Purification of *Escherichia coli* YoaA, a Putative Helicase

Master’s Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Biochemistry
Dr. Susan Lovett, Advisor

In Partial Fulfillment
of the Requirements for the Degree

Master of Science
in
Biochemistry

by
Mark Gregory

May 2019
Acknowledgements

I would like to thank my advisor Dr. Susan Lovett. Thank you for sharing with me the incredible opportunity to work in her laboratory. I am grateful for your patience, optimism, encouragement and knowledge throughout the entirety of my work at your laboratory.

Thank you to my mentor Vincent Sutera. Thank you for the time and support you gave me in allowing me to develop as a scientist. You instilled in me the importance of being organized in my thinking when carrying out my experiments and devising the next plans of action. These are qualities that go beyond the laboratory setting and will stay with me for the rest of my life.

I would also like to thank everyone who worked with me in the Lovett lab. You were all very welcoming and provided a supportive environment during my experience.

Thank you to all my friends. Special thanks to Elena for motivation and support.

Thank you to the faculty in the Biochemistry Department for nourishing my love of biochemistry throughout my research.

Thank you to my family for encouraging me and providing me with the support to always be ambitious.
ABSTRACT

Purification of *Escherichia coli* YoaA, a Putative Helicase

A thesis presented to the Department of Biochemistry

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Mark Gregory

All cells must maintain their genomic integrity to survive. Several repair mechanisms necessitate unwinding the damaged DNA to allow the recruitment of repair factors. Unwinding activity is achieved by helicases which temporarily destabilize the DNA duplex. When this function fails, human diseases such as Werner Syndrome, Xeroderma Pigmentosum, and Trichothiodystrophy can result. The role of helicases is integral to DNA repair in prokaryotic organisms as well. In *Escherichia coli* (*E. coli*), helicases including the RecBCD complex, RecQ, and DinG are utilized to unwind regions of double stranded DNA to make it accessible to repair factors. The SOS response is a system of inducing genes that aid in repairing damaged DNA in *E. coli*. Though this system has been well characterized, the picture may be incomplete. YoaA is a putative helicase because it shares conserved sequences with damage-inducible helicase DinG and has been shown genetically to be involved in DNA repair. The goal of this study is to develop a system that will allow the purification of YoaA. In our original system, YoaA expression in a high copy plasmid resulted in cell toxicity. In the expression system presented herein, YoaA was fused to a N-terminal RGS-His₆ tag (6XHisYoaA) and overexpressed in a low-copy vector under the *lacZ* promoter. This resulted in a stable form of YoaA. To purify
6XHisYoaA, the Nickel-NTA, single-stranded-DNA, and MonoQ columns were chosen and tested based on the protein’s biochemical and ionic properties. The method that proved to be the most promising in allowing us to purify 6XHisYoaA was utilizing the Nickel-NTA column then the MonoQ column.
Table of Contents

List of Tables .......................................................................................................................... VIII
List of Figures ........................................................................................................................ IX
Introduction .................................................................................................................................. 1
Materials and Methods .............................................................................................................. 9
  Strain List: ............................................................................................................................... 9
    Table 1: *E. coli* K-12 Strains Made/Used in this Study ...................................................... 9
Media ........................................................................................................................................ 10
Gibson Assembly ..................................................................................................................... 11
Gateway System ..................................................................................................................... 11
Bacterial Transformations ....................................................................................................... 11
Plasmid Purification ................................................................................................................ 11
Drop Dialysis .......................................................................................................................... 12
Sequencing .............................................................................................................................. 12
Primers ...................................................................................................................................... 12
Gel Electrophoresis ................................................................................................................ 12
  Construction of pWSK29GWLacRGS6XHisYoaA-HolC and pWSK29GWT7RGS6XHisYoaA-HolC Plasmids ........................................................................................................... 13
Protein Overexpression ......................................................................................................... 14
Crude Lysate Extraction ........................................................................................................ 15
SDS-PAGE .............................................................................................................................. 15
Western Blots .......................................................................................................................... 16
Ni-NTA Column Purification ................................................................................................. 16
MonoQ Column Purification ................................................................................................. 17
ssDNA-cellulose Column Purification ................................................................. 18
Quantification of Plasmid DNA and Protein ......................................................... 18
Preparation of Electrocompetent Cells ................................................................. 18
Results ..................................................................................................................... 19
Discussion ................................................................................................................. 28
References ............................................................................................................... 34
List of Tables

Table 1: *E. coli* K-12 Strains Made/Used in this Study ................................................................. 9
Table 2: Primers Used .......................................................................................................................... 12
Table 3: Amounts of 6XHisYoaA Yielded from Ni-NTA Column Chromatography ........... 25
Table 4: Amounts of 6XHisYoaA Yielded from MonoQ Column Chromatography ........... 27
List of Figures

Figure 1: Structure of Azidothymidine ................................................................. 2
Figure 2: Architecture of the Pol III Holoenzyme ................................................... 3
Figure 3: Schematic of the Replisome ................................................................. 4
Figure 4: Sequence Alignment Between YoaA and DinG ....................................... 5
Figure 5: [4Fe-4S Cluster] in DinG ....................................................................... 6
Figure 6: Model for the Role of YoaA in Tolerance to AZT ..................................... 7
Figure 7: Plasmid Map of pWSK29GWLacRGSwXHisYoaA-HolC .......................... 20
Figure 8: SDS-PAGE and Western blot analysis of the crude lysate from MG1655 cells after overexpression of 6XHisYoaA with HolC under the lacZ promoter .............. 20
Figure 9: SDS-PAGE and Western blot analysis of the crude lysate from MG1655 cells after varying conditions of overexpression of 6XHisYoaA under the lacZ promoter .... 21
Figure 10: SDS-PAGE and Western blot analysis of the crude lysate from MG1655 and BL21 (DE3) cells after varying overexpression of 6XHisYoaA under the lacZ and T7 promoters ........................................................................ 22
Figure 11: Results of Pilot Batch Purification of 6XHisYoaA by Ni-NTA agarose affinity chromatography .................................................................................. 23
Figure 12: Results of Large-Scale Batch Purification of 6XHisYoaA by Ni-NTA agarose affinity chromatography ....................................................................... 24
Figure 13: Results of ssDNA-cellulose column purification following Ni-NTA agarose affinity chromatography ........................................................................... 26
Figure 14: Results of MonoQ column purification following Ni-NTA agarose affinity chromatography .................................................................................. 27
Introduction

All cells carry their genetic information in the form of DNA which exists predominantly as a double helix. The cell must be able to replicate and repair its DNA to faithfully transmit its genetic material to its progeny. Both processes require single stranded DNA (ssDNA) intermediates. *Escherichia coli* (*E. coli*) has a network of repairing damaged DNA known as the SOS response. This system has been well characterized ([1]), but there remain aspects of it yet to be discovered. DNA damage can occur from a variety of sources. Many of these sources lead to the formation of single stranded DNA (ssDNA) and include double stranded breaks and physical impediments during replication. The SOS response is stimulated by the accumulation of ssDNA which causes the RecA nucleoprotein to assemble on it ([2]) then promote the autodigestion of transcriptional repressor LexA ([3]). This relieves its repression of SOS genes and signals the induction of the SOS response. Mechanisms of repairing damaged DNA leading to ssDNA formation include utilizing the ssDNA in different ways to ultimately recruit RecA to the site of damage.

One such pathway that involves the RecA protein to resolve double stranded breaks is the RecBCD pathway. The formation of double strand breaks normally occurs with a very low rate of 0.01 breaks per genome ([4]). But, it is stimulated by stresses such as oxolinic acid ([5]) which inhibits DNA gyrase, an enzyme that relieves supercoiling. These double strand break intermediates are converted to ssDNA by several proteins in the RecBCD complex: RecC initially binds and forms a point of contact with the double strand break DNA, then
helicase/nuclease RecB and helicase RecD unwind the DNA at each end. As RecB unwinds the DNA duplex it concomitantly degrades the 5’ tail faster than the 3’ tail which is moved into its nuclease active site. RecC moves along the DNA until it encounters and binds to a chi sequence which inhibits RecB’s degradation of the 3’ tail causing RecB to degrade the 5’ tail only. This creates a 3’ tail for RecA binding ([6]) followed by the SOS response.

Stalling at the replication fork also generates ssDNA. This can occur due to physical blocks that impede DNA replication such as UV-induced pyrimidine dimers ([7]) and azidothymidine (AZT) which is a chain-terminating nucleotide that halts replication and so is toxic when not removed (Figure 1) ([8-9]).

Two proteins that have been shown to be involved in the general process of resolving stalled replication forks to resume replication are a monomeric 3’-5’ helicase RecQ and a ssDNA-specific 5’-3’ nuclease RecJ ([10-11]). Based on in vitro data RecQ unwinds the template DNA downstream of the replication fork and removes physical blocks, then switches to the lagging strand ([11]). The ssDNA that is generated is initially degraded by RecJ ([10]) which leaves a complementary region of ssDNA available for RecA recruitment ([11]). This stabilizes the ssDNA against further degradation ([12]) and induces the SOS response. The activity of helicases permits RecA access to the site of DNA damage which renders them crucial to DNA repair.

Figure 1: Structure of Azidothymidine.
Azidothymidine has an azide moiety at its 3’ carbon, which prevents elongation of the nascent DNA strand. (A) Thymidine (obtained from sigmaaldrich.com) (B) Azidothymidine, adapted from Elwell 1987
Several forms of DNA repair require helicases because they can transiently destabilize DNA for downstream processing. Helicases are enzymes that translocate along and unwind their DNA substrate. Consequently, much insight has been gained into the mobility of these proteins by modeling them as motors moving along a track ([13]). Helicases achieve their unwinding activity either passively by trapping transient ssDNA species or actively by directly coupling NTP hydrolysis to base-pair destabilization ([14]). The polarity of unwinding can either be 5’-3’ or 3’-5’ and differs among helicases due to their ability to associate with one strand of the DNA duplex ([15]). Helicases do not often function in isolation but rather associate with other proteins and factors to constitute macromolecular machines.

In *E. coli*, the replisome is the enzyme complex predominantly involved in DNA replication. The Polymerase III (Pol III) Holoenzyme is one of the functional units of the replisome (Figure 2). It is made up of two cores (α, *dnaE*; ε, *dnaQ*; and θ, holE), each associated with the β clamp and tethered to a single clamp loader complex (γ, *dnaX*; τ₂, *dnaX*; δ, holA; δ’, holB; χ, holC; ψ, holD) through a flexible linker by the τ subunit. The clamp loader complex

![Figure 2: Architecture of the Pol III Holoenzyme.](image-url)

The Pol III Holoenzyme is comprised of two Pol III cores (α, ε, and θ, two β clamps) and the clamp loader complex (γ, τ₂, δ, and δ, and accessory subunits χ and ψ). Adapted from O'Donnell 2006
loads onto a DNA template the β clamp which binds the Pol III core and anchors the enzyme to the DNA. This promotes the enzyme’s processivity. The χ and ψ subunits are accessory to the β clamp loading activity of the clamp loader complex but play a role in stabilizing it. To complete the initial formation of the replisome (Figure 3), the C-terminal domains of each τ subunit bind the homohexameric DnaB. This helicase encircles single stranded DNA while it unwinds the duplex and interacts with DnaG primase which continually initiates discontinuous lagging strand replication by placing a primer on the lagging strand at the replication fork (reviewed in [14]).

**Figure 3: Schematic of the Replisome.** The replisome is comprised of the Pol III Holoenzyme, DnaB helicase and DnaG primase. The leading strand is synthesized continuously by one Pol III core. The synthesis of the lagging strand by the other Pol III core is discontinuous, involving the continual re-initiation of replication by DnaG, whose activity is dependent on its preliminary interaction with DnaB at the replication fork. Adapted from O’Donnell 2006.

While the assembly of the replisome has been well characterized at the initiation of DNA replication, an understanding of its composition during the progression of the replication fork has started to emerge. In *E. coli*, the replisome is highly dynamic and flexible as it migrates along the DNA template. Mechanisms underlying the fluidity of the replisome during DNA replication

4
include displacing a downstream impeding RNA polymerase and utilizing the mRNA as a primer
during leading strand synthesis ([15]), exchanging subunits of the core and the clamp loader with
those free in solution ([16]), and by enlisting translesion DNA polymerases Pol II and Pol IV to
regulate helicase activity, thereby slowing fork progression as a possible means to give more
time for DNA repair ([17]).

Recent work has implicated HolC in DNA repair by resolving stalling during replication.
In a previous study, azidothymidine (AZT) was used to generate ssDNA gaps. The holC gene
was involved in the promotion of tolerance to AZT by interacting with yoaA. The sensitivity of a
holC deletion mutant to AZT and its growth deficiency were lessened when yoaA was also
absent, indicating that yoaA is toxic in the absence, but not in the presence, of holC ([20]).

The yoaA gene encodes YoaA, a 70.4 kDa monomer comprised of 636 amino acids. The
paralog of YoaA is E.coli DinG. The expression of DinG was shown to be repressed by high
levels of LexA, and the promoter sequence of dinG contains a LexA binding site, indicating that
it is a SOS gene ([21]). The DinG protein is a 5’-3’ Iron-Sulfur (Fe-S) helicase that is a member
of the Rad3 family of helicases ([22-25]). YoaA and DinG share conserved sequences (Figure 4).
The two proteins have a Walker A Box (A(G/P)TGTGKT), associated with phosphate binding, and a DEAH Box, associated with ATP hydrolysis. Also, YoaA has four putative Fe-binding sites consisting of four cysteines which DinG possesses. In the crystal structure of DinG ([26]), these four cysteine residues are covalently attached to a [4Fe-4S] cluster (Figure 5). Based on these conserved sequences, YoaA is thought to be a 5’-3’ helicase with a [4Fe-4S] cluster essential for its activity. The yoaA and holC genes interacted to promote AZT tolerance ([20]), and the two protein products physically interact which was determined by mass spectroscopic analysis of a pull-down assay ([27]). This led to a proposed model of the role of YoaA in DNA repair. In this model, HolC recruits YoaA to the site of AZT incorporation. YoaA subsequently unwinds the DNA duplex to facilitate the removal of AZT ([20]) (Figure 6). That YoaA plays an important role specifically in DNA repair owing to its putative helicase activity led to the interest in determining whether this DNA repair protein is an actual helicase.
The goal of this study is to purify YoaA. Following purification of the protein, helicase assays will be performed to assess whether the protein is a helicase and to determine the DNA substrates on which it is active. YoaA is natively expressed at low levels, likely because it plays an accessory role in DNA repair and is cytotoxic when overexpressed. How much YoaA is required for helicase assays has yet to be determined and most of the protein is lost due to purification. Therefore, a large-scale purification of YoaA must be performed to obtain a sufficiently large amount of the protein to assay.

Figure 6: Model for the Role of YoaA in Tolerance to AZT. (A) The incorporation of the azidothymidine monophosphate (AMP) chain-terminating nucleotide halts replication, leading to the formation of a ssDNA gap. (B) SSB binds to the gap and recruits subunits HolC and HolD. HolC recruits YoaA to the site of AMP incorporation. (C) YoaA unwinds the local DNA duplex, thereby permitting the removal of the nucleotide by the exonuclease (ε) subunit DnaQ. Adapted from Brown, 2015.
Another complication in purifying YoaA is in its stability. Previous attempts to purify YoaA utilized a high copy number plasmid, which is a plasmid that exists in several copies per cell. The original high copy plasmid pCA24N-6XHisYoaA expressed YoaA at levels too high for the cell to manage. This may have led to the formation of several truncations of the protein, as detected by immunoblotting. It is possible that the protein was degraded by the cell soon after being expressed to prevent its toxic accumulation. Stabilizing YoaA against degradation is essential for helicase assays to correctly attribute the helicase activity to the native and full-length protein. Therefore, expressing YoaA from a low copy plasmid may prevent it from accumulating to toxic levels. Additionally, *yoaA* was toxic when *holC* was absent, but not *holC* was present ([20]) and the two protein products are binding partners ([27]). This suggests that HolC may protect YoaA from degradation through their physical interaction by occluding it from the cell’s protein degradation machinery. In principle, the presence of HolC with YoaA could render YoaA stable for a sufficiently long amount of time that it will not be degraded during overexpression.

The *yoaA* gene was fused to a N-terminal RGS-His6 tag (6XHisYoaA) for purification and cloned with *holC* into a low copy plasmid. The resulting expressing vector is denoted as pWSK29GWLacRGS6XHisYoaA-HolC. It appears that low copy expression of YoaA might result in a stable form of the putative helicase because only the full-length form of the protein was present. As a first step in the purification of 6XHisYoaA, the crude lysate was passed over a Nickel-NTA (Ni-NTA) affinity chromatography column. This was an effective first step, consistent with previous work in the lab to purify 6XHisYoaA. I then tested two subsequent columns in the purification. The isoelectric point of the primary structure of 6XHisYoaA is 6.62, suggesting that it is predominantly negatively charged at around the physiological pH. Since
6XHisYoaA is anionic, I tested the MonoQ column as a next purification step because it is a positively charged and strong anion-exchanger. YoaA is a putative helicase so it is expected to bind ssDNA as it translocates along its dsDNA substrate. Thus, I chose the ssDNA-cellulose column as another possible subsequent purification step. The MonoQ column bound and released 6XHisYoaA. 6XHisYoaA bound to the ssDNA-cellulose column but failed to elute in the elution conditions tested. The purification method that was the most successful utilized the Ni-NTA column then the MonoQ column. A linear elution gradient of 6XHisYoaA from the two columns will be the next step in hopefully isolating the protein to homogeneity, then helicase assays will be performed on this DNA repair protein to determine if it is a helicase.

Materials and Methods

Table 1: *E. coli* K-12 Strains Made/Used in this Study

<table>
<thead>
<tr>
<th>Strain Number</th>
<th>Strain</th>
<th>Genotype</th>
<th>Plasmid</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>242</td>
<td>MG1655</td>
<td>F- rph-1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2168</td>
<td>BL21(DE3)</td>
<td>F- ompT hsdS(rB-mB-) dcm- gal λ (DE3)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>STL 2895</td>
<td>DH5α</td>
<td>F- phi80d lacZ ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 hsd17 deoR supE44 thi-1 λ- gyrA96 relA1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>7473</td>
<td>DB3.1</td>
<td>F- gyrA462 endA1 D(src-recA) mcrB mrr hsdS20(rB-MB-) supE44 ara-14 galK2 lacY1 proA2 rpsL20(Smr) xyl-5 lambda - leu mtl-1</td>
<td>N/A</td>
<td>Provided by the Gateway System</td>
</tr>
<tr>
<td>15503</td>
<td>DB3.1</td>
<td>F- gyrA462 endA1 D(src-recA) mcrB mrr hsdS20(rB-MB-) supE44 ara-14 galK2 lacY1 proA2 rpsL20(Smr) xyl-5 lambda - leu mtl-1</td>
<td>pDONR221</td>
<td>Provided by Invitrogen</td>
</tr>
<tr>
<td>21418</td>
<td>DB3.1</td>
<td>F- gyrA462 endA1 D(src-recA) mcrB mrr hsdS20(rB-MB-) supE44 ara-14 galK2 lacY1 proA2 rpsL20(Smr) xyl-5 lambda - leu mtl-1</td>
<td>pWSK29GWLac</td>
<td>Gateway fragment in orientation of Lac promoter</td>
</tr>
<tr>
<td>21419</td>
<td>DB3.1</td>
<td>F- gyrA462 endA1 D(src-recA) mcrB mrr hsdS20(rB-MB-) supE44 ara-14 galK2 lacY1 proA2 rpsL20(Smr) xyl-5 lambda - leu mtl-1</td>
<td>pWSK29GWT7</td>
<td>Gateway fragment in orientation of T7 promoter</td>
</tr>
<tr>
<td>21661</td>
<td>NEB-5α</td>
<td>fhuA2Δ(argF-lacZ)U169 phoA glvV44 ϕ80ΔlacZΔM15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</td>
<td>pDONR221RG</td>
<td>Isolate 1</td>
</tr>
<tr>
<td>ISOLATE</td>
<td>ORIGIN</td>
<td>STRAIN CODE</td>
<td>STRAIN FEATURES</td>
<td>PLASMID</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>21699</td>
<td>MG1655</td>
<td>F-rph-1</td>
<td></td>
<td>pWSK29GWL</td>
</tr>
<tr>
<td>21700</td>
<td>MG1655</td>
<td>F-rph-1</td>
<td></td>
<td>pWSK29GWL</td>
</tr>
<tr>
<td>21987</td>
<td>NEB-5α</td>
<td>fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</td>
<td>pWSK29GWL</td>
<td>cRGS6XHisYo aA-HoIC</td>
</tr>
<tr>
<td>21990</td>
<td>NEB-5α</td>
<td>fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</td>
<td>pWSK29GWL</td>
<td>cRGS6XHisYo aA-HoIC</td>
</tr>
<tr>
<td>22032</td>
<td>DH5α</td>
<td>F- phi80d lacZ ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 hsd17 deoR supE44 thi-1 λ- gyrA96 relA1</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22033</td>
<td>DH5α</td>
<td>F- phi80d lacZ ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 hsd17 deoR supE44 thi-1 λ- gyrA96 relA1</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22034</td>
<td>DH5α</td>
<td>F- phi80d lacZ ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 hsd17 deoR supE44 thi-1 λ- gyrA96 relA1</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22035</td>
<td>DH5α</td>
<td>F- phi80d lacZ ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 hsd17 deoR supE44 thi-1 λ- gyrA96 relA1</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22036</td>
<td>BL21(DE3)</td>
<td>F-ompT hsdS(rB-mB-) dcm- gal λ (DE3)</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22037</td>
<td>BL21(DE3)</td>
<td>F-ompT hsdS(rB-mB-) dcm- gal λ (DE3)</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22038</td>
<td>BL21(DE3)</td>
<td>F-ompT hsdS(rB-mB-) dcm- gal λ (DE3)</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
<tr>
<td>22039</td>
<td>BL21(DE3)</td>
<td>F-ompT hsdS(rB-mB-) dcm- gal λ (DE3)</td>
<td>pWSK29GWT7</td>
<td>RGS6XHisYoa A-HoIC</td>
</tr>
</tbody>
</table>

**Media**

Luria-Bertani (LB) media was used for bacterial strains. The LB media was made utilizing 1% Bacto Tryptone, 0.5% Bacto Yeast Extract, and 0.5% NaCl (% w/v). Ampicillin LB media was made to a final concentration of 100 µg/mL from a stock concentration of 50 mg/mL. Kanamycin LB media was made to a final concentration of 30 µg/mL from a stock concentration of 30 mg/mL.
Gibson Assembly

The Gibson Assembly protocol from NEB was followed. Other members in the lab assembled the 6XHis-\(\text{yoaA}\) and \(\text{holC}\) gene fragments to produce the 6XHis-\(\text{yoaA-holC}\) dual construct (primers referenced in Table 2).

Gateway System

The BP Clonase II and LR Clonase II recombination reaction protocol from Invitrogen was followed with the exception that no Proteinase K was added to the solution to halt the reaction. The 6XHis-\(\text{yoaA-holC}\) dual construct was cloned into pDONR221 using attB1 and attB2 sites (primers reference in Table 2). LR Clonase II reactions then cloned \(\text{yoaA-holC}\) from the resulting pDONR221RGS6XHisYoaA-HolC plasmid into either pWSK2GWLac or pWSK29GWT7, which contain the Gateway fragment in either the orientation of the \(\text{lacZ}\) or T7 promoter, respectively.

Bacterial Transformations

The Dower \textit{et al} 1988 protocol was followed to transform electrocompetent MG1655, DH5-\(\alpha\) and BL21 (DE3) cells utilizing the BioRAD Micropulser set for 0.2 cm cuvettes and 2.5 kV. Following transformation, cells were rescued for 1 hour at 37 °C in the roller drum then plated on the appropriate selective media. Single colony isolates were picked for plasmid purification.

Plasmid Purification

The GeneJet Plasmid Miniprep Kit protocol from Thermo Scientific was followed to extract bacterial plasmids.
Drop Dialysis

Following BP Clonase II and LR Clonase II reactions, drop dialysis against ddH$_2$O was performed. A volume of 3 µL of the reaction mixture was loaded onto a 0.025 µm Millipore Membrane Filter (Thermo Scientific), and dialysis was allowed to take place for 2 minutes for every 1 µL of sample.

Sequencing

DNA was sequenced through Genewiz Sanger Sequencing according to their protocol. The pWSK29GWLac and pWSK29GWT7 plasmids were previously constructed by Vincent Sutera using the Gateway Vector Conversion System (Invitrogen). Primers are listed in Table 2.

Primers

Primers were generated from Sigma Aldrich. All primer sequences are in Table 2.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HisyoaAattB1FORD UAL (Forward)</td>
<td>GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGAACCATGAG</td>
</tr>
<tr>
<td>yoaAREVBegholCD UAL (Reverse)</td>
<td>TGTTGTCATTGTCCAGAAGGTTAGACGCAGTCGCGTTTTTCATTACCTGGAGGAT</td>
</tr>
<tr>
<td>NATHolCFORDUAL (Forward)</td>
<td>CCGTACCGGTTTTCTTGGATACCATCCTCCAGGTAATGAAAAACGCGACA</td>
</tr>
<tr>
<td>NATHolCREVattB2DUAL (Reverse)</td>
<td>GGGGACACTTTGTACAAGGAAGCTGGGTTACTTATTATTTCCAGGGTTCGTA</td>
</tr>
<tr>
<td>VASyoaA515Fseq</td>
<td>ACTGCCCGATGTATAAAGATTGCTTTTGTGG</td>
</tr>
<tr>
<td>DB 10 YoaA b817R</td>
<td>CAGTCGCTGCACAGACAGTTTGTCC</td>
</tr>
<tr>
<td>VASyoaA1120Fseq</td>
<td>CCGTCACGCGGCCAGAATTTCAAGAG</td>
</tr>
<tr>
<td>M13F</td>
<td>GTAAAACGACGGCCAG</td>
</tr>
<tr>
<td>M13R</td>
<td>CAGGAAAACGCTATGAC</td>
</tr>
</tbody>
</table>

Gel Electrophoresis

A 1.5% Agarose gel was used to confirm the identity of extracted plasmids. Each gel was prepared according to the protocol from New England Biolabs, loaded with a 2-log DNA ladder (0.1-10 kb) from NEB, and loading dye (4% sucrose, 0.25% Bromophenol blue) was added to
each sample. The gel was run at 100 Volts for 75 minutes, stained in 1 µg/mL ethidium bromide in 1X TAE for 10 minutes on a shaker, then detained for 10 minutes.

**Construction of pWSK29GWLacRGS6XHisYoaA-HolC and pWSK29GWT7RGS6XHisYoaA-HolC Plasmids**

The 6XHis-yoaA and holC gene fragments were generated as the first step towards constructing the pWSK29GWLacRGS6XHisYoaA-HolC and pWSK29GWT7RGS6XHisYoaA-HolC plasmids. Bacterial colony PCR utilizing MG1655 WT E.coli was used to construct the 6XHis-yoaA and holC fragments. The genomic template was prepared by resuspending the edge of a single colony in 10 µL doubly-distilled H₂O (ddH₂O), microwaving for 30 seconds at 1,000 Watts, and placing on ice. PCR of 6XHis-yoaA was performed with 2 units of Phusion DNA Polymerase (NEB), 5X Phusion HF Buffer to a final concentration of 1X, 10 mM dNTPs (NEB) to a final concentration of 10µM, 10 µl microwaved lysed colony template, 100 µM HisyoaAattB1FORDUAL and 100 µM yoaAREVBegholCDUAL each to a final concentration of 5 µM in a final volume of 50 µL. The PCR cycle for 6XHis-yoaA was as follows: initial denaturation at 98°C for 2 minutes, secondary denaturation at 98°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 10 minutes. The PCR cycle for 6XHis-yoaA was repeated 24 times with a final extension at 72°C for 5 minutes. PCR of holC was performed in the same way as for 6XHis-yoaA with the exception that NATholCFORDUAL and NATholCREVattB2DUAL were used as the primers. The PCR cycle for holC was the same as for 6XHis-yoaA with the exception that the extension step was 1 minute, 15 seconds. Following PCR, the 6XHis-yoaA and holC fragments were purified following the GeneJET PCR Purification Kit protocol (Thermo Scientific).
The 6XHis-yoaA and holC fragments were assembled to produce 6XHis-yoaA-holC following the Gibson Assembly protocol from NEB. The 6XHis-yoaA-holC product was purified following the GeneJET PCR Purification Kit (Thermo Scientific) and cloned into pDONR221 following the BP reaction protocol (Invitrogen). 3 µL of the BP reaction was drop dialyzed for 2 minutes for every 1 µL of sample, and the pDONR221RGS6XHisYoaA-HolC plasmid was transformed into NEB® 5-alpha E. coli (NEB). 6XHis-yoaA-holC was then cloned from pDONR221RGS6XHisYoaA-HolC into pWSK29GWLac or pWSK29GWT7 following the LR Clonase II reaction protocol (Invitrogen). The LR Clonase II reaction was drop dialyzed as previously described, then the LR Clonase II reaction product pWSK29GWLacRGS6XHisYoaA-HolC or pWSK29GWT7RGS6XHisYoaA-HolC was transformed into electrocompetent DH5α E. coli. The pWSK29GWLacRGS6XHisYoaA-HolC plasmid was extracted then transformed into electrocompetent MG1655 WT E. coli for overexpression. The same was done for pWSK29GWT7RGS6XHisYoaA-HolC plasmid except that it was transformed into home-made electrocompetent BL21 (DE3) E. coli. All electrocompetent cells were prepared following the Dower et al 1988 protocol.

**Protein Overexpression**

MG1655 WT E. coli containing pWSK29GWLacRGS6XHisYoaA-HolC, pWSK29GWT7RGS6XHisYoaA-HolC or the empty pWSK29 vector were grown shaking in LB media at 37°C with the appropriate antibiotic. The OD$_{600}$ was monitored using a BioRad spectrophotometer. When the OD$_{600}$ reached 0.6, cells were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 2 hours. The cells were then centrifuged at 4700 rpm for 30 minutes at room temperature, concentrated 1:100 in Tris-Sucrose (50 mM).
Tris(hydroxymethyl)aminomethane, 10% w/v sucrose, pH 7.5), flash frozen in dry ice with 95% ethanol, and stored at -80°C.

**Crude Lysate Extraction**

In the crude lysate extraction all steps were carried out on ice or at 4°C unless indicated otherwise. Cells were thawed from -80°C freezer on ice. For every 100 µL of cells, final concentrations of 1µM DTT and 0.1 mg/mL lysozyme (United States Biochemical) in Tris-Sucrose were added. After a 5-minute incubation, a final concentration of 0.25M NaCl was added, followed by an additional 25-minute incubation. The cells were then heat shocked at 37°C then rescued on ice, and this was repeated twice. For small volumes of crude lysate (less than 5 mL), the heat shock and rescue steps were 15 seconds and 30 seconds, respectively. For larger volumes of crude lysate (5 mL or more), the heat shock and rescue steps were 1 minute and 2 minutes, respectively. Following heat shock, the lysed cells were centrifuged at 12,000 rpm at 4°C, and the CL supernatant was collected. For small and large volumes of CL, the centrifugations lasted 15 minutes and 40 minutes, respectively.

**SDS-PAGE**

The Running buffers and polyacrylamide gels were prepared following the protocol in Laemmli et al. Nature 227, 680 1970. An exception was that the ammonium persulfate (Fisher BioReagents) was prepared at a concentration of 20% (v/v) instead of 10%. After the separating gels were poured, a thin layer of water-saturated butanol was applied on top. Gel ladders, gel casings, and plates were obtained from BioRad. All fractions analyzed by SDS-PAGE were combined with an equal volume of 2X FSB (0.125M Tris pH 6.8, 4% SDS 20% Glycerol, 1.43M Beta Mercaptoethanol, 1mg/ml Bromophenol), boiled for 2 minutes, and loaded onto 12% polyacrylamide gels. Molecular weight protein standards (BioRad Protein Precision Plus dual
color) were also loaded into each gel. The gels were run at 250 volts and 400 milliamps for 50 minutes. For Coomassie staining, gels were stained with Coomassie stain (0.2% Coomassie blue R-250, 40% Methanol, and 10% glacial acetic acid) for 15 minutes. The gels were then washed with high destain (40% methanol and 10% glacial acetic acid) twice for 15 minutes each time then once for 15 minutes to overnight in low destain (10% methanol and 5% glacial acetic acid).

**Western Blots**

Following SDS-PAGE, protein gels were equilibrated in 1X transfer buffer (Thermo Scientific) for 15 min before the transfer. A Polyvinylidene difluoride (PVDF) membrane (Perkin Elmer Life Science) was immersed in methanol, ddH₂O, then 1X Transfer buffer twice for 30 seconds each time. Transfer cassettes (Bio-Rad) were prepared following the protocol outlined in the “BioRad Mini Trans-Blot Cell Assembly and Preparation for Transfer” document. The transfer step ran at 100 volts and 300 milliamps for 75 minutes. After transfer, a Western Blot was performed according to the QIAexpress Detection and Assay Handbook (Qiagen) with the following modifications: the primary Penta-His Tag Monoclonal Antibody (ThermoFisher) was incubated at a 1:2500 dilution in 3% BSA in TBS and the secondary and the ECL Anti-mouse IgG, peroxidase-linked whole antibody (GE Healthcare) was incubated at a dilution of 1:5000 in 10% nonfat milk in TBS, each for 1 hour. Detection was performed by mixing equal volumes of SuperSignal® West Pico Chemiluminescent Substrate (PIERCE) reagents and incubating for 2 minutes. The PVDF membrane was then imaged with a Bio-Rad imager 5 evenly spaced times between 10 – 300 seconds.

**Ni-NTA Column Purification**

500 mL of MG1655 WT *E. coli* containing pWSK29GWLacRGS6XHisYoaA-HolC were grown, induced with IPTG and concentrated in Tris-Sucrose as previously outlined.
Ni Sepharose 6 Fast Flow resin (GE Healthcare) was added to a Econo-Column® Chromatography Column, 1.0 × 5 cm (Bio-Rad) to a final column volume (CV) of 0.5 mL. The column was then equilibrated overnight in 40 CV of Ni Wash Buffer (50 mM NaH$_2$PO$_4$ pH 8, 500 mM NaCl, 20 mM Imidazole) at 4°C. Immediately following crude lysate extraction, the crude lysate was applied to the equilibrated column, and the crude lysate flow through (FT) was collected. 40 CV of Ni Wash Buffer was applied to the column, and the Wash FT was collected. 5 CV of 100 mM Imidazole Elution Buffer (50 mM NaH$_2$PO$_4$ pH 8, 500 mM NaCl, 100 mM Imidazole) was applied to the column, and the eluted protein was collected in five 1 CV fractions. The elution was repeated with 250 mM Imidazole Elution Buffer (50 mM NaH$_2$PO$_4$ pH 8, 500 mM NaCl, 250 mM Imidazole), then 500 mM Imidazole Elution Buffer (50 mM NaH$_2$PO$_4$ pH 8, 500 mM NaCl, 250 mM Imidazole).

The Nickel column purification of YoaA was repeated with the following modifications: 5 L of MG1655 WT E. coli containing pWSK29GWLacRGS6XHisYoaA-HolC were grown, a final CV of 1 mL of Ni Sepharose 6 Fast Flow resin and an Econo-Column® Chromatography Column, 2 × 5 cm (Bio-Rad) were used, and only 100 mM Imidazole Elution Buffer was used to elute protein, which was collected in ten 500 µL fractions.

**MonoQ Column Purification**

The three elution fractions with the largest amount of YoaA were pooled and dialyzed overnight against MonoQ Wash Buffer (10 mM Tris-HCl, 0.1 M NaCl pH 8.0, 10% ethylene glycol, 1 mM DTT) at 4°C. Q Sepharose High Performance resin (GE Healthcare) was added to an Econo-Column® Chromatography Column, 1.0 × 5 cm (Bio-Rad) to a final column volume (CV) of 0.5 mL. The MonoQ column was equilibrated overnight in 40 CV of MonoQ Wash Buffer at 4°C. The dialyzed sample was applied to the equilibrated column, and the FT was
collected. 40 CV of MonoQ Wash Buffer was applied to each column, and the Wash FT was collected. Protein was then eluted from the Q column with 5 CV of 1.5 M NaCl Elution Buffer (10 mM Tris-HCl pH 8.0, 1.5 M NaCl, 10% ethylene glycol, 1 mM DTT) and collected in five 1 CV fractions.

**ssDNA-cellulose Column Purification**

The ssDNA-cellulose (USB) was mixed with ssDNA-cellulose Wash Buffer (10 mM Tris-HCl, 0.1 M NaCl pH 8.0, 10% ethylene glycol, 1 mM DTT) to a final CV of 0.5 mL and added to an Econo-Column® Chromatography Column, 1.0 × 5 cm (Bio-Rad). The dialysis, column equilibration, and purification were performed in the same way as for the MonoQ column with the exception that the elution buffer was 2 M NaCl Elution Buffer (10 mM Tris-HCl pH 8.0, 2 M NaCl, 10% ethylene glycol, 1 mM DTT).

**Quantification of Plasmid DNA and Protein**

The concentration of purified plasmid DNA was determined using the Nanodrop (Thermo Fisher) according to the protocols indicated by the manufacturer. The protein concentration was determined using a Pierce™ Coomassie Plus (Bradford) Assay Kit following the instructions provided by the manufacturer.

**Preparation of Electrompetent Cells:**

Electrompetent MG1655 WT, DH5α and BL21 (DE3) *E.coli* were prepared by according to the Dower *et al* 1988 protocol with the following modifications. All steps were carried out at 4°C and the cells were centrifuged at 4700 rpm for 30 minutes. Cells were grown shaking in (LB) media at 37°C until the OD₆₀₀ reached 0.6. Then, the cells were centrifuged, then resuspended in 1 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0, and this was repeated once. After another centrifugation, cells were resuspended in 7% dimethyl sulfoxide
(DMSO), re-centrifuged, and concentrated to a final volume of 1 mL in 7% DMSO. The cells were flash frozen in dry ice with 95% ethanol and stored at -80°C.

Results

Expression of YoaA

Previous attempts to purify YoaA in a high copy expression vector were unsuccessful in singularly isolating it. The original high copy plasmid used was pCA24N-6XHisYoaA which was obtained from the ASKA collection. This high copy plasmid was utilized to express high levels and attempts to purify 6XHis YoaA were unsuccessful in isolating the protein through affinity chromatography techniques. The high copy overexpression seemed to generate several truncated forms of 6XHisYoaA, which were detected through immunoblotting, possibly because the protein was broken down due to its toxic accumulation. Since HolC binds to YoaA ([27]) and has been shown to suppress its toxicity ([20]), expressing the two proteins together could stabilize YoaA. Thus, a low copy co-expression vector of YoaA and HolC was constructed.

To prevent YoaA from building up to toxic levels and to stabilize it, it was expressed from the low copy plasmid denoted as pWSK29GWLacRGS6XHisYoaA-HolC. The gene yoaA was placed with holC under the lacZ promoter in the pWSK29 plasmid (Figure 7) by linking the two genes by a TAATG pentanucleotide so that the TAA stop codon of yoaA overlapped with the ATG start codon of holC by one nucleotide. When I overexpressed 6XHisYoaA in MG1655 E. coli, I detected it exclusively in its full-length form (Figure 8). By expressing 6XHisYoaA in this way, we expect a 1:1 ratio of expression of the two proteins, which could lead to fewer breakdown products. To this date, it is indeterminate whether they are being expressed in a 1:1 fashion, but from our experiment the population of 6XHisYoaA truncations disappeared.
Figure 7: Plasmid Map of pWSK29GWLacRGS6XHisYoaA-HolC. yoaA, fused to a RGS6XHis tag, was placed downstream the lacZ promoter and upstream of holC such that the stop and start codons of yoaA and holC, respectively, overlapped by one nucleotide.

A.  
B.  

Figure 8: SDS-PAGE and Western blot analysis of the crude lysate from MG1655 cells after overexpression of 6XHisYoaA with HolC under the lacZ promoter. MG1655 cells co-expressing 6XHisYoaA with HolC under the lacZ promoter were grown at 37°C to an O.D. of 0.6 and induced with 1 mM IPTG for 2 hours. (A) Coomassie-stained gel and (B) Penta-His-probed Western blot of 6XHisYoaA: lanes 1A and 4A, Isolates 1A and 4A, respectively, containing pWSK29GWLacRGS6XHisYoaA-HolC; lane M, Precision-Plus Protein Standard molecular weight ladder; lane E, Isolate containing the empty pWSK29 vector. Arrow indicates position of 6XHisYoaA.
To try to elucidate the optimal growing conditions for overexpression, the plasmid was grown in MG1655 *E. coli* at different temperatures and induced with IPTG for various times (Figure 9).

The ability of the low-copy expression system to produce 6XHisYoaA depended on temperature and the time of induction. At 37 and 30°C, the levels of expression of 6XHisYoaA were comparable, but at 24°C, the levels went down. At 24, 30, and 37°C, the levels of 6XHisYoaA present decreased as the time of induction increased, with a sharp fall overnight at 24°C, and the absence of 6XHisYoaA after overnight induction at 30°C. Expressing 6XHisYoaA under the lacZ promoter at 37°C for 2 hours led to stable expression. We wished to further increase expression while maintaining stability to avoid toxic accumulation. The T7 promoter is a strong promoter utilized for high levels of gene expression. We reversed the orientation of yoaA and holC by placing them downstream of the T7 promoter to possibly increase expression of 6XHisYoaA (Figure 10).

**Figure 9**: SDS-PAGE and Western blot analysis of the crude lysate from MG1655 cells after varying conditions of overexpression of 6XHisYoaA under the lacZ promoter. MG1655 cells containing pWSK29GWlacRGS6XHisYoaA-HolC 1A were grown to an O.D. of 0.6 and induced with 1 mM IPTG for different times (2 hours, 4 hours, overnight) and at different temperatures (24, 30, and 37 °C). (A) Coomassie-stained gel and (B) Penta-His-probed Western blot of 6XHisYoaA: lane M, Precision-Plus Protein Standard molecular weight ladder; for each of the other lanes, the top and bottom entries denote the induction times and temperatures, respectively. “O/N” denotes overnight induction.
Expressing 6XHisYoaA under the T7 promoter did not appear to negatively affect its stability, as it was detected exclusively in its full-length form. However, the level of expression did not noticeably increase when 6XHisYoaA placed under the control of the T7 promoter. Inducing low-copy expression 6XHisYoaA from the lacZ promoter for 2 hours at 37°C seemed to be the best condition, as changing to the stronger T7 promoter didn’t increase expression.

In all future experiments, the optimal conditions to obtain significant amounts of 6XHisYoaA were established by inducing pWSK29GWLaC6XHisYoaA-HolC 1A in STL 21699 (see Strain List) with IPTG for 2 hours at 37°C.

**Batch Purification of YoaA**

After establishing the conditions for the expression of 6XHisYoaA, the next step was to determine which column chromatography would be best for purification. In an initial pilot experiment, where 6XHisYoaA was expressed in half a liter, a Ni-NTA agarose affinity
chromatography column was used as an initial step in the purification. To determine a range of Imidazole concentrations to be used for purification, elution buffers containing 100mM, then 250mM, then 500mM Imidazole were tested and applied to the column (Figure 11).

6XHisYoaA was detected in Fractions 100 B, C, D, and E, with Fractions B and D having the greatest and smallest amount of the protein, respectively. It was clear in Fractions 100 B, C, D that there was separation of 6XHisYoaA from other proteins and collection of 6XHisYoaA. No 6XHisYoaA was detected in Fractions 250 A-E and 500 A-E. Thus, 100 mM Imidazole was
enough to elute 6XHisYoaA from the Ni-NTA column and separate it from other proteins.

To increase the yield of 6XHISYoaA, scaling up the overexpression was necessary. 6XHisYoaA was expressed in 5 L of STL 21699, and the protein was eluted with 100 mM Imidazole (Figure 12). The yield of 6XHisYoaA is shown in Table 3.

Figure 12: Results of Large-Scale Batch Purification of 6XHisYoaA by Ni-NTA agarose affinity chromatography. Shown are fractions of the purification of 6XHisYoaA over a Ni-NTA agarose column. (A), (C) Coomassie-stained gels and (B), (D) Penta-His-probed Western blots of 6XHisYoaA: lane M, Precision-Plus Protein Standard molecular weight ladder; lane CL, crude lysate; lane CL FT, crude lysate flow through; lane W FT, wash flow through; lanes 100 A-J, 100 mM Imidazole elution, fractions A-J, respectively.
6XHisYoaA was detected in Fractions 100 B-H. The early elution fractions 100 B, C, and D contained the greatest amount of 6XHisYoaA, but contained other protein species, whereas the later elution fractions 100 E, F, G and H were the most enriched in the protein, but contained a low absolute amount. In Fractions 100 B, 100 C, and 100 D, two other proteins with molecular weights between 50 kDa and 75 kDa copurified with 6XHisYoaA in comparable amounts, and other protein species ranging from around 37 kDa to 20 kDa were present in smaller quantities. The total amount of 6XHisYoaA yielded from the Ni-NTA column was 57 μg. Since the Wash Buffer for this column contained 20 mM Imidazole, the optimal concentration of Imidazole for eluting 6XHISYoaA exclusively, or predominantly, is expected to be in the relatively narrow range from 20mM to 100mM, and the total yield expected is around 57 μg.

To attempt to further isolate 6XHisYoaA from other proteins, we decided to test other columns following the Ni-NTA agarose column. Since helicases dissociate the two strands of DNA making up the double helix, they remain bound to the ssDNA that results as they translocate along their double stranded DNA substrate. A ssDNA-cellulose column is an affinity-based column consisting of cellulose linked to ssDNA oligomers. In principle, this column could bind to 6XHisYoaA by providing ssDNA as a ligand. Affinity chromatography utilizing ssDNA-cellulose was employed to purify E. coli single-strand binding protein (SSB), the ssb protein product involved in DNA recombination and repair ([28]). In this preparation, SSB was eluted with 2M NaCl. For these reasons, we decided to test the ssDNA-cellulose following the Ni-NTA

<table>
<thead>
<tr>
<th>Identity</th>
<th>Total Protein</th>
<th>Total Amount of 6XHisYoaA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>170mg</td>
<td></td>
</tr>
<tr>
<td>100 mM Imidazole Elution</td>
<td>419μg</td>
<td>57.01μg</td>
</tr>
</tbody>
</table>

Table 3: Amounts of 6XHisYoaA Yielded from Ni-NTA Column Chromatography. Crude lysate was extracted from 5 L of STL 21699 overexpressing 6XHisYoaA, applied to a Ni-NTA agarose column, and 6XHisYoaA was eluted with 100mM Imidazole. “CL” denotes crude lysate.
purification of 6XHisYoaA and chose 2M NaCl for elution. I combined Fractions 100 B, 100C, and 100 D from the Ni column, dialyzed the pooled sample in ssDNA-cellulose Wash Buffer, and applied it to the ssDNA-cellulose column (Figure 13).

6XHisYoaA bound to the ssDNA-cellulose column, as it was neither detected in the flow through nor the wash flow through. 6XHisYoaA did not appear in any of the elution fractions for this column, indicating that it failed to elute.

The second way we decided to attempt to further isolate 6XHisYoaA was based on its charge. Based on its amino acid sequence, 6XHisYoaA has an isoelectric point of 6.62, suggesting that it is predominantly negatively charged at physiological pH. Thus, a positively charged anion-exchanger column would be expected to bind 6XHisYoaA, separating it from other, cationic proteins. The MonoQ resin consists of agarose beads linked to a constitutively positively charged group, which makes it a strong anion exchanger. The MonoQ column was
used as the next step after Ni-NTA affinity chromatography in the purification of DinG ([23]). DinG eluted from the MonoQ column at 1M NaCl. Thus, we decided to test the MonoQ column and chose 1.5M NaCl for the elution step to ensure elution of 6XHisYoaA. I combined Fractions 100 B, 100C, and 100 D from the Ni column, dialyzed the pooled sample against MonoQ Wash Buffer, and applied it to the MonoQ Column (Figure 14).

![Figure 14: Results of MonoQ column purification following Ni-NTA agarose affinity chromatography.](image)

Shown are fractions of Fractions 100 B, 100 C and 100 D from the Ni column purified over the MonoQ column. (A) Coomassie-stained gel and (B) Penta-His-probed Western blots of YoaA: lane M, Precision-Plus Protein Standard molecular weight ladder; lane CL, crude lysate; lane AD, pooled sample after dialysis; lane FT, flow through; lane W FT, wash flow through; lanes E1-E5, 1.5M NaCl elution, fractions 1-5, respectively.

The yield of 6XHisYoaA from the MonoQ purification is shown in Table 4.

<table>
<thead>
<tr>
<th>Identity</th>
<th>Total Protein</th>
<th>Total Amount of 6XHisYoaA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni AD (100 B, 100 C, 100 D pooled)</td>
<td>444 μg</td>
<td>57.7 μg</td>
</tr>
<tr>
<td>MonoQ 1.5M NaCl Elution</td>
<td>77.5 μg</td>
<td>10.63 μg</td>
</tr>
</tbody>
</table>

Table 4: Amounts of 6XHisYoaA Yielded from MonoQ Column Chromatography. Fractions 100 B, C, and D from Ni-NTA affinity chromatography of 6XHisYoaA (Figure 12) were pooled, dialyzed against MonoQ Wash Buffer, then applied to the MonoQ column. 6XHisYoaA was eluted with 1.5 M NaCl. “Ni AD” denotes pooled sample after dialysis against MonoQ Wash Buffer.
6XHisYoaA bound to the MonoQ column. Proteins in the 37-25 kDa molecular weight range flowed through. 6XHisYoaA eluted from the MonoQ column and was detected in elution Fractions 1 and 2. Of the 57 μg of 6XHisYoaA applied to the column, only around 11 μg of the protein eluted, leaving behind roughly 80% of the protein bound to the column. At a salt concentration between 0.1M NaCl, the concentration of the Wash Buffer, and 1.5M NaCl, 6XHisYoaA may elute to singularity or to further enrichment.

In summary, the batch purification of 6XHisYoaA presented makes use of Ni-NTA agarose, followed by MonoQ chromatography. The elution concentration for the MonoQ column was chosen to elute as much 6XHisYoaA as possible to assess the maximum amount of the protein that could be yielded. In this workflow, the final yield of 6XHisYoaA from 5 L of induced STL 21699 following Ni-NTA/MonoQ column purification is 11 μg. Since 6XHisYoaA was eluted in batches from each column, an intermediate elution concentration for each column could be optimal for eluting the protein by itself or for further enriching in it. The elution conditions in this batch purification system will thus be used to purify YoaA through a linear gradient to hopefully separate this protein to homogeneity from the rest for downstream applications in helicase assays.

Discussion

YoaA is the protein product of the yoaA gene which is implicated in DNA repair. Since YoaA shares several conserved sequences with its paralog DinG (Figure 4), it is thought to have phosphate binding and ATP hydrolysis motifs, and four putative Fe-binding cysteine residues that DinG possesses. Genetically, YoaA was also shown to be involved in the processing of ssDNA gaps ([20]) which suggests that it achieves its DNA repair function through its putative
helicase activity. For this reason, the goal of this study was to purify YoaA for helicase assays to assess whether this DNA repair protein is an actual helicase.

The work presented here established conditions for expressing and purifying 6XHisYoaA. When 6XHisYoaA was previously expressed from the high copy number plasmid, pCA24N-6XHisYoaA, a low yield of 6XHisYoaA was obtained following Ni-NTA column chromatography and the cells experienced a certain level of toxicity. Precipitation of YoaA was observed in the cellular extracts following overexpression. This may due to inclusion bodies which can occur when high levels of recombinant protein are expressed ([29]) to protect against their toxicity ([30]). This is further corroborated by DinG displaying similar phenotypes when overexpressed ([20]). This would suggest that overexpressing helicases in general can be toxic to the cell. The resulting accumulation of ssDNA could render the genome more unstable becoming prone to cleavage from endonucleases resulting in potentially lethal double strand breaks. Another cellular response to this toxicity is to degrade the protein itself or to form truncations. In addition to possibly representing protective inclusion bodies, the precipitation may have resulted because the concentration of 6XHisYoaA exceeded its solubility threshold. The small fraction of the total protein that did solubilize was the subpopulation that could elute from the Ni-NTA column. This may account for the relatively low elution yield of 6XHisYoaA.

Since high levels of YoaA resulted in a low yield of soluble protein, we decided to lower the expression levels of 6XHisYoaA to possibly increase yield and alleviate toxicity. The low copy expression system described here produced more stable 6XHisYoaA with no detectable precipitation in the cellular extract. Previous studies in our laboratory have shown that the low copy expression does not express YoaA in levels that cause it to cross the solubility threshold and therefore there is very little YoaA to be found in the pellet of lysed cells. This also indicates
that there is more soluble YoaA in the supernatant. The absence of precipitation in the crude lysate indicates that the majority of expressed 6XHisYoaA was soluble. This does not mean that YoaA never crosses this threshold of solubility, but more likely suggests that the lower copy expression takes longer to reach these levels. The data shows that the doubling time following induction increased from the normal duration of 30 minutes to 2 hours before starting to cross the solubility and toxicity threshold. The levels of the protein did appear to be less toxic than when in high levels since no precipitate was present in the cellular extract. Another perspective for the lack of precipitation is that 6XHisYoaA was less concentrated and so may have stayed soluble simply because its concentration was lower than its solubility threshold. 6XHisYoaA was soluble in all the growth conditions tested, but the levels of the protein did depend on temperature and the duration of expression.

The optimal condition to obtain the greatest amount of 6XHisYoaA was inducing expression for 2 hours at 37°C. Induction at 24°C yielded less 6XHisYoaA possibly due to a large deviation in the physiological temperature, which may be the best temperature for expression. The fact that the greatest expression was observed after 2 hours indicates that 6XHisYoaA may be stabilized against degradation for short, but not for long, induction times. Therefore, there may exist a time-dependent competition between the production of 6XHisYoaA and its subsequent intracellular degradation. The intracellular concentration of 6XHisYoaA soon after induction was likely low enough to not appreciably exceed native levels and lead to cell toxicity. As the protein accumulated over time, it may have become concentrated enough to aberrantly associate with ssDNA gaps, generating toxic amounts of ssDNA. At this point, its toxic levels likely initiated the activity of proteases such that their degradation activity occurred at a faster rate than the translation of 6XHisYoaA. Overnight may have triggered complete
cellular degradation of YoaA due to its constant accumulation. It is also reasonable to suppose that an overnight expression is taxing because the carbon food source has been depleted and cells are forced into stationary phase. This may explain why inducing for 2 hours was optimal: it was long enough to accumulate 6XHisYoaA but short enough to not exceed the threshold of being overly toxic and thereby minimize its degradation. After obtaining stable and soluble 6XHisYoaA, a Ni-NTA column was utilized to separate the protein from others based on affinity.

The first step that was chosen to purify 6XHisYoaA was Ni-NTA column chromatography. Since adjacently spaced histidine residues can form tight interactions with Ni$^{2+}$ through a coordination complex, a RGSHis$_6$ tag was fused to the protein to separate 6XHisYoaA from other proteins based on affinity. As expected, the Ni column had selectivity for 6XHisYoaA since lower amounts of other proteins were present. The column did not appear to bind appreciably tightly to the protein since it eluted completely with 100 mM Imidazole. Although the tag is at the N-terminal of the protein, the protein’s above average size of 70.4 kDa may have affected its binding to the Ni column. The large volume of 6XHisYoaA may have partially interfered with the ability of adjacent 6XHisYoaA molecules to bind the Ni column. The first several residues of YoaA itself are mostly hydrophobic or neutral so there may have been only a very weak interaction between the native protein and the column. This may also explain why other proteins eluted with 6XHisYoaA, as they had comparable affinities that were low enough that 100 mM Imidazole outcompeted all of them for binding. Since we could not detect HolC in our construct, we were unable to confirm that it was copurifying with 6XHisYoaA. Other constructs are being constructed in our laboratory to address this. Like the Ni-NTA agarose column, the subsequent ssDNA-column tested was also based on affinity.
The second step to purify 6XHisYoaA included testing the ssDNA-column to possibly separate the protein from the rest based on affinity. Since YoaA is thought to function in DNA repair by associating with ssDNA gaps (Figure 5, [20]), it would be expected to bind ssDNA. Thus, the ssDNA-cellulose column was chosen to be a source of ssDNA ligands for 6XHisYoaA. Affinity chromatography utilizing ssDNA-cellulose was also previously employed to purify *E. coli* ssDNA-binding protein SSB ([28]). Consistent with our prediction, 6XHisYoaA did bind the ssDNA-cellulose column. However, it failed to elute despite the relatively high concentration of 2M NaCl utilized. This could be because 6XHisYoaA bound exceedingly tightly to the column. Another possibility is that 6XHisYoaA may have oligomerized with itself after it associated with the ssDNA. This occurs before the unwinding activity of many helicases such as *E. coli* DnaB which assembles as a homo-hexamer before associating with ssDNA (Figure 3, [14]). The size of the oligomerized complex may have exceeded the pore size which was likely very small because all the other proteins present in the Ni column elution fractions also bound the ssDNA-cellulose column. Thus, the elution of 6XHisYoaA from the column may have been sterically hindered.

The alternative column tested was the MonoQ column which was selected based on the ionic properties of 6XHisYoaA. The MonoQ column was also tested as the second purification step of 6XHisYoaA. Based on the primary structure of the protein it has an isoelectric point of 6.62 so it is predominantly anionic at physiological pH. Thus, it would be expected to bind to a positively charged species. The MonoQ column has terminal tertiary amine groups with no titratable protons, so it is always positively charged regardless of the pH of the solution. The constitutive positive charge on the MonoQ column renders it a strong anion exchanger. For this reason, we tested the MonoQ column at pH 8 and expected a theoretical 95% negatively charged state. In
principle, this favorable ionic interaction could separate 6XHisYoaA from any other positively charged proteins. Indeed, 6XHisYoaA bound the MonoQ column. Also, some lower molecular weight proteins flowed through which likely represents positively charged proteins. However, only 20% of 6XHisYoaA loaded onto the column eluted. 6XHisYoaA may have bound tightly to the column, as virtually all the protein was negatively charged at pH 8. Also, 6XHisYoaA may have aggregated with itself on the MonoQ column which could be related to the 0.1M NaCl salt concentration of the Wash Buffer. Two competing attractive forces are those between 6XHisYoaA and itself, and its electrostatic attraction to the MonoQ column. The strength of each interaction depends on the dielectric constant of the medium. At 0.1M NaCl, the dielectric strength may have been low enough that 6XHisYoaA more favorably aggregated to itself than bound to the column. This possibly resulted in a complex too large to flow through the pore. A marginal increase in the dielectric strength of the Wash Buffer could prevent aggregation. A salt concentration of 0.15M NaCl may be high enough to prevent 6XHisYoaA from aggregating with itself on the column but low enough that it does not interfere with its binding to the column. Since 6XHisYoaA eluted from the MonoQ column constitutes the second step of the purification

In summary, the preliminary batch purification of 6XHisYoaA makes use of Ni-NTA agarose then MonoQ chromatography. The elution concentration for each column was chosen to yield the maximum amount of 6XHisYoaA. The final yield of 6XHisYoaA from 5 L of induced STL 21699 following Ni-NTA/MonoQ column purification is 11 μg. Since 6XHisYoaA was eluted in batches an intermediate elution concentration for each column could elute the protein to homogeneity or better separate it from other proteins. The elution conditions will thus be utilized through a linear gradient as the next step to hopefully separate 6XHisYoaA from all other proteins towards assessing whether this DNA repair protein is an actual helicase.
References