REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
03-29-2019	Final Technical Report	09/01/2015 – 12/31/2018

4. TITLE AND SUBTITLE
Final Technical Report: Plant Sentinels to Explosives

5a. CONTRACT NUMBER

5b. GRANT NUMBER
ONR-N00014-15-1-2472

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Medford, June
Baker, David

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Colorado State University
Campus Delivery 1878
Fort Collins, CO 80523-1878;
University of Washington
UW Box 357350, 1705 NE Pacific St
Seattle, WA 98195-7350

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Linda A. Chrisey, PhD
Program Officer, Naval Biosciences
Code 342, Warfighter Protection & Applications
875 N. Randolph Street
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Plants can be genetically engineered to act as sensors of external substances; here we made the components for plants to sense two organophosphate pesticides, malathion and diazinon. A computationally designed protein is first designed to bind the substance of interest. *De novo* protein scaffolds for computational protein design are an effective starting point that can improve the binding efficiency over modifying existing proteins. Currently, extensive testing and modification are required to produce binders with nanoscale affinity for a ligand. When assembling the complete genetic circuit for the plant sensor, addition of a genetic positive feedback module can provide critical amplification as well as memory of a brief exposure to the substance. Modeling of positive feedback circuits informs their rational design, and the dynamic range of each circuit module's output is engineered to align with the input for the readout module. Finally, we designed novel genetic components, synthetic 5'UTRs, that can tune the function of our sensor genetic circuit at the translational phase, providing additional regulation and control over the circuit's output.

15. SUBJECT TERMS
Synthetic biology, computational protein design, positive feedback

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			June Medford
U	U	U	UU	6	**19b. TELEPHONE NUMBER** *(include area code)* 970-491-7865

GRANT #: N00014-15-1-2472

PRINCIPAL INVESTIGATOR: June Medford and David Baker (Co-PI)

INSTITUTION: Colorado State University, University of Washington

GRANT TITLE: Plant Sentinels to Explosives

AWARD PERIOD: 09/01/2015 – 12/31/2018

OBJECTIVES:

1. Produce computationally designed binding proteins to two analytes: malathion and diazinon, both organophosphate pesticides (note: target compounds were modified from original proposal).
2. Produce ligand-dependent transcriptional sensors to malathion and diazinon.
3. Test and optimize function of transcriptional sensors in isolated plant cells and whole plants.
4. Design and test regulatory genetic circuits that permit fine-tuning of synthetic genetic circuits.

APPROACH:

COMPUTATIONAL PROTEIN DESIGN

Computational enzyme design is a multifaceted problem that tackles optimization of multiple parameters such as substrate binding, transition state stabilization and product release to allow enzymatic turnover. The David Baker lab (University of Washington) has successfully created a variety of *de novo* scaffolds that have several advantages over the native proteins; most critical is that these scaffolds explore a wider range of possible conformational space compared to native proteins and thus can potentially host an active site closer to the calculated transition states. We used such *de novo* scaffolds as starting templates for high affinity organophosphate capture reagents. In order to form a covalent and irreversible interaction with the target compounds (diazinon and malathion), we introduced a serine-histidine-aspartate catalytic triad, which is the covalent binding site of the organophosphate compounds in the native target acetylcholinesterase, to the stable *de novo* scaffolds. Once we found the appropriate scaffolds to host the catalytic triads, we will use Rosetta to optimize the sequence in order to pre-organize the active site towards a conformation favorable for the reaction with the organophosphate compounds to occur. Designs with micromolar affinity towards the target compounds diazinon and malathion will be further sequence-optimized using a combination of directed evolution and computational and structure guided design process. We will then iterated this process in order to obtain a binder with nanomolar affinity for the target compounds.

TUNABLE CONTROL OF GENETIC CIRCUITS

Positive Feedback Circuits

A positive feedback circuit can provide plants with the ability to remember a brief exposure to a substance of interest and amplify the response. The ability to precisely tune module outputs (e.g., readouts) to accommodate application-specific components will enhance modularity and efficiency of genetic circuit design. Connecting genetic circuit modules (sensing, controller, readout, etc.) requires that the dynamic range of each circuit module is compatible. For example, if the readout needs further amplification or memory, a component with a known input-output function can be introduced modularly to confer the desired properties and produce a predictable output. We designed and built two positive feedback circuits and used modeling and experimental results to improve their function. In this way, and developed a basis for tuning and adapting genetic modules to a variety of applications. Downstream tuning and modifications can then be based on knowledge of parameters and application needs.

5'UTRs

In collaboration with the Howard Salis lab at Pennsylvania State University, we have investigated an additional regulatory system that can provide circuits with tunable control. This system is based on rational design of 5' untranscribed regions (UTRs). 5'UTRs, located upstream of the mRNA start codon, confer translational efficiency and stability to mRNA transcripts. Using rationally designed synthetic 5'UTRs, gene expression can be modulated post-transcriptionally, tuning expression levels to a custom application with predictable output.

The Salis Lab designed 66 synthetic 5'UTR sequences for Arabidopsis, using general principles for robust translation initiation in eukaryotes. The UTRs were designed to be maximally informative with respect to thermodynamic properties known to influence ribosome/messenger cap binding, scanning, pausing, and translation initiation in plant mRNAs. The study of these translational features will enable a reliable *de novo* design of synthetic 5'UTRs. The Translation Initiation Rate (TIR) will be determined in Arabidopsis protoplasts, and the data will allow us to fully characterize 5'UTR features and their contribution to the TIR in a plant genetic circuit. While we have used transcriptional features to produce positive feedback and toggle switches, the 5' UTR and translational regulation will provide fine-tuning of our circuits.

ACCOMPLISHMENTS:

COMPUTATIONAL PROTEIN DESIGN

We have previously described our efforts in the computational design of a covalent-capture protein for the organophosphate pesticides diazinon and malathion. We obtained the geometric parameters for the interaction between the covalently bound ligand and the active site residues from Protein Databank (PDB) crystal structures of proteases or esterases with organophosphate ligands or ligands with tetrahedral geometry that are covalently bound to the

serine of the catalytic triad. These geometric parameters were then used with the Matcher algorithm to search for appropriate *de novo* scaffolds that can host the active site residues in the catalytic conformation. We focused on using *de novo* scaffolds that are primarily alpha-helical in nature since they were more stable and conducive to mutation than the *de novo* mixed-alpha-beta proteins.

We have characterized and tested the stability of the preliminary designs for malathion and diazinon. After computationally verifying that the protein sequences of the designs will fold to the designed protein conformation using the Rosetta *ab initio* folding protocol, we ordered ten preliminary designs from the helical repeat scaffolds. The goal in characterizing these designs is to show that the design pipeline will not detrimentally affect the stability of the scaffold and thus will show the feasibility of the design process for a large-scale production effort. All ten designs for malathion binding were successfully expressed and were purified in soluble form in *E. coli*. Using circular dichroism, all ten designs were found to be thermostable and predominantly alpha-helical in character, as designed. Seven of the ten designs were found to be in the monomeric state as ascertained by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) experiments, suggesting that the proteins folded as designed. Furthermore, the designs were characterized by small-angle X-ray scattering (SAXS) experiments. Three out of the seven monomeric malathion binder designs showed significant correlation between the design model and the SAXS profile, indicating that the design model was in agreement to the actual structure of the proteins. We are currently attempting to solve the structure of the three designs using x-ray crystallography.

To test the activity of the designs, we have successfully introduced the coding sequences of the designs into a yeast expression vector and fused it to the yeast aga1 coding sequence. This allows the designs to be expressed in the yeast cell surface and tested for ligand-binding using a biotinylated diazinon or malathion probe. The diazinon or malathion moiety of the ligand serves as a reactive handle that forms a covalent bond with the active site serine of the design while the biotin moiety allows binding of a fluorescent conjugate of streptavidin and acts as a reporter for binding of the ligand to the design. We have successfully expressed the design in yeast cell surface display using fluorescent-labeled organophosphate probes synthesized by collaborators from the Chemistry Department at Colorado State University. The methodology for biotinylating diazinon has been published (Nottingham et al., 2018). However, testing for binding of the biotinylated malathion probe at a concentration of 4 μM resulted in no detectable binding signal for the surface displayed designs.

TUNABLE CONTROL OF GENETIC CIRCUITS

Positive Feedback Circuits

A. Characterizing Zinc Finger Positive Feedback Circuits

We generated concentration curves for the zinc finger circuit and determined the minimum concentration at which the positive feedback circuit increases expression of luciferase. We then developed a mathematical model to describe the activity of the direct activation system, which

lacks a positive feedback step, and the positive feedback system. This was a Process Model, which mathematically describes the molecular interactions of the genetic circuit. The process model can be used to estimate the output of luciferase molecules as a function of time. We then performed experiments, using transiently transformed protoplasts and a luciferase readout, to confirm the model predictions and capture the full dynamic range of the circuit's response to the inducer. The experiments demonstrated that the addition of a positive feedback step confers significant amplification to the circuit after 24 hours, that the response increases with increasing concentrations of the OHT inducer, and that the dynamics displayed the expected sigmoidal shape with a low output state and a high output state. These results were then used to develop a Data Model. The data model differs from the process model in that it captures sources of variability that affect the genetic circuit and measurements of its function, including plasmid copy number, random effects, and in stably transformed plants (to be accomplished), different chromosome insertion points. Final experiments have been completed and analysis is being done to prepare the work for publication. We may need to perform additional work prior to publication.

B. Gal4-based Positive Feedback Circuit

The Gal4-based positive feedback circuit was introduced into plants and stable plants produced. This circuit was found to have a certain amount of circuit leakiness. We proposed several fixes to the circuit design. The first was a titration sponge, which would increase Gal4 binding sites in order to sequester Gal4-containing transcription factors and reduce the leakiness of the Gal4 binding site-containing promoter in the positive feedback step. Preliminary tests of the titration sponge constructs suggested that small numbers of Gal4 binding sites have minimal or no effect on reducing leaky expression. However, this work needs reproduction and critical analysis. Larger numbers of Gal4 sites, however, produced a reduction in leaky expression without significantly reducing the range of the high state. This resulted in a substantial overall increase in fold change as a result of the positive feedback circuit.

Design and analysis of synthetic regulatory 5' UTR elements for the manipulation of translation efficiency

The 66 synthetic 5'UTRs were rationally designed to test seven criteria: 1) Length of the 5'UTR; 2) Presence of a Stem Loop 1 (SL1) motif, a hairpin that confers stability or enhancement when positioned correctly; 3) RNA stability relative to CAP-proximity (RNA secondary structure can be inhibitory to CAP binding, affecting the translation initiation rate); 4) Ribosomal scanning influence (conserved hairpin lengths with different DNA sequences can affect the scanning context efficiency for the 40S ribosome subunit); 5) Evaluation of in-frame and out-of-frame start codon translation efficiency (the use of alternative start codons can reduce the translation rate of the true start codon); 6) Alternative Kozak sequences (mutations at each position of the canonical eukaryote sequence "AACAATGGC"; 7) Non-canonical start codons such as "CUG" and "ACG" as compared to the canonical "AUG" start codon.

We constructed the UTRs using a PCR extension technique. Nine of the 66 were problematic due to structural complexity or other reasons and could not be assembled, leaving 57. The 57 5'UTRs were divided into six experimental groups based on similar structure and function.

Plasmids were designed and assembled with the 5'UTRs placed between a pNOS promoter and a Firefly luciferase readout gene. Each plasmid also included a Renilla luciferase gene under control of the FMV promoter; this acts as a measure of plasmid transformation efficiency. Plasmids were transiently introduced into Arabidopsis protoplasts, and Firefly (Fluc) and Renilla (Rluc) were quantified using a luciferase camera.

We have collected luciferase data for each construct, and performed a preliminary analysis. We found that the dynamic ranges of the two luciferase reporters differed, due to the greater strength of the FMV promoter relative to the pNOS promoter. The luciferase data must be re-extracted from the camera software and adjusted by a constant factor in order to bring the dynamic ranges into alignment. The postdoc responsible for this work has left the Medford lab, so the work to complete this step has been delayed. Once the reanalysis is complete, we anticipate a manuscript on our results in collaboration with the Salis Lab.

CONCLUSIONS:

- *De novo* protein scaffolds for computational protein design are an effective starting point that can improve the binding efficiency over modifying existing proteins. However, extensive testing and modification are still required to produce binders with nanoscale affinity for a ligand.
- Modeling of positive feedback circuits informs rational design, and the dynamic range of each circuit module's output must be in alignment with the input for the next module downstream.
- Addition of a titration sponge provides some reduction in leaky expression in a Gal4-based positive feedback circuit. However, the contribution of a degradation tag to circuit leakiness has not been verified.
- Rationally designed synthetic 5'UTRs may provide a path for additional genetic circuit regulation and tuning at the translation stage.

SIGNIFICANCE:

COMPUTATIONAL PROTEIN DESIGN

Protein design methodologies previously used for other ligands are adaptable to designing binders for organophosphate pesticides (malathion and diazinon). We are exploring new scaffolds for catalytic triad designs in order to find more appropriate *de novo* starting designs for malathion and diazinon binding. This research has revealed the importance of exploring the interfaces of oligomeric scaffolds as active sites for catalytic triads. These interfaces might provide a more diverse shape of potential binding sites than the pockets of their monomeric counterparts and can potentially provide more stability due to a bigger hydrophobic core for the oligomers.

TUNABLE CONTROL OF GENETIC CIRCUITS

Positive Feedback

A goal of developing and optimizing various positive feedback circuits, and quantitatively characterizing their behavior, is to develop a "drop-in" module that can be added to any genetic circuit to produce a predictable change in expression. This quantification must include measurement of the dynamic range of the steps before (input) and after (output) the positive feedback step. We have treated the zinc finger (ZF) positive feedback system as a test system that permits development of a model of positive feedback circuits in general. In subsequent work, we will apply the ZF results and models to a well-characterized input/output system with a known KD and characteristic *in planta* transcriptional response, such as our digoxigenin-inducible transcriptional activator system. Application of the model to experimental data will provide a path forward as we improve the positive feedback circuit. The eventual goal is a genetic circuit with highly predictable function that results from connecting characterized modules in plants.

5'UTRs

We expect the results to help determine the rules governing ribosomal interactions on plant mRNAs. Once these have been established, a model of *Arabidopsis* translation initiation can be constructed. The model will be coupled with an automated design algorithm that generates UTRs for tunable control of gene expression with any DNA coding sequence. The results of this project will streamline the construction of practical genetic circuits from previously characterized components.

PATENT INFORMATION:

US Patent application 62/701,396 was partially supported under related ONR award N00014-07-1-0180.

AWARD INFORMATION:

N/A

PUBLICATIONS and ABSTRACTS (for total period of grant):

Nottingham, KG, McNally, A, and McNaughton, BR (2018) Synthesis of biotinylated diazinon: Lessons learned for biotinylation of thiophosphate esters. *Tetrahedron Letters* 59: 234-237. https://doi.org/10.1016/j.tetlet.2017.12.001