-di-GMP, even in the absence of metal chelators (Figure 4). Unfortunately, the chromatograms from this experiment

7

showed an erratic baseline that made accurate c-di-GMP quantification impossible. It is interesting to note that the chromatograms we obtained resembled those from previous phosphodiesterase activity assays for BdcA.¹ In light of the second study that did not observe c-di-GMP binding, the results based on HPLC chromatograms from the Ma, *et al* paper also may have been artifacts.



Figure 4. Chromatograms taken at 250nm for (A) pGpG standard, (B) c-di-GMP incubated with BdcA under PDE assay conditions, and (C) c-di-GMP standards reveal that no c-di-GMP degradation occurred in the PDE assay. C-di-GMP eluted at 9.34 minutes and pGpG at 8.81 minutes (indicated by arrows).

Despite several attempts to remove apparent contamination from the system, including several methods of washing, several new mobile phase combinations, and a new column, problems with accurately quantifying c-di-GMP persisted (Figure 5). These problems may have arisen in part due to contaminants in a failing HPLC seal.



Figure 5. C-di-GMP chromatograms measured at 250nm under various conditions. (A) Mobile phases of methanol with 0.1% formic acid (buffer A) and water with 0.1% formic acid (buffer B) lead to reduced absorbance for c-di-GMP samples (10.45 minutes) that makes quantification imprecise, especially for low concentrations. (B) Mobile phases of 0.1M KH₂PO₄, pH=6 (buffer A) and 90% buffer A with 10% methanol (buffer B) showed a c-di-GMP peak at 22.13 minutes, but lead to problems with the HPLC. Specifically, pressure spiked and rose above tolerable limits. (C) Mobile phases of 0.1M triethyl amine, pH=6 (buffer A) and 50% buffer A with 50% methanol led to an uneven baseline, although a small peak could be seen for c-di-GMP at 18.6 minutes.

The initial binding studies of mutants using rapid equilibrium dialysis (RED) and analysis via HPLC were inconclusive because of these problems reliably quantifying c-di-GMP. In retrospect, these issues were likely exacerbated due to the lack of c-di-GMP binding by BdcA. However, as the only report in the literature at the time had quantified cdi-GMP binding to BdcA with a K_d of ~12 μ M,¹ the lack of observed c-di-GMP binding was attributed to issues with the HPLC. Therefore, to sidestep potential HPLC issues, c-di-GMP was quantified using UV-Vis analysis following RED. Surprisingly, these experiments showed that BdcA had no affinity for c-di-GMP (data not shown). In order to verify these results, c-di-GMP affinity was also measured via isothermal titration calorimetry (ITC, Figure 6). ITC was chosen because it measures binding directly by quantifying heat of binding, as opposed to techniques that correlate a measurement of ligand concentration to binding affinity, and thereby reduces the potential for other factors of an experimental setup to skew the results.



Figure 6. Raw ITC data for BdcA E50Q titrated with c-di-GMP (A) and buffer (B) are equivalent.

Contrary to previous results, ITC also showed that BdcA lacked affinity for c-di-GMP. Testing under a variety of conditions, including the use of high-salt buffers, produced the same result. Although our efforts were aimed toward characterizing a novel c-di-GMP receptor, our data suggested that BdcA has no affinity for c-di-GMP and is not a novel c-di-GMP receptor.

2. Work based on NADPH-specific binding

A second paper characterizing BdcA recently reported binding to NADPH specifically, with no observable affinity of BdcA for c-di-GMP.¹³ The lack of c-di-GMP affinity corroborated with our own results, and in light of these new findings we were forced to reevaluate our aims. Fortunately, some mutants we created to probe c-di-GMP binding were also relevant to exploring the correlation between NADPH interactions and biofilm dispersal mediated by BdcA. Since c-di-GMP sequestering is no longer a valid hypothesis for BdcA's biofilm dispersal capabilities, we turned our attention towards understanding this phenomenon.

Among the proteins identified by DALI during the bioinformatics analysis, a βketoacetyl-CoA reductase from *Staphylococcus aureus* termed FabG (PBD: 3SJ7) that has been shown to bind to NADPH shares a 32% sequence identity with BdcA.²⁴ Alignment of one asymmetric unit of FabG with the BdcA model shows how NADPH might bind to BdcA (Figure 7A).



Figure 7. BdcA model (green) aligned with FabG asymmetric unit (cyan). NADPH is shown in orange. (A) Overall protein structures align to a high degree. (B) The lysine at position 150 (magenta) in BdcA is in prime position to engage in catalytic activity. Side chains for the rest of the potentially catalytic triad are shown in green

FabG oxidizes NADPH using the characteristic Asn-Ser-Tyr-Lys catalytic tetrad of proteins that contain a Rossman fold.²⁴ The model of BdcA contains these same residues in the same positions when it is structurally aligned with FabG (Figure 7B). The catalytic role of these residues was explored over twenty years ago, and it is well known that mutating either the tyrosine or lysine residues totally abolishes catalytic activity in otherwise active proteins, except for rare substitutions that exhibit a fraction of the original activity (substitution with cysteine or arginine).²⁵ Therefore, the K150I mutation should have no catalytic activity towards NADPH, whether or not WT BdcA does. Any difference in biofilm dispersal between this mutant and WT BdcA would provide evidence as to the role of catalysis in biofilm dispersal. The placement of the other mutations that were generated on BdcA are shown in Figure 8, in which BdcA was aligned with FabG bound to NADPH and then FabG was hidden.



Figure 8. Model of BdcA aligned with FabG (hidden) complexed to NADPH with mutated residues highlighted in magenta

It is interesting to note that the glutamic acid residue at position 50 is extremely distal to the likely NADPH binding pocket, but that mutating this residue to glutamine in the previous Ma, *et al* study dramatically improved BdcA's ability to disperse biofilms.¹ Although E50 may affect protein dynamics or stability, leading to changes in biofilm formation due to altered BdcA-dependent catalysis, it also raises the possibility that BdcA may somehow disperse biofilms independently of its interaction with NADPH.

ITC was used to determine the NADPH affinity of BdcA. Previously, this technique was used to establish a K_d of 25.9±4.1 μ M for WT BdcA.¹³ Our preliminary data suggest a value for NADPH binding that falls within the error range of this literature value. In a single trial, WT BdcA was found to bind NADPH with a K_d of 27.5±2.6 μ M (Figure 9A). Further trials are necessary to determine the quality of this data, but as a qualitative measure it is sufficient to verify NADPH binding.



Figure 9. Binding isotherms for (A) WT BdcA, (B) Bdca E50Q, and (C) BdcA E50Q with reference data from the negative control subtracted. Because it is constructed by subtracting one set of data from another, there is no raw binding isotherm to show.

Although BdcA was engineered to better disperse biofilms by means of an E50Q mutation, it is unknown how this mutation affects cofactor binding. Any differences in NADPH affinity for BdcA E50Q compared to WT BdcA would have implications for the method by which BdcA disperses biofilms. Initials trials to determine BdcA E50Q affinity

for NADPH are inconclusive (Figure 9). ITC experiments thus far have yielded binding isotherms that exhibit a high heat of dilution, which manifests in peaks that level off well below 0 kcal per mol of injectant. In other words, injections of NADPH still produce relatively large amounts of heat even after BdcA has become saturated or almost completely saturated with NADPH. This makes the data difficult to fit, because the model for data fitting is built on the assumption that the addition of ligand into a solution of fully saturated protein will results in a negligible change in the heat content of the system. As a result, the fitting in Figure 9B is much worse than 9A, which have Chi² values that differ by more than an order of magnitude. Varying the time between injections from 330 seconds to 200 seconds resulted in no significant difference in data fitting (data not shown), further strengthening our hypothesis that the ITC issues are due to heats of dilution.

Fitting for the E50Q binding isotherm can be marginally improved by subtracting the reference data of buffer titrated into protein without NADPH (Figure 9C). The resulting isotherm yields a K_d value of $11.2\pm3.1 \mu$ M, which is a promising result for correlating cofactor binding and biofilm dispersal. However, several issues suggest that this value could be inaccurate. The Chi² still differs substantially from data fitting to the WT BdcA binding isotherm in Figure 9A. In addition, it is unsettling that Figure 9C is the result of a binding isotherm from which reference data has been subtracted, and no reference data has been subtracted from 9A. Consequent ITC trials all resulted in heats of dilution comparable to those obtained for E50Q, suggesting that the single binding isotherm for WT BdcA used to calculate its affinity for NADPH might be anomalous (data not shown), possibly due to better buffer matching between BdcA and NADPH in the first ITC trial. In future experiments, it will be worthwhile devote effort to standardizing the technique and obtaining reliable values for the heat of dilution. For figure 9C, reference data used resulted from the titration of buffer into protein, but it will also be worthwhile to determine the effect of NADPH titration into a solution of buffer. Finally, reducing the concentration of titrated NADPH may help to resolve the issue, because a greater portion of peaks in the binding isotherm will show heat release that is truly dependent on binding rather than dilution effects.

The E50Q mutant of BdcA has previously been shown to disperse biofilms better than WT BdcA.¹ However, it is not known how any of the other mutants used in this work affect biofilm dispersal relative to WT BdcA. Any trends between NADPH affinity and biofilm dispersal capability could provide valuable insights into the mechanism of BdcAdependent biofilm dispersal. Additionally, differences in the ability of the K150I mutant to disperse biofilms would suggest that catalytic ability for NADPH is important in its mode of function.

Expression of BdcA in *E. coli* during biofilm dispersal assays did not affect cell growth compared to uninduced controls (Figure 10A). Beyond this, it would be difficult to extract any conclusions from the data obtained from biofilm dispersal assays. The crystalviolet staining method of biofilm quantification is highly sensitive to small variations in technique. The method involves washing of the assay plate in order to remove planktonic cells and then later washing again to remove crystal violet dye, but washing that is too rigorous will disrupt biofilms within the wells of the plate.



Figure 10. Biofilm dispersal assay. (A) A₆₀₀ readings of biofilm assay wells before staining with crystal violet at 12 and 23 hours after IPTG induction show that BdcA expression has no significant effect on cell density. (B) A₆₃₀ readings of assay wells after staining with crystal violet at 12 and 23 hours post-induction. Results are the average of three wells for each colony, with two colonies used for WT and each mutant. WT3 and WT4 are negative controls that were not induced with IPTG. Errors shown are one standard deviation.

It is likely that the differences observed between 12 and 23 hours post-induction are a result of different amounts of washing rather than a drastic reduction in overall biofilm levels, especially considering that these results reflect biofilm levels from a single trial. In one subsequent experiment, washing was so vigorous that biofilms were removed completely before crystal staining was carried out (data not shown). In future experiments, the technique will require refinement to reduce error in the data, and multiple trials will be necessary to determine the variation among different washes within otherwise identical samples.

Conclusion

Initially, we focused on investigating BdcA, a protein that disperses biofilms in a variety of gram-negative bacteria,¹⁰ as a novel c-di-GMP receptor. Although the enzymes involved in making and breaking down c-di-GMP are relatively well characterized, less is known about c-di-GMP receptors, and the search for such binding partners is ongoing. BdcA contains a Rossmann fold, which had never been identified in a c-di-GMP receptor, and suggested that it bound to c-di-GMP through an uncharacterized mechanism. Just as other c-di-GMP binding domains are conserved, we hypothesized that the c-di-GMP binding domain within BdcA might also be conserved, and set out to identify potential homologs with a similar binding domain. We used structural similarity rather than sequence identity to identify potential homologs because proteins with a Rossmann fold show low sequence similarity. No obvious homologs emerged, and therefore we turned to mutagenesis to initially probe the determinants of c-di-GMP binding. We found that BdcA did not interact with c-di-GMP at all, either by degrading or binding it, and recent literature corroborated our findings.¹³

We then turned to probing the mechanism by which BdcA disperses biofilms, since the only explanation behind this process was refuted. We demonstrated that BdcA binds to NADPH. Several of the mutants that we created to investigate c-di-GMP binding are relevant to studies with NADPH. Based on comparison of BdcA with other Rossmann fold proteins, these mutations should interfere with either binding or catalysis specifically. In addition, the mutation E50Q has previously been shown to enhance biofilm dispersal, but its effect on NADPH affinity is unknown.¹ ITC experiments to determine affinity of BdcA mutants are ongoing. Large heats of dilution in mutant isotherms, which render data inconclusive, have hampered current efforts. Assays to correlate NADPH interaction to biofilm dispersal capability were undertaken. Before producing reliable data, the technique by which we quantify biofilms must be improved.

Methods

1. Genomic isolation and cloning

E. coli K-12 MG1655 was cultured overnight and genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit as specified in the protocol supplied by Promega. The primers used to amplify *bdcA* from the genomic DNA via PCR and incorporate restriction sites for BamHI and XhoI are listed in table S1. Amplified DNA and pET28a were digested using BamHI and XhoI. Ligation was carried out using the Fast-Link[™] DNA Ligation Kit. Ligation product was transformed into *E. coli* DH5α cells. These cells were cultured overnight and plasmid was isolated using the GenElute[™] Plasmid Miniprep Kit according to the protocol from Sigma-Aldrich Co. Cloning success was verified via PCR amplification followed by gel electrophoresis (Figure S3).

2. Site-directed mutagenesis

Mutations were incorporated into the constructed plasmid using the Quickchange® Lightning Site Directed Mutagenesis Kit as specified by the corresponding instruction manual. DNA sequencing then confirmed successful mutation. Primers used and mutants created are listed in table S1.

3. Protein expression and purification

pET28a plasmid containing the BdcA gene was transformed into *E. coli* Tuner pLysS. Cultures were induced with 0.1mM IPTG after reaching an OD₆₀₀ of 0.6. All growth media was supplemented by 40 µg/mL kanamycin and 30 µg/mL chloramphenicol. Cells were harvested by centrifugation after 24 hours, homogenized, and purified via affinity chromatography using a Ni-NTA column. Purified protein was immediately dialyzed into the appropriate buffer (specific buffers for each experiment are given in subsequent sections), flash frozen, and stored at -80°C. The viability of this protocol was verified via gel electrophoresis (Figure S4).

4. Phosphodiesterase Activity Assay

BdcA that had been dialyzed into PDE Assay Buffer¹ (50 mM Tris, 5 mM MgCl₂, 50 mM NaCl, pH=6) immediately following affinity chromatography purification was incubated with 10 μ M c-di-GMP (BioLog) for 1 hour. EDTA was added and samples were heated at 95°C for 5 minutes to denature protein. Supernatant was collected and analyzed via HPLC. Traces were compared to c-di-GMP and pGpG standards (BioLog), as well as standards containing a mixture of the two.

5. HPLC analysis of nucleotides

HPLC was performed using an Agilent 1260 Infinity system with photodiode array detector on reverse phase C18 columns. Flow rates were 1 mL/min and samples were run at room temperature. A variety of conditions were used to optimize conditions for quantifying c-di-GMP.

For runs with a supelcosil (Sigma-Aldrich) LC-18 column (25 cm × 4.6 mm, 5 µL particle size): Gradients of buffer A (50 mM triethylamine, pH 5.2 with acetic acid) and buffer B (60% buffer A and 40% acetonitrile) were used as follows. Starting with 100% A, the composition was changed to 65% A at 10 minutes with a constant gradient, 0% A at 16 minutes, 100% A at 22 minutes, and maintained at 100% A until 25 minutes. These specifications were used for the phosphodiesterase assays. Gradients of buffer A (water with 0.1% formic acid) and buffer B (methanol with 0.1% formic acid) were also tested on this column with the following gradients: 100% A at 0 minutes, 98.5% A at 0.1 minutes, 92% A at 12 minutes, 98% A at 4.1 minutes, 98% A at 8 minutes, 92% A at 10 minutes, and 98.5% A at 25 minutes.

For runs on a Microsorb-MV 100 C18 column: Gradients of buffer A (0.1M KH₂PO₄, pH=6) and buffer B (90%A, 10% methanol) were used as follows: 100% A at 0 minutes, 100% A at 9 minutes (the flow was increased to 1.3 mL/minute at this time point and maintained for the remainder of the run), 75% A at 15 minutes, 10% A at 17.5 minutes, 0% A from 19 to 23 minutes, and 100% A at 24 minutes. In addition, runs with gradients of buffer A (0.1M triethylamine, pH=6) and buffer B (50% A and 50% methanol) were performed with the following specification. Starting with 100% A, the percentage of A was

transitioned to 50% with a constant gradient until 25 minutes, and then transitioned to 100% A until 30 minutes.

6. Sequence and structure alignments

BdcA FASTA sequence was submitted to NCBI's Basic Local Alignment Search Tool (BLAST), specifically using the protein blast and excluding *Escherichia coli* from the search results to obtain superfamily data. The same sequence was submitted to the SWISS-MODEL server using an automatically generated template and an unspecified template.^{17–20} Structural homologs were found using the Dali server, with specifications as detailed previously.²¹ MacPyMOL was used to align FabG and BdcA and create a variety of corresponding images.

7. Rapid Equilibrium Dialysis (RED)

Single-Use Rapid Equilibrium Dialysis (RED) Plate with Inserts, 8K MWCO (Pierce) were used to measure c-di-GMP (BioLog) affinity for BdcA. In all experiments, buffers used were the same as those that the protein had been dialyzed into.

For c-di-GMP binding to WT BdcA and analysis via HPLC: Sample chambers were filled with 10 µM c-di-GMP and BdcA in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8mM KH₂PO₄). Buffer chambers were filled with PBS only. Control experiments were performed with no c-di-GMP added. Each experiment was performed in triplicate using multiple wells on the same plate. Plates were covered with sealing tape and incubated at 37°C for four hours. Samples were collected and analyzed via HPLC.

22

For c-di-GMP binding to WT BdcA and analysis via UV-Vis spectroscopy: Sample chambers were filled with 10 μ M BdcA in PDE assay buffer and c-di-GMP sufficient to reach 20 μ M total across both sample and buffer chambers. Control experiments were performed with no c-di-GMP added. Each experiment was performed in triplicate using multiple wells on the same plate. Plates were covered with sealing tape and incubated at 30°C for four hours. Absorbance at 251nm was measured using a Cary 100 UV-Vis Spectrophotometer.

For c-di-GMP binding to BdcA E50Q in sample chambers: Sample chambers were filled w 7.5 µM BdcA E50Q. Buffer chambers were filled with c-di-GMP (kerafast) in PDE assay buffer sufficient to reach 10 in µM both sample and buffer wells. Control experiments were performed with no c-di-GMP added. Each experiment was performed in triplicate using multiple wells on the same plate. Plates were covered with sealing tape and incubated at 30°C for four hours. Absorbance at 251nm was measured using a Cary 100 UV-Vis Spectrophotometer.

For c-di-GMP binding to BdcA E50Q in buffer chambers: Buffer chambers were filled with 7.5 μM BdcA E50Q in triplicate. Corresponding sample chambers were filled with 2.5, 5, or 7.5 μM of c-di-GMP (total concentration between sample and buffer wells) in PDE assay buffer to establish a gradient. Sample and buffer wells were filled in tandem with no BdcA to serve as a negative control. Plates were covered with sealing tape and incubated at 30°C for four hours. Absorbance at 251nm was measured using a Cary 100 UV-Vis Spectrophotometer

8. Isothermal titration calorimetry

A VP-ITC (GE Healthcare) ITC was used to titrate ligand into samples of protein under constant stirring. Bio-Rad Protein Assay was used to determine protein concentration. All samples were degassed using a Microcal Thermovac before loading.

For binding studies of c-di-GMP to E50Q BdcA: Protein samples that had been dialyzed in PDE assay buffer immediately following purification and c-di-GMP in PDE assay buffer were loaded at concentrations of 10 µM and 300 µM, respectively. Each titration was carried out at 20°C and with stirring at 270 rpm, 18 injections of µL. Protein was omitted in one negative control, and c-di-GMP in another. In one experiment, BdcA was dialyzed into high salt buffer (50mM Tris, 0.5mM EDTA, 25mM MgCl₂, 250mM NaCl) immediately after purification, and this buffer was also used in place of PDE assay buffer for the rest of the experiment.

For binding studies of NADPH: Immediately after protein purification via affinity chromatography, buffer used during dialysis (20mM HEPES) was collected for each protein in order to create specific NADPH solutions that matched each protein solution exactly. Concentrations of NADPH solutions were verified via absorbance at 340nm. NADPH at 600 uM was titrated into a 12uM sample of protein at 25°C under constant stirring at 310rpm. Each titration was carried out using 30 NADPH injections of 10uL with 200 seconds between injections. After a single titration with WT BdcA verified NADPH binding, the first injection was changed to 2 uL for subsequent experiments. In a single trial that was a repeat of a previous trial in every other aspect, the spacing between injections was changed to 330 seconds.

9. Biofilm dispersal assay

Overnight cultures of WT BdcA and mutants were diluted in LB to 0.05AU at 600nm in 96-well polysterene plates. After 19 hours, sample wells were inoculated with 0.1mM IPTG and positive controls were left undisturbed. Plates were shaken for 1 minute at 150 rpm before incubating further. Crystal Violet endpoint assays²⁶ were used to quantify biofilms at 12 and 23 hours post-induction.

Works Cited

- 1. Ma, Q., Yang, Z., Pu, M., Peti, W. & Wood, T. K. Engineering a novel c-di-GMP-binding protein for biofilm dispersal. *Environ. Microbiol.* **13**, 631–42 (2011).
- 2. Donlan, R. M. Biofilm formation: a clinically relevant microbiological process. *Clin. Infect. Dis.* **33**, 1387–92 (2001).
- 3. Blenkinsopp, S. A. & Costerton, J. W. Understanding bacterial biofilms. **9**, (1991).
- 4. Potera, C. Biofilms Invade Microbiology. *Science (80-.).* 273, (1996).
- 5. Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–33 (2010).
- 6. Park, A., Jeong, H.-H., Lee, J., Kim, K. P. & Lee, C.-S. Effect of shear stress on the formation of bacterial biofilm in a microfluidic channel. *BioChip J.* **5**, 236–241 (2011).
- 7. US National Institutes of Health. No Title.
- 8. Kostakioti, M., Hadjifrangiskou, M. & Hultgren, S. J. Bacterial Biofilms: Development, Dispersal, and Therapeutic Strategies in the Dawn of the Postantibiotic Era. *Cold Spring Harb. Perspect. Med.* 1–24 (2013).
- 9. Worthington, R. J., Richards, J. J. & Melander, C. Non-Microbicidal Control of Bacterial Biofilms with Small Molecules. *Anti-Infective Agents* **12**, 120–138 (2014).
- 10. Ma, Q., Zhang, G. & Wood, T. K. Escherichia coli BdcA controls biofilm dispersal in Pseudomonas aeruginosa and Rhizobium meliloti. *BMC Res. Notes* **4**, 447 (2011).
- 11. Robert-Paganin, J., Nonin-Lecomte, S. & Réty, S. Crystal Structure of an EAL Domain in Complex with Reaction Product 5'-pGpG. *PLoS One* **7**, 1–13 (2012).
- 12. Spurbeck, R. R., Tarrien, R. J. & Harry, L. T. Enzymatically Active and Inactive Phosphodiesterases and Diguanylate Cyclases Are Involved in Regulation of Motility or Sessility in Escherichia coli CFT073. (2012). doi:10.1128/mBio.00307-12.Editor
- 13. Lord, D. M., Baran, A. U., Wood, T. K., Peti, W. & Page, R. BdcA, a protein important for Escherichia coli biofilm dispersal, is a short-chain dehydrogenase/reductase that binds specifically to NADPH. *PLoS One* **9**, e105751 (2014).

- 14. Bijtenhoorn, P. *et al.* A novel metagenomic short-chain dehydrogenase/reductase attenuates Pseudomonas aeruginosa biofilm formation and virulence on Caenorhabditis elegans. *PLoS One* **6**, e26278 (2011).
- 15. Lesk, a M. NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* **5**, 775–83 (1995).
- 16. Kallberg, Y., Oppermann, U. & Persson, B. Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. *FEBS J.* **277**, 2375–2386 (2010).
- 17. Biasini, M. *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, 252–258 (2014).
- 18. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37**, 387–392 (2009).
- 19. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, 162–173 (2009).
- 20. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: A webbased environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
- 21. Holm, L. & Rosenström, P. Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, 545–549 (2010).
- Schmidt, A. J., Ryjenkov, D. A. & Gomelsky, M. The Ubiquitous Protein Domain EAL Is a Cyclic Diguanylate-Specific Phosphodiesterase : Enzymatically Active and Inactive EAL Domains The Ubiquitous Protein Domain EAL Is a Cyclic Diguanylate-Specific Phosphodiesterase : Enzymatically Active and Inactive E. Society 187, 4774–4781 (2005).
- 23. Sundriyal, A. *et al.* Inherent Regulation of EAL Domain-catalyzed Hydrolysis of Second Messenger Cyclic di-GMP. *J. Biol. Chem.* **289**, 6978–6990 (2014).
- 24. Dutta, D., Bhattacharyya, S. & Das, A. K. Crystal structure and fluorescence studies reveal the role of helical dimeric interface of staphylococcal fabg1 in positive cooperativity for NADPH. *Proteins Struct. Funct. Bioinforma.* **80**, 1250–1257 (2012).
- 25. Joernvall, H., Persson, B. & Krook, M. Short-Chain Dehydrogenases/Reductases (SDR). *Biochemistry* **34**, (1995).
- 26. Jackson, D. W. *et al.* Biofilm Formation and Dispersal under the Influence of the Global Regulator CsrA of Escherichia coli. *J. Bacteriol.* **184**, 290–301 (2002).

Supplementary

					1	10	:	20	30 40	
Bdca					MGAF	FGKTVLI I	GSRGI	GAAIVRRFVT	DGANVRFTYAGSK	DAA
3V2G	. МННННН	IHSSGVDI	GTENLY	FQSMI	MTSISL	AGKTAFVI	GGSRGI	GAAIAKRLAL	EGAAVALTYVNAA	ERA
3RRU 3RSH			SNAM		SOFMAL	EGKVALVI	GASRGI	CKAIAELLAE	RGAKV.IGTATSE	SGA
3TZK			SNAM		SOFMNL	EGKVALVI	GASRGI	GKATAELLAE	RGAKV. IGTATSE	SGA
3TZH			SNAM		SQFMNL	EGKVALVI	GASRGI	GKAIAELLAE	RGAKV.IGTATSE	SGA
30P4			M		SQFMNL	EGKVALVI	G <mark>ASRG</mark> I (GKAIAELLAE	RGAKV.IGTATSE	SGA
3FTP	MAHHHHH	IHMGTLEA	QTQGPG		SMDKTL	DKQVAIVI	GASRGI	GRAIALELAR	RGAMV.IGTATTE	AGA
4DMM 100 S	MGSSHHH	INNHSSGI	VPRGSH	• • • • •	MTALPL	LDKIAPA,	GASEGI	GRAIALELAA ODATITITINA	RCART VACDIER	CDT
3NUG					TERL	AGKTALVI	GAAOGI	GKATAARLAA	DGATVIVSDINAF	GAK
3AWD			.GSHMY		MEKLRL	DNRVAIV	GAQNI	GLACVTALAE	AGARV. IIADLDE	AMA
2WDZ			MDY	1	RTVFRL	DGACAAVI	GAGSGI (<mark>G</mark> LEICRAF A A	SGARL.ILIDREA	AAL
3LQF			MDY	••••	RTVFRL	DGACAAV	GAGSGI	GLEICRAFAA	SGARL.ILIDREA	AAL
310X 38K4			GSH	••••	ACTEDL	SCREATVI	GASSGI	GRAAALLFAR	A CATVALADI.DVM	
consensus>50					m.1	egkval!t	GasrgI	G.aiala.	.GA.v.iae	a
						- j	j -			
	50		6	•	7	•	8	•	90 100	
Bdca	KRLAQEI	'G	TAVFTD	SADR	DAVIDV	VRKS	GALDIL'	V V N <mark>A</mark> G I G V F G	. EALELNADDIDR	LFK
3V2G	QAVVSEI	EQAGGRA	VAIRAD	NRDA	EAIEQA	IRETVEAI	GGLDIL	VNSAGIWHSA	. PLEETTVADFDE	VMA
3RRU 3RSH	QAIS	DYLCDNC	KGMALN	VTNP.	ESIEAV	LKALTDER	GGVDIL	VNNAGITRDN VNNAGTTRDN	. LLMRMKEEEWSL	TME
3TZK	OAIS	DYLGDNO	KGMALN	VTNP	ESIEAV	LKAITDEE	GGVDIL	VNNAAITRDN	.LLMRMKEEEWSD	IME
3TZH	QAIS	DYLGDNG	KGMALN	VTNP	ESIEAV	LKAITDEH	GGVDIL	VNN <mark>A</mark> GITRDN	. LLMRMKEEEWSD	IME
30P4	QAIS	DYL <mark>G</mark> DNG	KGMALN	VTNP	E S I E A V	LKAITDEH	GGVDIL	V N N <mark>A</mark> G I T R D N	. LLMRMKEEEWSD	IME
3FTP	EGIGAAF	KOAGLEO	RGAVLN	VNDA	TAVDAL	VESTLKEN	GALNVL	VNNAGITODO	. LAMRMKDDEWDA	VID
4DMM 1ULS	DEVVAAI	AAAGGEA	HDVVMD	VADD	ASVERL	FAAVIERV FAFATAHI	GREDVE	VNNAGITRDT VHVAGTTRDN	- LLLKMKKDDWQS	VLD.
3NUG	AAA	ASIGKKA	RAIAAD	ISDP	GSVKAL	FAEIOALI	GGIDIL	VNNASIVPFV	.AWDDVDLDHWRK	IID
3AWD	TKAVEDI	RMEGHDV	SSVVMD	VTNT	ESVQNA	VRSVHEQI	GRVDIL	VAC <mark>A</mark> GICISE	VKAEDMTDGQWLK	QVD
2WDZ	DRAAQEI	GAAV/	ARIVAD	VTDA	E A M T <mark>A</mark> A	AAEA.EAN	/APVSIL'	V N S <mark>A</mark> G I A R L H	. DALETDDATWRQ	VMA
3LQF	DRAAQEI	GAAV	ARIVAD	VTDA	EAMTAA	AAEA.EAV	APVSIL	VNSAGIARLH	. DALETDDATWRQ	VMA
3TOX 3AK4	AELTDEI	AGLENGO	FAVEVD	VGDE	ALHEAL	VELAVREI	GGLDTA	CANAGALGAM	. PAVDITDEEWDE	NFD
consensus>50	a.i	q.na	1v.l#	vtd.	esvea.	i.d.	aavdil	vnnAgitrdn	m.m.eeew.d	vmd
	-	-						•		
	110	, 1	20		130	14	٠ <u>٩</u>	150	160	17 ọ
Bdca	110	HASVEA	ROMPE .	G <mark>G</mark> I	130 RILIIG	14 SVNGDRMI	VAGM . A	150 AYAASKSALQ	160 GMARGLARDFGPR	170 GIT
Bdca 3V2G 3PRO	110 INIHAPY VNFRAPF TNLTSIE	HASVEAA VAIRSAS	ROMPE.		130 RILIIG RIITIG RIITIG	14 SVNGDRMI SNLAELVI	VAGM.A	150 A <mark>y</mark> aasksalQ L <mark>y</mark> saskaala	160 GMARGLARDFGPR GLTKGLARDLGPR CFTKSLAPFUASP	170 GIT GIT
Bdca 3V2G 3RRO 3RSH	110 INIHAPY VNFRAPF TNLTSIF TNLTSIF	I HASVEA VAIRSAS RLSKAVI	RQMPE. RHLGD. RGMMKK	GG GG ROG ROG	130 RILIIG RIITIG RIINVG RIINVG	14 SVNGDRMI SNLAELVI SVVGTMGI SVVGTMGI	VAGM.A WPGI.S AGQ.A I	150 AVAASKSALQ LYSASKAALA NYAAAKAGVI VAAAKAGVI	160 GMARGLARDFGPR GLTKGLARDLGPR GFTKSMAREVASR GFTKSMAREVASR	170 GIT GIT GVT
Bdca 3V2G 3RRO 3RSH 3TZK	110 INIHAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF	HASVEAA VAIRSAS RLSKAVI RLSKAVI RLSKAVI	RGMMKK RGMMKK RGMMKK	G G . R Q . R Q G I . R Q G I	130 RILIIG RIITIG RIINVG RIINVG RIINVG	14 SVNGDRMI SNLAELVI SVVGTMG1 SVVGTMG1 SVVGTMG1	VAGM.A WPGI.S VAGQ.A IAGQ.A IAGQ.A	150 AVAASKSALQ LYSASKAALA YAAAKAGVI NYAAAKAGVI NYAAAKAGVI	160 GMARGLARDFGPR GLTKGLARDLGPR GFTKSMAREVASR GFTKSMAREVASR GFTKSMAREVASR	GIT GIT GVT GVT GVT
Bdca 3V2G 3RRO 3RSH 3TZK 3TZH	1 1 0 INIHAPY VNFRAPE TNLTSIE TNLTSIE TNLTSIE	HASVEAA VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI	R Q M P E . R H L G D . R G M M K K R G M M K K R G M M K K R G M M K K		130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG	14 SVNGDRMI SNLAELVI SVVGTMGI SVVGTMGI SVVGTMGI SVVGTMGI	VAGM.A WPGI.S IAGQ.A IAGQ.AI IAGQ.AI	150 AYAASKSALQ LYSASKAALA NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI	160 GMARGLARDFGPR GLTKGLARDLGPR GFTKSMAREVASR GFTKSMAREVASR GFTKSMAREVASS GFTKSMAREVASS	GIT GIT GVT GVT GVT
Bdca 3V2G 3RRO 3RSH 3TZK 3TZH 30P4 2870	110 INIHAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF	I VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI	RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK	GG GG .RQG .RQG .RQG .RQG	130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG	14 SVNGDRMI SNLAELVI SVVGTMGI SVVGTMGI SVVGTMGI SVVGTMGI	VAGM.A WPGI.S JAGQ.AI JAGQ.AI JAGQ.AI JAGQ.AI JAGQ.AI	150 A YAASKSALQ LYSASKAALA NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI	160 GMAR GLARDFGPR GLTK GLARDLGPR GFTK SMAREVASR GFTK SMAREVASR GFTK SMAREVASR GFTK SMAREVASR GFTK SMAREVASR	170 GIT GIT GVT GVT GVT
Bdca 3V2G 3RRO 3RSH 3TZK 3TZH 3OP4 3FTP 4DMM	110 INIHAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLGGVF	I 'HASVEA 'RLSKAVI 'RLSKAVI 'RLSKAVI 'RLSKAVI 'RLSKAVI 'RLSKAVI 'RLSKAVI	RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RFMMKA	G G . RQG . RQG . RQG . RQG . RQG . RQG . RGG . RSG	130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG RIVNIT RIVNIT	14 SVNGDRMI SNLAELVI SVVGTMGI SVVGTMGI SVVGTMGI SVVGTMGI SVVGSAGI	VAGMA WPGISS AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGQAA AGAAA AGQAA AGAAA A AGAAA A AGAAA A AGAAA A AGAAA A A AGAAA A A A A A A A A A A A A A A A A A A	150 AVAASKSALQ LYSASKAALA NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI	160 GMARGLARDFGPR GFTKSMAREVASR GFTKSMAREVASR GFTKSMAREVASR GFTKSMAREVASR GFTKSMAREVASR GFTKSMAREVASR GTTRALAREIGSR GMTRALAREIGSR	170 GIT GVT GVT GVT GVT GVT GVT
Bdca 3V2G 3RRO 3RSH 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS	110 INIHAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLGGVF VNLTGSF	I VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI I LCSRAA LCSRAA	RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RMMKA KIMIKQ SEAMREK		130 RILIIG RIITIG RIINVG RIINVG RIINVG RIVNIT RIVNIT SIVLTA	14 SVNGDRMI SNLAELVI SVVGTMGN SVVGTMGN SVVGTMGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN	VAGM.A WPGI.S IAGQ.A IAGQ.A IAGQ.A IAGQ.A IAGQ.A I IAGQ.A I IPGQ.A I IPGQ.A	150 AYAASKSALQ LYSASKAALA NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVA YSAAKAGVA YSAAKAGVA YAAAKAGVA	160 GMAR GLARDIG PR GF TK SMAREVASR GF TK SMAREVASR GF TK SMAREVASR GF TK SMAREVASR GF TK SMAREVASR GF TK SMAREVASR GT TK SMAREVASR GLT KTVA KELASR GLT KTVA KELASR GLT KTVA KELASR	170 GIT GVT GVT GVT GVT GVT GVT GIT GIT
Bdca 3V2G 3RR0 3RSH 3TZK 3TZH 30P4 3FTP 4DMM 1ULS 3NUG	I 1 0 INIHAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLKAVF LNLGGVF VNLTGGF	HASVEAA PAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSRAVI LSRAVI LVAKAAS LVAKAAS	RQMPE SRHLGD RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RDMKK RDMK KIMIKQ EAMREK	GG GG RQG RQG RQG RQG RQG RGG RSG NPG GKAG	130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG RIVNIT RIINIA SIVLTA RVISIA	14 SVNGDRMI SNLAELVI SVVGTMGI SVVGTMGI SVVGTMGI SVVGTMGI SVVGSAGI SVVGSAGI SVVGEMGI SVVGEMGI SNTFFAGI	VAGM.A WPGI.S IAGQ.AI IAGQ.AI IAGQ.AI IAGQ.AI IAGQ.AI IAGQ.AI ICPGQ.AI ICPNM.AI	150 AYAASKSALQ LYSASKAALA NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVA NYSAAKAGVA NYAAKAGVA YAASMAGVA	160 GMARGLARDJCPR GF1TKGLARDLCPR GF1TKSMAREVJASR GF1TKSMAREVJASR GF1TKSMAREVJASR GF1TKSMAREVJASR GF1TKSMAREVJASR GF1TKSMAREVJASR GL1TKTVARELJASR GL1TKTVARELJASR GL1TKTLALELGRM	170 GIT GUT GVT GVT GVT GVT GUT GIT GIT NIT
Bdca 3v2G 3RRO 3RSH 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3NUG 3AWD	III INIHAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLTGVF VNLTGSF VNLTGSF INLNGMF	HASVEAA VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LSSRAVI LVAKAAS IVVAKAAS SCOAVC	RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RCMMKA KIMIKQ EAMREK CDQMRAA RIMLEQ	GG GG RQG RQG RQG RQG RGG RGG NPG GKAG KQG	130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG RIVNIT RIVNIT SIVLTA RVISIA VIVAIG	I A SVNGDRMI SNLAELVI SVVGTMGN SVVGTMGN SVVGTMGN SVVGSAGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SNTFFAG7	VAGM.A WPGI.S: NAGQ.A NAGQ.A NAGQ.A NAGQ.A NAGQ.A NAGQ.A NPGQ.A NPGQ.A NPGQ.A NPGQ.A NPGQ.A	150 AYAASKSALQ LYSASKAALA NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYAAAKAGVI NYSAAKAGVI AYAAKAGVI AYAAKGVI	160 GMARGLARDFGPF GLUKGLARDFGPF GLUKGLARDFGPT GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GTKSMAREVASF GLTKTLALELGSW GLTRTLALELGSW GPTRALATELGSV GVTRSLAREWAPF	170 GIT GVT GVT GVT GVT GVT GIT GIT GIR NIT GIR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3NUG 3AMD 2WDZ 2T OF	I 1 0 INI HAPY VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLTGIF VNLTGFF VNLTGFF VNLTGFF	HASVEAA VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LSRAVI LCSRAA LVAKAAS I VTRAGI KSCQAVC	RQMPE. RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RDMMKA SEAMREK CDQMRAA RIMLEQ RAMVAR	GG GG .RQG .RQG .RQG .RQG .RQG .RQG	130 RILIIG RIINVG RIINVG RIINVG RIINVG RIINVG RIVNIT RIVNIT RVISIA VIVAIG AIVNIC	I G SVNGDRMI SVVGTMGN SVVGTMGN SVVGTMGN SVVGTMGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN SNTFFAGT SMSGLIVN SMSGTIVN	VAGM.A WPGIS AGQ.A IAGQ	150 CARAKASKALQ CSASKALA NAAAKAGVI VAAAKAGVI WAAAKAGVI NAAKAGVI VAASKAGVI VAASKAGVI VAASKAGVI VAASKAGVI VAASKAGVI MASKAGVI MASKAGVI MASKAGVI	160 GMAR GLARD JCPF GLTK GLARD LIGPF GFTK SMAREV JASF GFTK SMAREV JASF GFTK SMAREV JASF GFTK SMAREV JASF GFTK SMAREV JASF GFTK SMAREV JASF GLTK TLALELIGFK GFTR RLARE LIGSF GLTR TLALELIGFK GFTR RLARE LIGSF GFTR RLARE LIGFK GFTR RLARE SMAPP	170 GIT GUT GVT GVT GVT GVT GUT GIT GIR GIR GIR GUR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZH 30P4 3FTP 4DMM 1ULS 3MVG 3AWD 2MDZ 3LQF 3TOX	110 INIHAPY VNFRAPH TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLTGGF VNLTGGF VNVDGMF VNVDGMF VNVDGMF	I HASVEAA VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI IRLSKAVI ILSRAVI ILSRAVI ILVAKAAS IVTRAGI ISCQAVG WASRAFG	RQMPE RHLGD RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RCMMKK RCMMKK RCMMK RCMK RC	GG GG .RQG .RQG .RQG .RQG .RQG .RQ	130 RILIIG RIINVG RIINVG RIINVG RINVG RIVNIT RIVNIT RIVNIT RIVISIA VIVAIG AIVNLG AIVNLG	I 4 SVNGDRMI SNLAELVI SVVGTMGN SVVGTMGN SVVGTMGN SVVGSAGN SVVGSAGN SVVGEMGN SNVGEMGN SNVGEMGN SNSGLIVN SMSGLIVN SMSGTIVN SMSGTIVN	VAGM.A WPGI.SI AGQ.AI AGQ.AI AGQ.AI AGQ.AI PGQ.VI ILGQ.AI ILGQ.AI ILGQ.AI ILGQ.AI ILGQ.AI IRPQQA IRPQQAS IRPQFASI	150 SAXKALA SAXKALA XAAXKAGVI XAAXKAGVI XAAXKAGVI XAAXKAGVI XAAXKAGVI XAAXKAGVI XAXKAGVI XAXKAGVI XAXKAGVI XAXKAGVI XAXKAGVI XAXKAGVI XAXKAGVI	160 GMARGLARDJCGPR GF1TKGLARDLCGPR GF1TKSMAREVJASF GF1TKSMAREVJASF GF1TKSMAREVJASF GF1TKSMAREVJASF GF1TKSMAREVJASF GF1TKLAREVJASF GL1TKTJAELJASF GL1TKTJAELJASF GL1TKTJAELJASF GL1TKTJAELGPX QYIRSLAAEWACF QJITRALAAEWACF QJITRALAAEWACF QJITRALAAEWACF QJITRALAAEWACF QJITRALAAEWACF	170 GIT GUT GVT GVT GVT GVT GUT GUT GIR GIR GUR GUR GUR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZH 3PTP 4DMM 1ULS 3NUG 3AWD 2NDZ 3LQF 3TOX 3AR4	III INIHAP VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLTSIF UNLTSUF VNLTGTF VNLTGTF VNVDGMF VNVDGMF VNVDGMF VNVDGMF	I VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI I LSRAVI LVAKAAS VI VRARAFC VASRAFC VASRAFC LAAKYQV LANQIAC	ROMPE RELGD. RCGMKK RGMKK RGMKK RGMKK RGMKK RPMKA KIRGMKK RPMKA RIMKA RIMKA RIMKA RAMVAR RAMVAR PAIAAL		130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG RIINVG RIVNIG AIVNLG AIVNLG AIVNLG LTFTSS	14 SVNGDRMI SNLAELVI SVVGTMGN SVVGTMGN SVVGTMGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGSAGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SVVGEMGN SMSGTIVN SMSGTIVN SVGHTAGN SLAAKVG/	VAGM.A WPGI.S WAGO.AI WAGO.AI WAGO.AI WAGO.AI WAGO.AI WPGO.AI WPGO.AI WPGO.AI WPGO.AI WPGO.AI WPGO.AI WRPOFAS WRPOFAS WRPOFAS	150 AMAASKSALQ CSASKAALA CAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI SMASKGAVH SMASKGAVH SMASKGAVH SMASKGAVH SMASKGAVH	160 GMARGLARDFGPF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GTKRLAELGSK GTTRLALELGSK GPTRALATELGSK GPTRALATELGSK GPTRALATELGSK GPTRALATELGSK GLTRLAEWASF GLTRLAEWASF GLTRLAEWASF	IT GIT GVT GVT GVT GVT GIT GIR GIR GIR GIR GVR GVR GVR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZH 30P4 3FTP 4DMM 1ULS 3NUG 3NUG 3NUG 3NUG 3ND 2VD2 3LQF 3TOX 3AK4 consensus>50	I 1 9 INIHAPY VNFRAPP TNLTSIF TNLTSIF TNLTSIF TNLTSIF UNLTSIF VNLTGGF VNLTGGF VNLTGGF VNLTGGF VNLTGGF VNLTGGF VNVDGMF TNLTGF VNVDGMF	HASVEA VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI LSSKAVI LCSRAAP LVAKAAS IVTRAGI WASRAFC VASRAFC LANQIAC	ROMPE. RRGMKK RGKK RGKKK RGKKK RGKKK RGKKK RGKKK RGKKK RGKKKK RGKKKK RGKKKK RGKKKKK RGKKKKKK RGKKKKKKKKKK		130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG RIINVG RIVNIG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG	LA SVNGDRMI SNLAELVI SVVGTMGN SVVGTMGN SVVGTMGN SVVGSAGT SVVGSAGT SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SNTFFAG SUDALVG SUVG SUDALVG SUDALVG SUDALVG SUVG SUDALVG SUDALVG SUDALVG SUDALVG SUDALVG SUDALVG SUDALVG SUVG SUDALVG SUVG SUVG SUVG SUVG SUVG SUVG SUVG SU	VAGM. A WPGI.S. MAGQ. A AGQ. A A RPOPAS: AGV. A A GQ. A A GQ. A A GQ. A A C A C A C A C A C A C A C A C A C A	150 SASKALA SASKALA VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVA SAAKAGVA SAAKAGVA SAAKAGVA SAAKAGVA SAAKAGVA SMASKGAVH PAASKAGLI HS	160 GMAR GLARDICPE GLING GLARDICOPE GFINS MAREVIASE GFINS MAREVIASE GFINS MAREVIASE GFINS MAREVIASE GFINS MAREVIASE GFINS MAREVIASE GLINT LALELIGRW GFINS MAREVIS GLINT LALELIGRW GFINS MARE IGSS GLINT LALELIGRW GFINS MALATELIGRW GFINS MAREVIS GFINS MALATELIGNW GFINS MALATELIGNW GF	170 GIT GVT GVT GVT GVT GIT GIR GIR GVR GVR GVR GVR GVR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZK 3PTP 4DMM 1ULS 3NUG 3AWD 2WDZ 3LQF 3TQX 3AK4 consensus>50	I 1 0 IN I HAPH VNF HAPH TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLLTSUGGE UNLTGGE VNLTGGE VNLTGGE VNLTGGE VNLTGGE VNVLGGE VNVLTGE VNVLGGE VNVLTE VNVLTE VNVL	HASVEA VAIRSA VAIRSA RISKAVI RISKAVI RISKAVI RISKAVI RISKAVI LCSRAA IVTRACT RSCQAVC WASRAFC LAAKYO LAAKYO LAAQUAC	ROMPE. RHLGD. RGMMKK RGMKK RGMKK RGMKK RGMKK RGMKK RDMKA K RDMKA RAMKA R	GG GG RQG RQG RQG RQG RQG RSG RSG RSG 	I39 RILITIG RIITIG RIINVG RIINVG RIINVG RIINVI RIVNIT RIVNIT RIVNIG AIVNLG AIVNLG AIVNLG AIVNLG XIVNTA	14 SVNGDRMK SNLAELVI SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAG SVVGE SVVSCHG SVVSCHG SVVSCHG SVVSCHG SVVSCHG SVVGE SVSCTVV SNSGTVV SNSGTVV SVGHTAGI SVVGEMGI	9 VAGM.A WPPGI.S AGCO.A	150 SASKALQ SASKALA YAAAKAGVI YAAAKAGVI YAAAKAGVI YAAKAGVI YAAKGGVI YAAKGGVI YAAKGGVI YAAKGAVH YAAKGAVH YAAKGAVH YAAKGAVH YAAKGAVH YAAKGAVH	160 GMARGLARDJCGPR GFITKGLARDJCGPR GFITKSMAREVJASR GFITKSMAREVJASR GFITKSMAREVJASR GFITKSMAREVJASR GFITKSMAREVJASR GLITRTLALEIGST GLITRTLALEIGST GUTRLALEIGST QJITRLAREMJAG UITRLAREMJAG UITRLAREMJAG GITRLALAREMJAG GITVQLAREJGAR	170 GIT GVT GVT GVT GVT GUT GIR GIR GVR GVR GVR GVR GVR
Bdca 3y2G 3RRO 3RSH 3TZH 3OP4 4D3M 4D3M 1ULS 3NUG 3NWD 2WDZ 3LQF 3TOX 3AK4 consensus>50	110 IN THAP IN	HASVEA VAIRSAS RISKAVI RISKAVI RISKAVI RISKAVI CSRAAP LVAKAAS ISKAVI VAKAAS VAKAAS VASAS V	ROMPE. RRHLGD. RGMMXK RGMMXK RGMMKK RGMMKK RGMMKK RRGMMKK RCMMXK RDMMXA RTIMLKQ EAMREK DOMRAA RRMVAR RAMVAR RAMVAR RAMVAR RAMVA R		130 RILIIG RIITIG RIINVG RIINVG RIINVG RIINVG RIVNIT RIINIA SIVLTA SIVLTA AIVNIG LIFTTSS VIVNTA riinig	14 SVNGDRMH SVLAELVI SVVGTMG SVVGTMG SVVGTMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG SVVGSMG	10 VVAGM.A. VVAGM.A. VAGM.A. VA	150 AMAASKSALQ USASKAAL MAAKAGVI MAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI MASKAKAVA MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH	160 GMARGLARDFGPF GLMKGLARDLGPF GPMKSMAREVASB GPMKSMAREVASB GPMKSMAREVASB GPMKSMAREVASB GPMKSMAREVASB GPMKSMAREVASB GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREISS GLMRTLALAREWAGF GLMQALAVELGAR GLMQALAVELGAR GLVQALAVELGAR GLVQALAVELGAR	170 GIT GVT GVT GVT GVT GVT GIT GIT GIR GVR GVR GVR GVR GVR
Bdca 3V2G 3RRH 3RSH 3TZK 3TZK 3TZH 3OP4 3PTP 4DMM 1ULS 3NUG 3AWD 2WDZ 3LQF 3TOX 3AR4 consensus>50	110 TITHAP VN PRAPI TN TSTFF TN TSTF TN TSTF TN TSTF VN LTGFF VN LTGFF VN LTGFF VN LTGFF VN LTGFF VN LTGFF VN VDGMA TSTFFF VN VDGMA TSTFFF VN L.18	HASVEA, VAIRSAS VAIRSAS RISKAVI RISKAVI RISKAVI RISKAVI RISKAVI LVAKAAS LVAKAAS UVAKAAS UVAKAAS LVAKAAS LAKYOV LAAKYOV LAAKYOV LAAKYOV	ROMPE. RHLGD. RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RMMK RMM		130 RILIIG RIJING RIJNVG RIJNVG RIJNVG RIJNVG RIJNJG AIVNIG AIVNLG AIVNLG AIVNLG AIVNLG L TFTSS VIVTA	14 EVNGDRMU SNLAELVI SVVGTMGI SVVGTMGI SVVGTMGI SVVGTMGI SVVGSAG SVVGS SVVGSAG SVVGSAG SVVGSAG SVVGSAG SVVGSAG SVVGSAG	40 WAGM.A. WPGI.S: AGQ.A.	150 AMAASKSALQ SASKAALA MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAKAGVI MAAKAGVA MASKAGVH MASKAGVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH	160 GMARGLARDJCPF GLTKGLARDJCPF GFTKSMAREVJASF GFTKSMAREVJASF GFTKSMAREVJASF GFTKSMAREVJASF GFTKSMAREVJASF GTTKSMAREVJASF GLTRTLALELGRW GFTRALATELGRW GFTRALATELGRW GFTRALATELGRW GTTRALAREWACF GLTVALAEWACF GLTVALAEWACF GLTVALAEWACF GLTVALAEWACF GLTVALAEWACF	IT O GIT GVT GVT GVT GUT GUT GIT SUT GIT GIR GIR GVR GVR GVR GVR GVR GVR GVR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3WDG 3MMD 2WDZ 3LQF 3TOX 3AK4 <i>consensus>50</i> Bdca	110 VNFRAP TNLTSIF TNLTSIF TNLTSIF TNLTSIF TNLTSIF VLLTGFF VLLTGFF VLLTGFF VNVDGMF VNVDGMF VNVDGMF VNVDGMF	HASVEA VAIRSAS VAIRSAS KISKAVI RISKAVI RISKAVI RISKAVI RISKAVI RISKAVI VTRAV VASRAS VVRAV VASRAS VI VASRAS V VASRAS V VASRAS V VASRAS V VASRAS V VASRAS V VASRAS V V V V V V V V V V V V V V V V V V V	ROMPE. RHLGD. RGMMKK RGMMKK RGMMKK RGMMKK RGMMKK RDMMKK RDMMKA RIMLEQ EAMVAR RAMVAR RAMVAR RAMVAR RAMVAR PAIJAL RHFLAS T. mm.	RQG RQG RQG RQG RQG RQG RQG S RQG S S A G A G G G G G G G G G G G G G S N T K G G G G G G S N C G C C C C C C C C C C C C C C C C C	130 RILIIG RIITIG RIINVG RIINVG RIINVG RIVNTR RIVNIT SIVLIA VIVAIG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG	14 SVNGDRMI SVVGTMCI SVVGTMCI SVVGTMCI SVVGTMCI SVVGTMCI SVVGTMCI SVVGEMCI SNVGEMCI SNSGIV SNSGIV SNSGIV SNSGIV SNSGIV SNSGIV SNSGIV SNSGIV SVVGHTACI SVVGTMGI SVVGTMGI	10 WPGI.S WPGI.S MAGO.A MA	150 SAXKSALQ SAXKACVI CAAAKACVI CAAAKACVI CAAAKACVI CAAAKACVI CAAXXACVI CAAX	160 GMAR GLARD LGPF GLTK GLARD LGPF GLTK GLARD LGPF GFTKS MARE VIASF GFTKS MARE VIASF GFTKS MARE VIASF GFTKS MARE VIASF GFTKS MARE VIASF GLTK TVARE LIGSF GLTK TVARE LIGSF GLTR TLALE LIGSK GFTR ALARE HIGSF GLTR TLALE LIGSK GFTR ALARE MAGF QLTR ALARE WAGF QLTR ALARE WAGF QLTR ALARE WAGF GLT VIELGAR LIGST CONSTRUCTION GT LIGST CONSTRUCTION GT LIGS	ITO GIT GUT GUT GUT GUT GUT GUT GIT GIT GUR GUR GUR GUR GUR GUR GUR GUR GUR
Bdca JV2G JRSH JT2K JT2K JT2K JT2H JPTP 4DMM 1ULS JNUG JAWD 2WDZ JLQF JT0X JAK4 consensus>50 Bdca JV2G	110 111 HAP V P RAP 11151 1151	I ASVEA VAISKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAPI LVAKAASI VTRACT NASCAVC WASRAFC LAAKYOV SAAC SAAC LAAKYOV SAAC SAAC LAAKYOV SAAC SAAC SAAC SAAC SAAC SAAC SAAC SAA	ROMPE. RHLGD. RGM XK RGM XK	RQG RQG RQG RQG RQG RQG RQG RGG GKQG GKQ	130 RILITIG RIITIG RIINVG RIINVG RIINVG RIVNIG RIVNIG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG	14 SVNGDRM SNLAELVI SNUGTMGS SVUGTMGS S	10 VVAGM.A. VVAGM.A. VADPGI.S VAGO.A. VARPOPAS:	150 AMAASKSALQ USASKAALA MAAKAGVI MAAKAGVI AAAKAGVI AAAKAGVI MAAAKAGVI MAAAKAGVA MASKAGVA MASKAGVA MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKAGAVH MASKGAVH MASKAGAVA MASKAGAVA	160 GMARGLARDFGPF GLAKGLARDFGPF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GPTKSMAREVASF GTKSMAREVASF GTKSLAREIGSF GLTRTLALELGRW GPTRALATELGRW GPTRALATELGRW GTTRALAREWASF GLTRTLAEMWASF GLTRALAREWASF GLTRALAREWASF GLTRALAREWASF GLTRALAREWASF GLTRALAREWASF GLTRALAREWASF GLVGALAVELGAN GLV	170 GIT GUT GVT GVT GVT GUT GIT GIT GIT GIR GVR GVR GVR GUR
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3NUG 3NUG 3NUG 3NUG 3NUG 3NUG 3NUG 3NUG 3AV4 2VDZ 3C 3AV4 2VDZ 3C 3AV4 2V2G 3AV4 Bdca 3V2G 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3V2G 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3RSH 3TZK 3TZK 3TZK 3TZK 3TZH 3DTH	110 VNPRAPF VNPRAPF TNLTSFF TNLTSFF TNLTSFF VNLTGFF VNLTGFF VNLTGFF VNLTGFF VNVVQPG VNLTSAF VNLTSFF VNLTSFF VNLTSFF VNLTSFF VNLTSFF VNLTSFF VNLTVPG	HASVEA VAIRESA VAIRESA RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAP IVTRAGT NSCQAV WASRAF LAAKYO VASRAP SCAV SCAV SCAV SCAV SCAV SCAV SCAV SCAV	I 2 0 REM I GD . REHLGD . RGM MK K RGM MK K RGM MK K RGM MK K RGM MK K RGM MK K RGM MK K LO RAM VAR RAM VAR	RQG RQG RQG RQG RQG RQG RQG RQG RQG RGG GGAG GGA	130 RILITIG RIITIG RIINVG RIINVG RIINVG RIVNVG RIVNIG SIVLTA VIVAIG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG AIVNLG	14 SVNGDRM SNLAELVI VVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAG SVVG SVV	10 10 10 10 10 10 10 10 10 10	150 CANARAS KALA CAAAKAGVI CAAAKAGVI CAAAKAGVI CAAAKAGVI CAAAKAGVI CAAAKAGVI CAAAKAGVI CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAAAKAGVA CAACACAA CACACAACAA CACACAACAACAACAACA	160 GMARGLARDFCPF GLTKGLARDLCPF GLTKGLARDLCPF GFTKSMAREVVASF GFTKSMAREVVASF GFTKSMAREVVASF GFTKSMAREVVASF GFTKSMAREVVASF GTTKSMAREVVASF GLTRTLALELGRN GTTRALAREVVASF GLTRTLALELGRN GTTRALARELGSN GLTRTLALELGRN GTTRALAREWASF GLTRTLALAREWASF GLTRTLALAREWASF GLTRTLALST GLTRALAREWASF GLTRTLALST GLTRALAREWASF GLTRTLALST GLTRALAREWASF GLTRTLAGT GLTRALAREWASF GLTRTLAGT GLTRALAREWASF GLTRTLAGT GLTRALAREWASF GLTCGNMHTT GLTST GLT	170 GIT GUT GVT GVT GVT GUT GIT GIT GIR GVR GVR GVR GVR GVR GVR GVR GVR GVR GV
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3WD 2WDZ 3LQF 3TOX 3LQF 3TOX 3AK4 consensus>50 Bdca 3V2G 3RSH 3RSH	110 1NTHAPT VNFRAPF TNLTSIF TNLTSIF TNLTSIF TNLTSIF VNLTGFF VNLTGFF VNVDGHF VNVDGHF VNVDGHF VNVDGHF VNVDGHF VNVDFF VNVDFF VNVDFF VNVDFF VNVDFF VNTVAFG	HASVEA VAIRSAS VAIRSAS KISKAVI RLSKAVI RLSKAVI RLSKAVI RISKAVI RSCOAVC WASRAS I VTRAGG VTRAGG VARAGG NASSAS STDTDAN STDTDAN STDTDAN FIETDMI	ROMPE. RHLGD. RGMMKK RGMKK RGMKK RAK RGMKK RGMKK RAK RGMKK RAK RGMKK RAK RGMKK RAK RAK RAK RAK RAK RAK RAK R	GG RQG RQG RQG RQG RSG RSG GAGG GAGG GAGG GAGG GAGG GAGG GAGG GAGG GAGG GGAGG GGAGG GGG GGG RQGG 	130 RILITIG RILITIG RIINVG RIINVG RIINVG RIINVG RIINVG AIVNIG	14 SVNGDRMI SVVGTMGI SVVGTMGI SVVGTMGI SVVGTMGI SVVGSAG SVVGSAG SVVGSAG SVVGSAG SVVGSAG SVVGSAG SVVGSAG SVVGTMGI SVVGTMG	10, WPGI.8, AGO.4,	150 ANAASKSALQ SASKAALA VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAKAGVA VAAKAGVA VAAKAGVA VAAKAGVA VAAKAGVA MASKAAVH MAS	160 GMAR GLARD JCGPF GGT KGLARD LGGPF GFTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GTTKSMAREVASS GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTRALAFELGSK GTTALAAEWAGS GTT	170 GIT GGVT GGVT GGVT GGVT GGVT GGVT GGVT
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZH 3OP4 4DMM 1ULS 3WUG 3AWD 2WDZ 3LQF 3TOX 3AK4 consensus>50 Bdca 3W2G 3RSH 3Y2G 3RSN 3RSK 3TZK 3TZK	110 TITHAP VIPRAP TITSIF TITSIF TITSIF TITSIF TITSIF TITSIF TITSIF TITSIF V	I ASVE A VAIRSAS VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LSRAAR IVTRAST SCAV WASRAS LAAKYOV JANOIA JANOIA STDTDMN FIETDMM FIETDMM	REOMPE. REGMAKK RGMAKKK RGMAKK RGMAKKKK RGMAKKKK RGMAKKKK RGMAKKKK RGMAKKKK RGMAKKKKK RGMAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK		130 RILIG RILING RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIING RII	14 SVNGDRM SNLAELVI SVUGTMG SVUGTMG SVUGTMG SVUGTMG SVUGTMG SVUGSAG SVUSSAG SV	10 10 10 10 10 10 10 10 10 10	150 AMAASKSALQ SASKAALA GAAAKAGVI GAAAKAGVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAKAGVI MAAKAGVI MASKAGVI MASKAGVI MASKAGVI MASKGAVI MASKAGVI MASKGAVI MASKAGVI M	160 GMARGLARDJCGPT GLAKGLARDJCGPT GLAKGLARDJCGPT GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GTTKSMAREVASF GLTRTLALELGRM GFTRALATELCGN GTTRLAREMASF GLTRTLALELGRM GTTRLAREMASF GLTRTLAREMASF GLTRTLAREMASF GLTCALAREMASF	170 GIT GGVT GGVT GGVT GGVT GGVT GGVT GGVT
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZH 3PTP 4DMM 1ULS 3NUG	110 VN PRAPF VN PRAPF TN LTSIF TN LTSIF TN LTSIF TN LTSIF VN LTGFF VN LTGFF VN LTGFF VN LTGFF VN VDGMF VN VDGMF VN VDGMF VN VDGMF VN VVGPC VN VVQPC VN VVGPC VN VVQPC VN VVGPC VN VVQPC VN VVQC VN VVC VN VVC	HASVEA VAIRESAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAP IVTRAGT RSCAVC WASRAP LAAKYOV JANOJIAC STDTDAN STDTDAN STDTDAN FIETDMI FIETDMI	RANDAR REAL REAL REAL REAL REAL REAL REAL RE		130 RILIIG RILING RILINVG RIINVG RIINVG RIINVG RIINVG RIINVIG GIVNIG GIVNIG GIVNIG GIVNIG CIVNIG	14 SVNGDRMI SNLAELVI SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSMG SVVS SVVS	10 WPGI.S: MPGI.S: MAGO.A: AGO.A:	150 CAN ASKSALQ SASKAALA VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVA VAAXAGGVI VAASKAGVI SMASKGAVI VAASKAGVI SMASKGAVI VAASKAGVI SMASKGAVI SMASK	160 GMAR GLARD LCPF GLINK GLARD LCPF GLINK GLARD LCPF GFINK SMAREVIASF GFINK SMAREVIASF GFINK SMAREVIASF GFINK SMAREVIASF GFINK SMAREVIASF GLINT LALELIGAN GFINK SMAREVIASF GLINT LALELIGAN GFINK SMAREVIASF GLINT LALELIGAN GFINK SMAREVIASF GLINT LALELIGAN GFINK SMAREVIASF GLINT LALELIGAN GFINK SMAREVIASF GLINT LALE MACTON GFINK SMAREVIASF GLINT LALE MACTON GFINK SMAREVIASF GLINT LALE MACTON GINT ALLAREWACTON GLINT LALE MACTON GLINT LALE MACTON GLINT LEGN GLINT LEGN GLINT LGENT GLINT LGENT SMART GLINT GLINT LGENT GLINT LGENT GL	ITO GUT GUT GUT GUT GUT GUT GUT GUT GUT GUT
Bdca 3y2G 3RRO 3RSH 3TZH 3OP4 4DMM 1ULS 3NUG 3AWD 2WDZ 3LQF 3TOX 3AK4 consensus>50 Bdca 3y2G 3RSO 3RSH 3TZK 3TZK 3TZK 3TZK 3FTP	110 VIP RIAP F VIP RIAP F VI ITSL ITSL ITSL ITSL ITSL ITSL V	I ASVEA VAIRSAS VAIRSAS RISKAVI RISKAVI RISKAVI RISKAVI RISKAVI RISKAVI LSRAVI VARAS IVTRACT RSCOAVC WASRAC LAAKIYOV SAAC ISO VASRAC STOTDMA FIETDMA FIETDMA FIETDMA FIETDMA	REGNER RELIGD. RELIGD. REGMMXE RGMMXE RGMMXE RGMMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGMXE RGPM RA RGMXE RGPM RA RA RA RA RA RA RA RA RA RA RA RA RA		130 RILIIG RILINVG RIING RIIN	14 SVNGDRM SNLAELVI SVNGTMG SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAC SVVGSAC SVVGSAC SVVGSAC SVVGSAC SVVGTMG S	0 VVAGM.A. VVAGM.A. MapgI.S MacO.A. Mac	150 AMAASKSALQ USASKAAL MAAKAGVI MAAAKAGVI MAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI MAAKASKA SAAKAGVI MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKAPLAS NSAVAFLAS	160 GMARGLARDFGPF GLATKGLARDLGPF GPTKSMAREVASB GPTKSMAREVASB GPTKSMAREVASB GPTKSMAREVASB GPTKSMAREVASF GTKSMAREVASF GTKSLAREIGSF GLATRLARELGR GTRTALARELGR GTRTALARELGR GTRTALARELGR GTRTALARENAG GTTR	ITO GUTT GUTT GUTT GUTT GUTT GUTT GUTT GU
Bdca 3V2G 3RRH 3RSH 3TZK 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3NUG 3AWD 2WDZ 3LQF 3TOX 3AK4 consensus>50 Bdca 3V2G 3V2G 3RRO 3RSH 3TZH 3OP4 3FTP 4DMM	110 TITHAP VN PRAPI TITST TITST TITST TITST TITST VN VGPC VN VGPC VN VGPC VN VGPC VN VGPC VN VAPC VN VAPC	HASVEA VAIRESA VAIRESA VAIRESA VAIRESA VAIRESA VAIRESA VAIRA VAIRESA VAIRA LSKAVI VIRLSKAVI VIRLSKAVI VIRLSKAVI VISCAVC WASRAFC WASRAFC VISCAVC WASRAFC VISCAVC WASRAF VISCAVC	I 2 0 REOMPE. RELEDI REMIXER RGMMKK RGMKK RGMKK RGMMKK RGMKK RGMMKK RGKK RGK	GG GC RQG 	130 RILIIG RILITIG RILIVG RILIVG RILIVG RILIVG RILIVG RILIVG RILIVG RILIVG RILIVG RILIVG RILING RILI	14 SVNGDRM SNLAELVJ SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAG SV	10 10 10 10 10 10 10 10 10 10	150 AMAASKSALQ SASKAALA MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAKAGVI MASKAGVH MASKAGV MASKAGVH M MASKAGVH M MASKAGVH M MASKAGVH M M M	160 GMARGLARDJCGPF GLTKGLARDJCGP GTTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GFTKSMAREVASS GTTKSMAREVASS GLTRTLALELGRW GTRTALARELGSS GLTRTLALELGRW GTRALARELGSS GLTRTLALELGRW GFTRALARELGSS GLTRTLALELGRW GFTRALARELGSS GLTVGALAVELGAR GLTVGALAVELGSS GLVGALAVELGSS GLVGALAVELGSS GLVGALAVELGSS JUTGETLAV PEAAYIGETLAV PEAAYIGETLAV PEAAYIGETLAV PAAAYIGETLAV	ITO GUTT GUTT GUTT GUTT GUTT GUTT GUTT GU
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZK 3PTP 4DMM 1ULS 3NUG 3AMD 2WD2 3LOF 3TOX 3AK4 consensus>50 Bdca 3V2G 3RR0 3RSH 3TZK	110 VNFRAPF VN	THASVEA VAIRESAS VAIRESAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAA IVTRAGT RSCAVC WASRAPC WASRAPC WASRAPC V V V V V V V V V V V V V V V V V V V	REMARKS REMARKS REMARKS REGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK RGMAKK DOMRAA RAMVAR RAMVA		130 RILIIG RILITIG RILITIG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG HR HR HR HR HR HR HR HR HR HR	14 SVNGDRMI SNLAELVI SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAG SVVS SVGSAG SVVS SVGSAG SVVS SVGSAG SVVS SVVGSAG SVVS SVGSAG SVVS SVVGSAG SVVS SVGSAG SVVS SVVS	10 VV/AGM.A WPGI.S AAGO.A PPN.A PPN.A PPN.A PPN.A AAGV.A PPLIC.A PPLI.A PP.A PPLI.A PP.A COP.EE. COP.REI.	150 ANASKSALQ SASKALA VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVI VAAAKAGVA VAAKAGVA VA	160 GMAR GLARDICOP GLINGLARDICOP GLINGLARDICOP GFINSMAREVVAS GFINSMAREVVAS GFINSMAREVVAS GFINSMAREVVAS GFINSMAREVVAS GFINSMAREVVAS GLINTLALELGAR GFINSMAREVVAS GLINTLALELGAR GFINSMAREVVAS GLINTLALELGAR GFINSMAREVVAS GLINTLALAREWACS GLINTLALAREVVAS GLINTLALAREVVAS GFINSMAREV GFINSM	170 GUTT GUTT GUTT GUTT GUTT GUTT GUTT GUT
Bdca 372G 3RSH 372K 372H 30P4 3PTP 4DMM 1ULS 3NUG 3AWD 2WDZ 3LQF 3TOX 3AK4 consensus>50 Bdca 3R2G 3R72B 3R70 3R7	110 117 HAP V PRAP TL TISL TL TISL TL TISL TL TISL TL TISL TL TISL VN LGGVF VN LGGVF VN VL GGVF VN VDGM VN VDGM VN VDGM VN VL TISL VN VDGV VN VL TISL VN VV VD VD VD VL TISL VN VV VD VD VD VL TISL VN VV VD	I ASVE A VAIRSAS VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAP IVTRAS SCAVC WASRAS IVTRAS SCAVC WASRAS ISTOTDAN FIETDMM FIETDMM FIETDMM FIETDMM FIETDMM FIETDMM FIETDMM	REGMAK REGMAK REGMAK REGMAK REGMAK REGMAK REGMAK REGMAK REMA	GG: GG: 	130 RILIG RILIG RILINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG RIINVG AIVNIG AIVNIG AIVNIG AIVNLG AI	14 SVNGDRM SNLAELVI SNUGTMG SVVGTMG S	10. 10. 10. 10. 10. 10. 10. 10.	150 AMAASKSALQ LSASKAALA AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVA SAAKAGVI AAAKAGVA SAAKAGVI AAAKAGVA AYAAKGAVI SAAKAGVI AAAAKAGVI AAAAKAGVI AAAAKAGVI AAAAKAGVI AAAAAKAGVI AAAAKAGVI AAAAKAGVI AAAAKAGVI AAAAKAGVI AAAAKAGVI AAAAAKAGVI AAAAAKAGVI AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	160 GMARGLARDFGPT GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GTTKSMAREVASF G	170 GITT GGUT GGVT GGVT GGVT GGVT GGVT GGVT GG
Bdca 3V2G 3RRH 3RSH 3TZK 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3NUG 3AND 2WDZ 3COF 3TOX 3AR4 Consensus>50 Bdca 3V2G 3RSH 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 1ULS 3NUG 3AR4 3FTP 4DMM 3COF 3TZH 3COF 3TZH 3COF 3TZH 3COF 3TZH 3COF 3TZH 3COF 3TZH 3COF 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3RSH 3TZH 3COF 3COF 3RSH 3TZH 3COF 3COF 3COF 3RSH 3TZH 3COF	110 VM PRAPT VM PRAPT TN TST TN TST TN TST VM PRAPT VM PRAPT VM PRAPT VM PRAPT VM PRAPT VM PRAPT VM PRAPT VM VM VM PRAPT VM VM VM PRAPT VM VM VM PRAPT VM VM VM VM PRAPT VM VM VM VM PRAPT VM VM VM PRAPT VM VM VM VM PRAPT VM VM V	HASVEA VAIRESA VAIRESA RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAA IVTRAGT NSCQAVC WASRAFC WASRAFC WASRAFC BIDTDAN STDTDAN STDTDAN STDTDAN FIETDMI FIETDMI FIETDMI FIETDMI FIETDMI FIETDMI FIETDMI FIETDMI FIETDMI FIETDMI	I 2 0 REGMIX K RGMIX K RGMIX K RGMIX K RGMIX K RGMIX K RGMIX K RGMIX K RGMIX K COMMX K COMX	GG:GG:GG:GG:GG:GG:GG:GG:GG:GG:GG:GG:G	130 RILIG RILITIG RILITIG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG RILINVG AIVVNLG LTFTSS VIVAIG LTFTSSS VIVAIG LTFTSSS VIVAIG LTFTSSS VIVAIG LTFTSSS VIVAIG LT	14 SVNGDRM SNLAELVI SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAG SVVG SVV	10 10 10 10 10 10 10 10 10 10	150 AMAASKSALQ SASKAALA MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAKAGVA MSAAKAGVA MSAAKAGVA MSAAKAGVA MSAAKAGVA MSAAKAGVA MASKAGAYH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKAGV MASKGAVH MASKGAV MASKGAV MASKGAV MASKGAVH MASKGAV	160 GMARGLARDJCPF GLTKGLARDJCPF GLTKGLARDJCPF GFTKSMAREVJASF GFTKSMAREVJASF GFTKSMAREVJASF GFTKSMAREVJASF GFTKSMAREVJASF GTTKSMAREVJASF GLTRTLALELGRN GFTRALAREVJASF GLTRTLALELGRN GFTRALARELGSS GTTRLLALELGRN GFTRALAREVJSF GTTRSLAREVJSF GLTRTLALELGRN GFTRALAREVJSF GTTSS GTTSS GTTRTLALELGRN GFTRALAREVJSF GTTSS	170 GUTT GGUTT GGUT GGUT GGUT GGUT GGUT GG
Bdca 3V2G 3RR0 3RSH 3TZK 3TZH 3OP4 4D3M 1ULS 3NUG 3AWD 2WDZ 3LQF 3CONSENSUS 5O Bdca 3V2G 3RC0 3RSH 3TZK 3TZK 3TZK 3TZK 3TZH 3DY2 4DMM 1ULS 3WUG 3AWD 2WDZ 3LQF	110 VIP RAP VIP RAP VIP RAP VIT LISSI VITS	THASVEA VAIRESAS VAIRESAS KAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LCSRAAA IVTRAGT RSCAVC WASRAPC WASRAPC UVAKA SRAPC IVTRAGT STDTDAN STDTDAN STDTDAN STDTDAN STDTDAN FIETDMI	REGMYCH REGMYK REGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RGMYK RA RA RA RA RA RA RA RA RA RA RA RA RA		130 RILIIG RILING RIINVG RIING RII	14 SVNGDRM SNLAELV: SVNGTMG SVVGTMG	0 VVAGM.A. VVAGM.A. MVPGI.S MAGO.A. MPGO.A. MPE	150 AMAASKSALQ USASKALS MAAXASVI MAAAKAGVI MAAAKAGVI MAAAKAGVI MAAKASVI MAAKASVI MAS	160 GMAR GLARD FGPF GLATGLARD FGPF GLATGLARD FGPF GPTCSMAREVASB GPTCSMAREVASB GPTCSMAREVASB GPTCSMAREVASB GPTCSMAREVASB GTTCSMAREVASB GTTCSMAREVASB GLATTLALELGOW GPTRSMAREVASB GLATTLALELGOW GTTRALAREWAGF GLATALAREWAG	179 GITT GUTT GUTT GUTT GUTT GUTT GUTT GUTT
Bdca 3726 3789 3724 3724 3724 30P4 3FTP 4DMM 1ULS 3WUG 3AWD 2WDZ 3AWD 3AK4 consensus>50 Bdca 3R24 3728 3728 3724 30P4 3FTP 4DMM 3R4 3728 37788 3778	110 TITHAP VIPRAP TITISII TITISII TITISII TITISII TITISII TITISII TITISII TITISII TITISII TITISII VIVOPC VITISII VIVOPC VITISII VIVOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VIVOVC VICOPC VICOPC VICOPC VICOPC VICOPC VICOPCC VICOPC VICO	I ASVE A A VAIRSAS VAIRSAS RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LSRAAR IVTRAS SCAV WASRAS LAAKYOV WASRAS LAAKYOV WASRAS LAAKYOV JANOITA STDTDM FIETDMI	REGMER REGMERT REGMERT REGMERT REGMERT REGMERT REGMERT REGMERT REGMERT REGMERT REGMERT REGMERT REME REAMERT REAME	GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	IJOG RILIIG RILINVGG RILINVGG RILINVGG RILINVGG RILINVG RILING RILINVG RILING RI	14 SVNGDRM SNLAELVI SVUGTMG SVUGTMG SVUGTMG SVUGTMG SVUGTMG SVUGTMG SVUGSAG SVUSSAG SV	10 10 10 10 10 10 10 10 10 10	150 HAASKSALQ CAAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVI AAAKAGVA SAAKAGVI SAAKAGVI SAAKAGVI SAAKAGVI SAAKAGVI AYAAKGGVI HS AYAAKGGVI HS AYAAKAGVI AYAAKGGVI HS AXAYAAKGGVI HS AXAYAAKAGVI AGAVAWLAG AGAVAWLAG AGAVAWLAG AGAVAFLAS ASAVAFLAS	160 GMARGIARDJCGPF GLARGIARDJCGPF GLARGIARDJCGP GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GFTKSMAREVASF GTTKSMAREVASF GTTSLAREVASF	179 GITT GGUTT GGUTT GGUTT GGUTT GGUT GGUT
Bdca 3V2G 3RR0 3RSH 3TZK 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3MUG 3AND 2WDZ 3TOX 3AR4 2VDZ 3CON 3AR4 3V2G 3RSH 3TZK 3TZK 3TZK 3TZK 3TZK 3TZK 3TZK 3TZH 3OP4 3FTP 4DMM 1ULS 3MUD 3AND 2WDZ 3LQP 3TOX 3AR4	I I O VI THAP VI THAP TH THIS I THIS	HASVEA VAIRESA VAIRESA RLSKAVI RLSKAVI RLSKAVI RLSKAVI RLSKAVI LSKAVI LSKAVI LSKAVI VRLSKAVI LVAKAAS IVTRAGT SCAVY WASRAFC WASRAFC VMASKAVI SCAVY VASKAVI FIDTDAN STDTDAN FIETDMI	I 2 0 REG MY E REG MY K RG MY K K RG MY K K K RG MY K K K K K K K K K K K K K K	CGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	I J O I L I G R I L I TIG R I L I TIG R I L I V G R I L I V G R I L I V G R I L I V J R I L N V G R I L V I J R I I N V G R I I N V G R I I N V G R I I N V G I L T F T I A I L R V I J I L R V I L R V I J I L R V I L R V I L R V I L R V I L R V I L	LARY SVNGDRM SNLAELVI SVVGTMG SVVGTMG SVVGTMG SVVGTMG SVVGSAG SVVG SVV	10 10 10 10 10 10 10 10 10 10	150 CAN A S K S A LO S A S K A A LA A A A K A S K S A LO A A A K A S K I A A A K A S K I A A A K A S K I M A A K A S K A M S A A K A S K A M S A A K A S K A M S A K A S K A S K A M S K A S K A S K A S K A M S K A S K A S K A S K A M S K A S K A S K A S K A S K A M S K A	160 GMARGLARDJCPF GLTKGLARDJCPF GLTKGLARDJCPF GTKSMAREVVASF GFTKSMAREVVASF GFTKSMAREVVASF GFTKSMAREVVASF GFTKSMAREVVASF GTTKSMAREVVASF GTTKSMAREVVASF GTTKSMAREVVASF GTTRSMAREVVASF GLTRTLALELGRN GTTRALAREVVASF GLTRTLALELGRN GTTRALAREVVASF GLTRTLALELGRN GTTRALAREVVASF GLTRTLALELGRN GTTRALAREVVASF GLTRTLALELGRN GTTRALAREVVASF GLTRTLALEGNN GTTRALAREVVASF GLTRTLATELGRN GLTRTLATELGNN GTTRALAREVVASF GLTCTTLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV PEAAYITCETLEV DAASYVTCAILAV DAASYVTCAILAV PAASYVTCAILAV	170 GITT GUTT GUTT GUTT GUTT GITR GIRR GIR GIR GIR GIR GIR GIR GIR GIR G



Figure S1. Sequence alignment of BdcA with top hits from DALI²¹ search.



Figure S2. Alignment of BdcA with structurally similar proteins show a 32% identity with 3SJ7 and a 34% identity with 3OP4



Figure S3. PCR amplification of the BdcA gene from final plasmid product followed by gel electrophoresis verifies that the gene was successfully cloned. Multiple aliquots are shown.

Cloning	Forward	GACTGGATCCATGGGCGCTTTTACAGGTAAGACA
Cloning	Reverse	GACTCTCGAGCGTAGTCGGTTATGCGCCAAA
G14V	Forward	CAGTTCTCATCCTCGGTGTCAGTCGTGGTATCGGTG
	Reverse	CACCGATACCACGACTGACACCGAGGATGAGAACTG
	Forward	CTAAACGCCTGGCACAACAGACTGGAGCGACAGC
ESUQ	Reverse	GCTGTCGCTCCAGTCTGTTGTGCCAGGCGTTTAG
	Forward	CGACAGCAGTATTCACAGCTAGTGCTGACAGAGACGC
D39A	Reverse	GCGTCTCTGTCAGCACTAGCTGTGAATACTGCTGTCG
	Forward	GTATTGGCGTCTTTGGCGCGGCCCTGGAATTAAATG
E90A	Reverse	CATTTAATTCCAGGGCCGCGCCAAAGACGCCAATAC
D136T	Forward	GGCTCCGTGAATGGCACTCGTATGCCTGTTGC
	Reverse	GCAACAGGCATACGAGTGCCATTCACGGAGCC
	Forward	GCTTATGCCGCCAGCATATCTGCCCTGCAAGG
K150I	Reverse	CCTTGCAGGGCAGATATGCTGGCGGCATAAGC
	Forward	GCCAGGGCCAATTGCTACCGACGCTAATC
D180A	Reverse	GATTAGCGTCGGTAGCAATTGGCCCTGGC

Table S1. Primers used for cloning and site-directed mutagenesis, listed 5' to 3'



Figure S4. Products from each step of the protein purification process. Lane 1: Ladder; Lane 3: pellet after large cell culture; Lane 5: cell lysate after homogenization; Lane 6: Column wash; Lane 8: collected fraction from the column containing purified protein.