Title
Stress-responsive biosensors for agricultural applications

Author
Hines, George

Publication Date
2013

Peer reviewed|Thesis/dissertation

Stress-Responsive Biosensors for Agricultural Applications

by

George Herbert Hines

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering – Mechanical Engineering

and the Designated Emphasis

in

Computational Science & Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Andrew K. Packard, Co-chair
Kameshwar Poolla, Co-chair
Lewis J. Feldman

Fall 2013

Stress-Responsive Biosensors for Agricultural Applications

Abstract

Stress-Responsive Biosensors for Agricultural Applications

by

George Herbert Hines

Doctor of Philosophy in Engineering – Mechanical Engineering and the Designated Emphasis in Computational Science & Engineering

University of California, Berkeley

Andrew K. Packard, Co-chair

Kameshwar Poolla, Co-chair

Plants have internal stress response mechanisms. These mechanisms are activated in response to environmental stimuli. Measuring the activation of these mechanisms directly gives a clearer picture of how an environmental stimulus is actually affecting the stress state of the plant. This information is valuable in an agricultural context because these internal mechanisms are generally upstream of observable physiological manifestations of stress. But measuring the activity of these pathways is often expensive and time-consuming. We will demonstrate an optical biosensor system reporting the activity of a heat- and drought-sensitive gene pathway in *Arabidopsis thaliana*. The marginal cost per data point of this system is very low, and the rate of data acquisition is faster than the dynamics of the gene pathway, so the biosensor system can easily and quickly detect transient changes in pathway activation, opening a real-time window into the heat- and drought-stress state of the plant. With this real-time information in hand, we proceed to close the loop around the biosensor output using a computer-based feedback policy to maintain a constant biosensor expression level using the temperature in a small greenhouse as our control input. Along the way we will discover that as complicated and unknown as the internal dynamics of this process may be, a mechanistic model is not necessary to design a feedback policy that is robust to biological variability in the plants.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Plants can neither walk nor talk. These obervations may seem banal, but they have profound evolutionary consequences. Immobility has forced the emergence of internal mechanisms for surviving climatic extremes; lack of precise communication between humans and plants has made the domestication of crops a heuristic exercise. These inabilities have not hamstrung commercial agriculture so far, but as food production expands to marginal land, adaptation is required of crops and growers. The adaptation of crops occurs in the laboratory and the nursery—technicians genetically modify and crossbreed plants in order to boost yield and quality, and to suppress susceptibility to unfamiliar environmental stresses. Crop adaptation is inherently rate-limited by the length of the growth cycle, on the scale of months. The grower adapts by adjusting the allocation of nutrients and pesticides to allow the crop to thrive in non-native conditions. Such interventions have time scales of hours or even days. But determining the appropriate intervention is complicated by the lack of real-time information about the health of the plant. Remember, plants don't talk.

The physiological effects of an environmental stress are not coincident with the onset of that stress, but rather are downstream (in time) of stress-induced genetic and chemical responses. The time from stress onset to observable manifestation is fundamentally limited by the temporal length of the mediating biochemical pathways. Because of this intrinsic delay and the biological variability from plant to plant, quantifying the health of a crop based on physiological properties of individual plants is challenging. Not surprisingly plant stress measurement is a vibrant research area. A theme is the "biosensor." The term itself applies broadly to any mechanism that enables measurement of a biochemical process before that process has effected a visible change in the host organism. To narrow our field of view we will consider biosensors in the context of feedback control of the growing environment. What will unfold is our perspective on integrating plant stress biosensors into a climate control system. Our intent is to encourage others to discuss and study a topic that we believe is important to the future of commercial agriculture.

Growing environments come in different shapes and sizes. It is tempting to separate them into *field* and *greenhouse* environment. In this classification field environments expose the crop directly to the outside world, while greenhouses offer at least some protective covering to separate the crop from the elements. But on the "greenhouse" side of that divide there is a broad spectrum ranging from "row cover," which is often simply a long sheet of translucent plastic placed over low plants to protect them from frost, to fully actuated "hothouses" for growing decorative plants, with the climate controlled down to the gaseous content of the air inside. Rather than trying to classify growing environments, we will consider them in terms of the *available actuation*, the environment parameters over which the grower has control. In a field environment the grower often has control over the allocation of water and pesticides. In a hothouse the grower may also control the temperature, humidity, or even the carbon dioxide content of the air. The nature of the available actuation will determine how, or even if, the information provided by a stress-responsive biosensor can be used.

Regardless of the forms they take, growing environments all share the purpose of turning profit. The grower accomplishes this by attempting to lower operating costs and boost yield. This is a tradeoff that penalizes stinginess and delay. If resources are applied too sparingly or too late in the quest for lower costs the crop will die, a consequence that encourages preemptive overcorrection. The latter strategy indeed works well, albeit inefficiently, for staple crops. But there are other classes of crops whose marketable qualities are harmed by exposure to excess nutrients. These are predominantly "flavor" crops, of which grapes and peppers are good examples. The quality of these crops improves when they are carefully stressed at key phases of the growth cycle, forcing the grower toward riskier intervention strategies. The risks are mitigated by the experience of the grower and observations of the physiology of individual plants, observations that are delayed with respect to the onset of stress. And so we arrive at an application of biosensors: speeding up the accurate discernment of plant stress in the field to enable better resource allocation. In other words, biosensor-based feedback control of the growing environment.

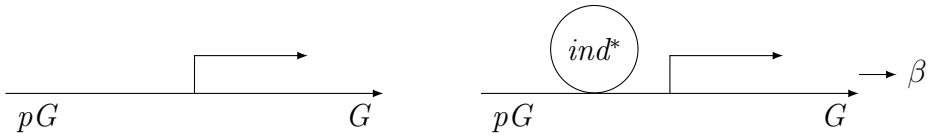
1.2 Cell Biology Background

The Central Dogma

The relationship between genes and proteins in a cell is described by the "central dogma" of biology, which holds that the DNA composing a gene is *transcribed* into chains of RNA, which are then *translated* into amino acid chains, which then fold into (nominally) functional proteins.

A gene is a sequence of DNA that is coherent in the sense that it is the code for a specific protein. A gene's transcription is regulated by a protein called a *transcription factor* (TF), or possibly several TFs. These regulators bind to a DNA sequence upstream of the gene and either *promote* or *repress* transcription. The DNA sequence to which the TF binds is referred to as a *promoter* or *repressor*, respectively. We will be dealing exclusively with

promoters in this work, and will therefore occasionally refer to the associated transcription factor as an *inducer*. The interaction between genes, promoters, and inducers is illustrated in a cartoon in Figure 1.1.



(a) Inactive (b) Active

Figure 1.1: In general a gene's promoter (pG) is physically upstream of the gene G relative to the direction of transcription, which is denoted by the direction of the arrows. The bent arrow emphasizes the genetic distinction between the promoter and the gene to be transcribed. When the inducer ind^* binds to pG, transcription of G (production of mRNA that will be translated into G's protein) proceeds with rate β. The translation occurs downstream in time and not shown.

When a promoter is bound by an inducer the DNA helix splits open, granting a large molecule called RNA polymerase (RNAP) access to the DNA sequence. The RNAP facilitates the formation of a chain of *messenger RNA* (mRNA) whose sequence is complementary to that of the gene. This process (transcription) takes place entirely in the cell nucleus, after which the mRNA chain is transported to the cytoplasm. In the cytoplasm there is a matrix of individual *transfer RNA* (tRNA) molecules, each of which exists bound to a corresponding amino acid (the fundamental component of a protein). When the mRNA chain arrives it is bound by another large molecule called a *ribosome*. The ribosome is a moving scaffold that couples each successive base in the mRNA chain to its complementary tRNA and then binds the amino acids associated with neighboring tRNAs. This is the translation process, and the result is an amino acid sequence that then folds into a functional protein.

The protein produced by one gene may act as a transcription factor for a different gene. In this case the genes are said to be part of a *cascade*. When many genes are mutually interacting in this way, the collection is referred to as a *network*.

Transformation

Physically, a gene G_1 and its promoter pG_1 are roughly consecutive sequences of nucleic acids in the DNA. Modern synthesis techniques make it possible to assemble sequences of our own construction, so the central dogma can be hacked by constructing a nucleic acid sequence composed of the sequence for pG_1 followed by some other gene sequence G_2. This new sequence, call it pG_1/G_2 can be inserted into the DNA of a host organism in which pG_1/G_1 is already functional. This insertion process is called *transformation*. An organism which has never been thus transformed is referred to as a *wild type*.

Once a host organism has been transformed as described above, there are two occurrences of pG_1 in the hosts's DNA: one promotes G_1 and the other promotes G_2. So while in the

wild type host binding of pG_1 would only promote transcription of G_1, in the transformed host binding of pG_1 promotes the transcription of both G_1 and G_2.

Fluorescent Proteins

(This section is a colloquial summary of [34], the definitive review of the history and development of fluorescent proteins for use in the lab.) Fluorescent proteins (FPs) are light-emitting proteins with two characteristic wavelengths: an *excitation* wavelength at which it absorbs light and a longer (lower-energy) *emission* wavelength at which it fluoresces. Gene sequences for a variety of different FPs are readily available and can be inserted into the DNA of host organisms with relative ease. A canonical use for a genetically encoded FP is as a signal of the activation of a known promoter. This involves the construction of a single DNA sequence composed of the promoter sequence followed by the FP gene sequence. After transformation with this construct, as described above, subsequent observation of FP fluorescence indicates the activation of the associated promoter.

The activity of fluorescent proteins was first observed in *Aequorea*, a particular species of jellyfish. The wild type is green FP (GFP), but since that discovery an ongoing effort of isolation, sequencing, and characterization has yielded many variants with emission wavelengths that span most of the visible spectrum. While FPs have been identified natively in many organisms besides *Aequorea*, that wild type is unique because all the information necessary to produce an active fluorophore is contained in its gene. This is in contrast to other wild-type FPs that require "third-party" helper compounds to activate the fluorophore after the FP has been produced. It is this self-containment, or *autonomy* that allows *Aequorea* GFP to be active in a wide variety of organism, from bacteria up to mammals, multiplying its utility as a research tool.

1.3 Plant stress

The stresses to which a plant is exposed very with geography and time. Generally temporal variation (seasonal or diurnal) acts periodically around a baseline stress profile that is characteristic of the local climate. Climatic variation is responsible for regional agricultural specialization. Diurnal and seasonal variation can profoundly influence crop development in a variety of ways, and plants have been shown to respond to a broad palette of environmental stressors including, but very much not limited to, cold, heat, drought, salinity, and light intensity. Certain characteristic patters of physiological response are known. For example, cold stress induces flowering (the initiation of reproduction) in a process called *vernalization* [2, 12].

It is important to emphasize that the physiological changes induced by environmental stresses are not necessarily bad. In the case of grapes grown for wine (*Vitis vinifera*), centuries of experience and decades of supporting scientific evidence have codified the need for managed stress at certain points in the growth cycle in order to amplify desirable charac-

teristics of the fruit [9]. Specifically, carefully withholding water has the doubly positive effect of simultaneously discouraging excessive foliage (so more sunlight can reach the lower branches and fruit clusters) and encouraging the concentration of sugars and other flavor compounds in the berries.

Water stress is notable because in most arable regions irrigation is available in fields in normal circumstances. When a region is not experiencing a general drought, water stress can be reduced in exchange for money. This sets up a tradeoff that we can hope to inform with better quantification of plant stress profiles and their relationship to economic yield. In harsher climates where field agriculture is not viable, greenhouses or other types of protected growing environments introduce more actuation authority. When there are not hard limits on the availability of resources (electricity, CO_2, water, etc.) then tradeoffs are possible in any dimension of stress that is associated with an environment control input. Finding the sweet spot of these tradeoffs of course relies on an accurate model of the marginal cost and benefit of a given climate control input. Such models are still a long way off, but building an accurate real-time picture of the crop stress state is an important step.

Because many different stressors can produce the same response, physiological manifestations cannot necessarily be used to infer the stresses that caused them. This many-to-one phenomenon is called stress *integration*. Integration occurs at cellular scales as well: while one gene may be induced by five different environment stresses, there may be five different genes that are each induced *selectively* by one of those stresses. A meta-analysis of plant stress research (conducted specifically with an eye toward hydrogen peroxide sensitivity) illustrates this point nicely, showing that the transcript level of the protein HsfA2 is strongly induced by at least five different stresses (including ozone, hydrogen peroxide, and salt), while HsfC1 is only strongly induced by one (cold) [18]. (A note on terminology: for historical reasons, most proteins that are active under stressful conditions are referred to as "heat shock" proteins/factors, even though they may be induced by other stresses. Consequently, many stress-responsive proteins will be abbreviated with the prefix *Hsf*.)

Some stresses often occur in combinations, a notable example being the pairing of drought stress with heat stress. This statement must be accompanied by the warning, explored in [20, 26], that coupling of stresses leads in some cases to unique responses that are not observed when the stresses are applied individually. Indeed, there are some transcripts that are produced under combined drought and heat stress even though they are not produced when the plant is under those stresses individually. A select few, on the other hand, are produced under heat stress, drought stress, and their combination. As a preview of things to come, it is notable that one of these transcripts codes for the protein DREB2A.

1.4 *Arabidopsis thaliana*

Becuase living organisms are complex, many disciplines have reached domain-specific consensus that certain representative *model* organisms will serve as proxies for larger classes of organisms. A well-known example is the use of Sprague-Dawley rats in early stage biomed-

ical research. For plant biology, especially the study of small flowering plants, the model organism is *Arabidopsis thaliana*, a member of the mustard family native to the hilly regions of eastern Europe and Asia. Referred to briefly as just *Arabidopsis*, the plant is small and has a relatively rapid life cycle. These properties render *Arabidopsis* suitable for efficient experimentation. Consequently recent years have found its entire genome sequenced numerous times, and vast databases compiled of functional genes, promoters, and binding (activation) stimuli. The most common method of genetically transforming *Arabidopsis* is called the "floral dip" method [4], and its simplicity further recommends *Arabidopsis* as an efficient research tool.

1.5 Biosensors

A biosensor, broadly defined, is a mechanism that allows the quantification of a biochemical process. Biosensors are differentiated by their mechanisms and their quantification methods. We will speak of biosensors as systems in terms of their *transduction*, *signaling*, and *detection* units. (This is a slight modification of the division used in [22], where they refer to a biosensor as the composition of a "biological component" and a "physicochemical transducer.") Transduction units can be chemical, e.g. induced by binding specific molecules, or genetic, e.g. encoded in DNA with production induced by response to a stimulus. Stably transformed genetically encoded biosensors are persistent across generations, so in the case of plants a successful biosensor gene insertion can establish a stable line of biosensor hosts available for the effort of growing them. Signaling units come in a variety of forms but we will only consider *optical* biosensors, which emit detectable levels of light when excited by an external energy source. Specifically, we will consider biosensors that include a fluorescent protein signaling unit. Detection units for optical biosensors measure the light intensity emitted thereby. These also come in a variety of form factors, most of which include an excitation light source that provides the energy input to the fluorescent protein, along with a photodetector that gathers the emitted light. Often there is a computer image processing step. Importantly, the marginal cost of a data point with many fluorescence meters is high, as a data point either takes a very long time to collect or requires extensive human intervention.

Practicality considerations

The idea *per se* of practical biosensors is not new. A recent review of optical detection schemes for plant stress discusses at length the characteristics possessed by a hypothetical practical biosensor [13]. It will be useful to refer to the list that accompanies this discussion:

1. There should be no or at most weak basal (uninduced, background or leaky) reporter gene expression.
2. The fusion gene should be highly inducible.

3. The range of compounds or conditions that elicit gene expression may be broad or narrow; in other words, specificity is according to user defined objectives.
4. The intensity of the signal should be well correlated with the concentration of the inducer compound or the severity of the physical/biological condition.
5. The signal response should be easily measurable with a high dynamic range of inducer concentration or conditions.
6. The spatial distribution of the signal should be uniform throughout the plant or a specific tissue depending on the specific application.
7. The temporal response should be appropriate to user objectives; for example, for sensing water-deficity conditions in plants, a measurable response should be elicited at early onset of water stress before plant entry into permanent wilting point.
8. "Switch-off" of the signal should be possible once the inducer compound, triggering event or conditions is removed.
9. The optical characteristic of the signal should be distinct from any interfering background noise within the plant environment to give high signal-to-noise ratio.

I argue that points 1 and 2 together are just a different way of expressing the sentiment in point 9. Point 8 deserves special mention because, as we will see, "switch-off" of production does not necessarily lead to switch-off of the biosensor output. However at least in the case we present, the production state (on or off) can still be inferred from the dynamic response of the biosensor. Furthermore, we will show that a biosensor that does not comply with point 8 may indicate that a plant has memory of the history of stress. (A similar roundup appears in [28], but those authors exclude some of the more questionable items.)

I further stress that a practical biosensor for agricultural applications must also have a low marginal cost of data collection. In the case of a stable genetically encoded optical biosensor the marginal cost of data is almost entirely concentrated in the detection step. Making this step cheap while preserving the biophysical meaning of the resulting measurement is an important step towards practicality.

Biosensors have obvious applications in agriculture, as the biological characteristics of a plant are of profound significance to the economic viability of a farming operation. The unique pressure on a biosensor in an agriculture setting is that the information gleaned from it should correlate with a metric of economic value. (Measurements from a drought biosensor should allow determination of required cost of irrigation, for example.)

Detection: remote or not

Detection mechanisms are either remote or non-remote. Remotely detectable biosensors are indicators that can be observed without making physical contact with the biosensor

host organism. The exact meaning of "contact" is context-dependent, but generally if the absolute detection distance is small (e.g. the detector is in close physical proximity to the organism) then "remote detection" carries the connotation of being visible with the naked eye. Measurement instruments may be required if the absolute detection distance is large, for example when aerial imagery of a field is used to measure leaf canopy extent. Remote detection from long distances has the benefit of a wide field of view, allowing rapid collection of a large dataset. This is one way to significantly reduce the marginal cost of data.

Some biosensors measure the natural physiological responses to environmental factors. An example is the "pressure bomb" described in [31] and in regular use in modern vineyards. The device is used for measuring the leaf water potential, a proxy for water stress, and consists of a small pressure vessel into which a plucked leaf is inserted with its stem sticking out through a seal. The pressure at which sap flows from the cut stem is the leaf water potential. Flow induced by low pressure is an indicator of excess, or *luxuriant*, watering. Flow that can only be induced by high pressure indicates under-watering, and the target irrigation level is some midpoint that is heuristically associated with good crop outcomes.

Plants themselves have innate remotely detectable biosensors in the form of visible physiological manifestations of environmental factors. For example, under prolonged exposure to high levels of heat and drought plants will wilt. This is worth pointing out because although the delay associated with wilting is long and the damage already done is possibly considerable, the "measurement" is free. This is a common theme with remotely detectable biosensors: although the marginal cost of collecting the data is very small, the delay between the introduction of the inducing factor and the appearance of meaningful output signal is long.

One notable example of a highly engineered remotely detectable biosensor is presented in [3]. The authors modify a tobacco plant to indicate the presence in the surrounding air of molecules of the explosive trinitrotoluene (TNT). The work is especially interesting because it highlights the modularity of the transduce/signal/detect biosensor paradigm. In this case the transduction mechanism is quite complex, with a trans-membrane protein spanning the cell wall so that the (large) TNT molecules find an external binding site. (They are too large and of the wrong composition to pass through the wall into the cell.) The binding of TNT molecules to the transmembrane protein induces an intracellular transport mechanism that shuttles a signaling compound into the cell's nucleus. This transport activates one or the other of two signaling mechanisms. The first, used for verification of the transduction mechanism's activity, is the production of a fluorescent protein; this signal is not remotely detectable, but is rapidly measurable. The second, and remotely detectable, signaling mechanism is the activation of a "degreening" pathway that causes the leaves of the plant to blanch. This physiological change takes several hours to manifest itself, which is often the case with remotely detectable signals. But it does have the virtue of being resettable (*a la* item 8 above), with the leaves of the plant eventually turning green again after the removal of the TNT stimulus. There are two take-home points from this example:

1. The same transduction unit can be used in conjuction with a palette of signaling units

depending on the constraints of the application.

2. Even with intricately designed biosensors, there is generally a tradeoff between remote detection and response time.

The sense/transduce/detect framework is both intuitive and instructive, as it highlights the potential modularity of biosensors in any context, although we will consider the particular case of stress biosensors. Indeed, we think that this is a necessary way to consider *practical* biosensors. Many genes have been identified as responding, solely and in combinations, to different environmental stresses. In theory the associated promoters could be used as transduction units for biosensors of their associated stress state(s), and linked to any of a palette of signaling units.

Stress biosensors

An important consideration in designing a biosensor to quantify an organism's perception of its environment is that the measurement should reflect something significant to the organism. In the transduce/signal/detect paradigm, this suggests that a biosensor designed to quantify an organism's response to an environmental factor should have a transduction unit that is endogenous to the organism. The trouble is that the endogenous stress response pathways of many organisms (and plants in particular as previously discussed) are highly complex with many interacting transcription factors in interconnected gene networks. This makes *specificity*, or designing a biosensor whose output is a correlate of only one environment factor, a difficult design goal to attain.

Heat stress is a good and notable example of this difficulty because it is initially induced neither by the elevation or depletion of a chemical agent. Elevated heat causes "denaturation" of proteins, which generally means that they unfold into non-functional configurations. Because these proteins are generally necessary for successful operation of cellular machinery, it is important that denaturation be corrected, a task which falls to a group of proteins that are collectively referred to as "heat shock proteins" that act as "chaperones," preventing denaturation [16]. This nomenclature is somewhat unfortunate, since the members of the heat shock protein family are known to be produced in response to an array of stresses and indeed many occur normally [14].

Water stress also has both biological and mechanical characteristics. Water is absorbed into the roots of the plant and then climbs through capillaries in the stalk to the leaves, where it is either used to facilitate biological processes or simply transpired out in the air. Plants control the rate of transpiration to some extent by opening and closing small pores in the leaves, but severe lack of water or very high temperature can both result in an inability to retain sufficient water to keep everything running smoothly, at which point wilting is observed. What makes water stress analysis slightly easier is a history of detailed transport models based on potential energy balance, for example [21].

Real-time applications

Our final bit of context regards the use of biosensors in real time. Biosensors have proven useful as a means of identifying the parameters of mathematical models describing the dynamic behavior of certain well-understood biochemical processes. For a first example, [6] uses carefully designed fluorescent proteins to study the activation dynamics of c-Jun N-terminal kinases (JNK), members of a mitogen-activated protein kinase (MAPK) cascade.

The transduction mechanism is a pair of adjacent binding sites that are folded together by the selective binding of JNK. The signaling mechanism is the Förster resonance energy transfer (FRET)-induced YFP fluorescence. Specifically, a cyan FP (CFP) is attached on one side of the pair of binding sites, and a YFP is attached on the other side. CFP and YFP have distinct excitation wavelengths, but the CFP's emission wavelength coincides with the excitation wavelength of the YFP. In this use case the CFP is continually excited, and when the JNK binds and folds the sites together, the CFP and YFP are brought into close enough proximity that the CFP's emission excites the YFP, causing an increase in the YFP fluorescence relative to the CFP fluorescence. This is detected by an optical microscope taking images at a rate of 1/minute. The fundamental time scales of transcription/translation are estimated to be on the order of 10 minutes [1], so the one-minute sample time in [6] is below the Nyquist sampling limit.

The fine time resolution and sensitive signaling mechanism allow very accurate estimates of model parameters assuming a model exists. In the case of the various component reactions of the MAPK cascade, validated models do exist (see [11] for a focused treatment). For the stress-response systems that we consider in high organisms (as opposed to the mammalian cell colonies considered in [6] for example) models do not exist because some links in the causal chain from environment factor to physiological manifestation are unknown. I offer this example to acknowledge that certain real-time applications of biosensors already exist, but they are significantly dissimilar to the contributions that we will present.

Even more recently, a second example of a real-time biosensor application is demonstrated in [17]. A light-sensitive promoter with active ON-OFF control is linked to YFP and genetically encoded in a yeast colony. The promoter takes advantage of the PhyB/PIF3 interaction so that when a pulse of near-infrared (IR) light is applied production of the YFP begins, and when a pulse of far-IR light is applied production ceases. The measurements, taken with a flow cytometer every 30 minutes, are passed in real time to a model-predictive controller (MPC) running on a benchtop computer, which computes a schedule of near/far-IR light pulses to regulate the YFP fluorescence to a specified setpoint in closed loop. From this example I want to highlight the following observations:

1. A mechanistic model is not used. However, since MPC by definition requires mathematical representation of the dynamics ("model") a four-state linear model is postulated and its parameters identified on the fly.

2. The transduction mechanism is not endogenous to the yeast cells.

3. As a result, the goal is setpoint regulation *per se*, without the setpoint taking on any deeper significance to the host organism.

This is the first published demonstration of such a computer-based feedback control system regulating a biosensor level by means of an external environment factor.

1.6 Contributions in this work

My primary contribution is in creating the nexus of the characteristics that the previous examples demonstrate individually. My vision for agricultural stress biosensors involves:

- stable genetically encoded biosensors
- fast enough response time to enable real-time decision making about resource allocation;
- a nondestructive, *in situ*, real-time measurement protocol;
- low marginal cost of data collection;
- correlation to characteristics that are relevant to crop yield;

In the following I will describe the development of a biosensor system that meets all these criteria, and will further demonstrate its use by implementing a computer-based feedback policy in a test greenhouse that regulates the biosensor output to a time-varying reference assuming only a simple heuristic representation of the system's dynamics.

Chapter 2

Biosensor Design

We have developed a heat stress biosensor that uses a fluorescent marker protein induced by an endogenous response pathway in mature *Arabidopsis thaliana*.

2.1 Promoter: pDREB2A

The DREB2A promoter (pDREB2A) is involved (relatively) selectively in the physiological response of plants to periods of drought, elevated temperature, and their combination [26, 29, 35]. Importantly, under stress pDREB2A is active in the leaves of the mature plant, affording rapid and nondestructive access to sample tissue. Since the behavior of this promoter has been extensively validated, we chose it as the transduction unit of our biosensor.

2.2 Marker: Histone2B:YFP

The signaling unit is a fluorescent protein, termed a "marker fusion" because it is a molecule of yellow fluorescent protein (YFP) joined to a molecule of the protein Histone2B (H2B). This fusion, abbreviated "H2B:YFP," exploits the fact that H2B is a protein that does its work in the nucleus of the cell and is therefore transported into the nucleus as it is produced in the cytoplasm. Because the YFP does not interfere with these transport mechanisms, it (YFP) gets carried into the nucleus on H2B's coattails. Thus the YFPs are spatially concentrated in the nucleus and the signal intensity is correspondingly increased. Together the DNA sequences of the transduction and signaling units compose the biosensor's DNA sequence, which is inserted into the DNA of a host plant. That plant and all its offspring are then transgenic hosts for the biosensor, and when the biosensor is activated its signal is detectable in the leaves. The signaling unit can be expressed throughout the growth cycle of the host plant, an important practical consideration suggesting the feasibility of deploying transgenic plants in the field. Yellow fluorescent protein fluoresces upon excitation by ultraviolet (UV) light.

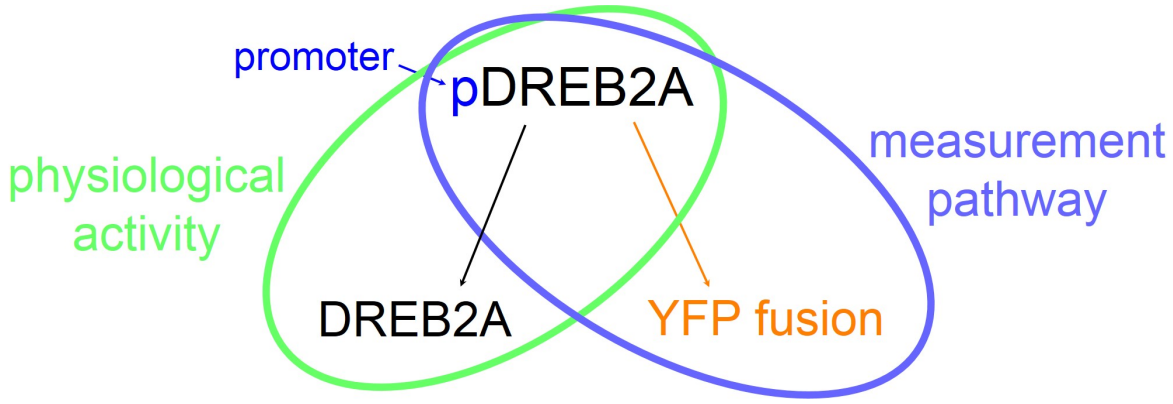


Figure 2.1: Transformation with the biosensor construct adds the measurement pathway in parallel with the endogenous pathway.

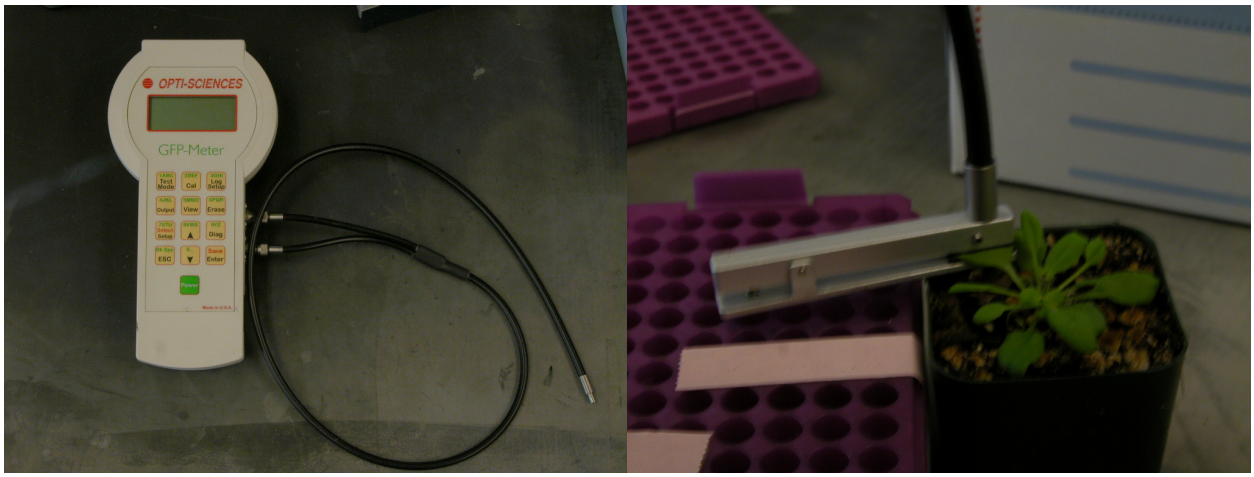
It is important to note that after the *Arabidopsis* host is transformed with the H2B:YFP biosensor construct there are *two* occurrences of pDREB2A in the host DNA for every one pre-transformation (endogenous) occurrence. The endogenous pDREB2A instances are still functional and promote the production of the DREB2A protein under heat stress. The pDREB2A instances that are inserted by transformation promote the production of our marker fusion under heat stress. These parallel pathways are shown in Figure 2.1, and it is important to stress that we are not "overwriting" the endogenous stress response pathways with our construct.

2.3 Fluorescence Measurement

To measure the amount of YFP present in the leaves of our biosensor hosts, we deploy two different COTS fluorescence detectors manufactured by Opti-Sciences, Inc., the GFP Meter (Figure 2.2a, now discontinued) and the GFP III Meter. Both devices are self-contained contact fluorometers, each with a UV source and a photodetector that are connected to the plant by a commingled fiber-optic cable clipped to a single leaf (Figure 2.2b). The excitation spectrum of the GFP Meter peaks at 465 nm, and its detector is most sensitive at 520 nm. For the GFP III Meter, the excitation and detection peaks are similar.

The GFP (III) Meter completes our biosensor system:

- Transduction unit — pDREB2A
- Signaling unit — H2B:YFP, dsH2B:YFP
- Detection unit — GFP Meter, GFP III Meter.



(a) Opti-Sciences, Inc. GFP Meter, shown with the fiber-optic cable attached.

(b) A biosensor host plant with the fiber-optic cable clipped to the youngest mature leaf.

Figure 2.2: The detection unit of our biosensor system.

2.4 Functionality

To confirm the functionality of our biosensor system, host plants were compared to a variety of negative controls. A table of the design objectives with respect to experimental controls is shown in Table 2.1. The "no sample" controls are designed to show that the detector itself will not introduce time- or temperature-dependent measurement errors that could be mistaken for measurements of the signaling unit fluorescence. These are important because some characterizations presented in [19] indicated that the GFP Meter measurements drifted on short time scales. Also, it was personally communicated by Opti-Sciences that temperature variations of more than a few degrees C would be visible in the measurement.

As is evident In Figures 2.3 and 2.4 it turns out that neither of these phenomena are visible in the "no sample" scenarios. This is because during calibration of the meter a certain photodetector count is defined to be zero and this serves as a lower saturation. E.g., if c_1 is the photodetector count defined to be the zero reference, then any photodetector count less than c_1 will be reported as 0 by the meter. It turns out that in these cases the time- and temperature- variation all occurred around a value sufficiently below the calibration zero bound, so the measurement was identically zero throughout.

However, upon introduction of a background fluorescence signal (either by inserting a biosensor host sample with no heat stress profile as in Figure 2.3 or a wild type sample exposed to a heat stress profile as in Figure 2.4) the mean background signal is elevated above the zero bound far enough that fluctuations in the measurement are apparent. The upward drift observed in the wild type in Figure 2.4 is because of yellowing of the leaves that occurred after the long exposure to heat stress. This behavior was observed in many of the wild types we tested, and we did not consider it profound, mainly because it is smaller

Table 2.1: A series of negative controls is designed to demonstrate that the biosensor's transduction and signaling units are functionally expressed in the host plant and the signal is distinct from the background fluorescence measurement.

Control scenario	Expected outcome	Shown in	What it demonstrates
no sample, no heat stress profile	0 output	Fig. 2.3	the detector zero doesn't drift over long time scales
biosensor host, no heat stress profile	steady low-level sensor output	Fig. 2.3	the biosensor doesn't show false positives
no sample, heat stress profile	0 output	Fig. 2.4	detector does not show measurable temperature dependence (in this setup)
wild type, heat stress profile	steady low-level sensor output	Fig. 2.4	no confounding signal under stress
biosensor host, heat stress profile	distinguishable signal	Fig. 2.4	biosensor is functional

by a large margin than the signal detected from a biosensor host in response to the same heat stress profile. But additionally, plants from different seed lines, even wild types, may have very different phenotypic responses to adverse growing conditions. Some wild types we tested yellowed under long heat stress, others did not. Some biosensor hosts yellowed under long heat stress, others did not. It seemed to have more to do with age of the plants than transformation status.

What is stark in these experiments is the obvious functionality of our biosensor system. It does not show false positives (Figure 2.3), and the magnitude of the signal relative to the background is large compared to that of the noise. Also interestingly, the baseline fluorescence of a never-stressed biosensor host plant is indistinguishable from the baseline fluorescence of a wild type. On that point, these time series presentations use arbitrary units for absolute fluorescence intensity in order to facilitate quantitative comparisons across the datasets. Eventually, we will begin using various detectors and calibrations, and introduce a method of adjusting the resulting data in a way that allows quantitative comparisons of the disparate data sets.

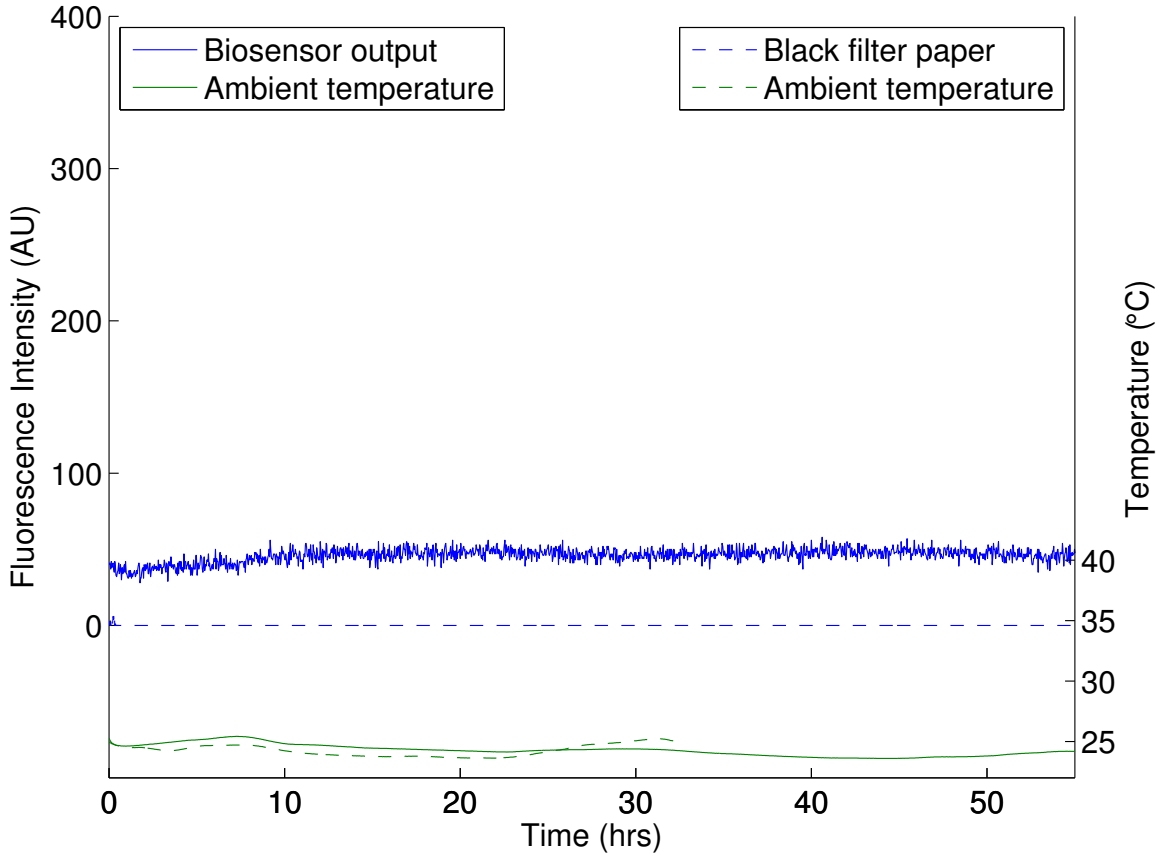


Figure 2.3: The biosensor output exhibits a basal expression above the black filter paper we used as a background in the GFP Meter's leaf clip. The ambient temperature fluctuates with the daily temperature profile of the building.

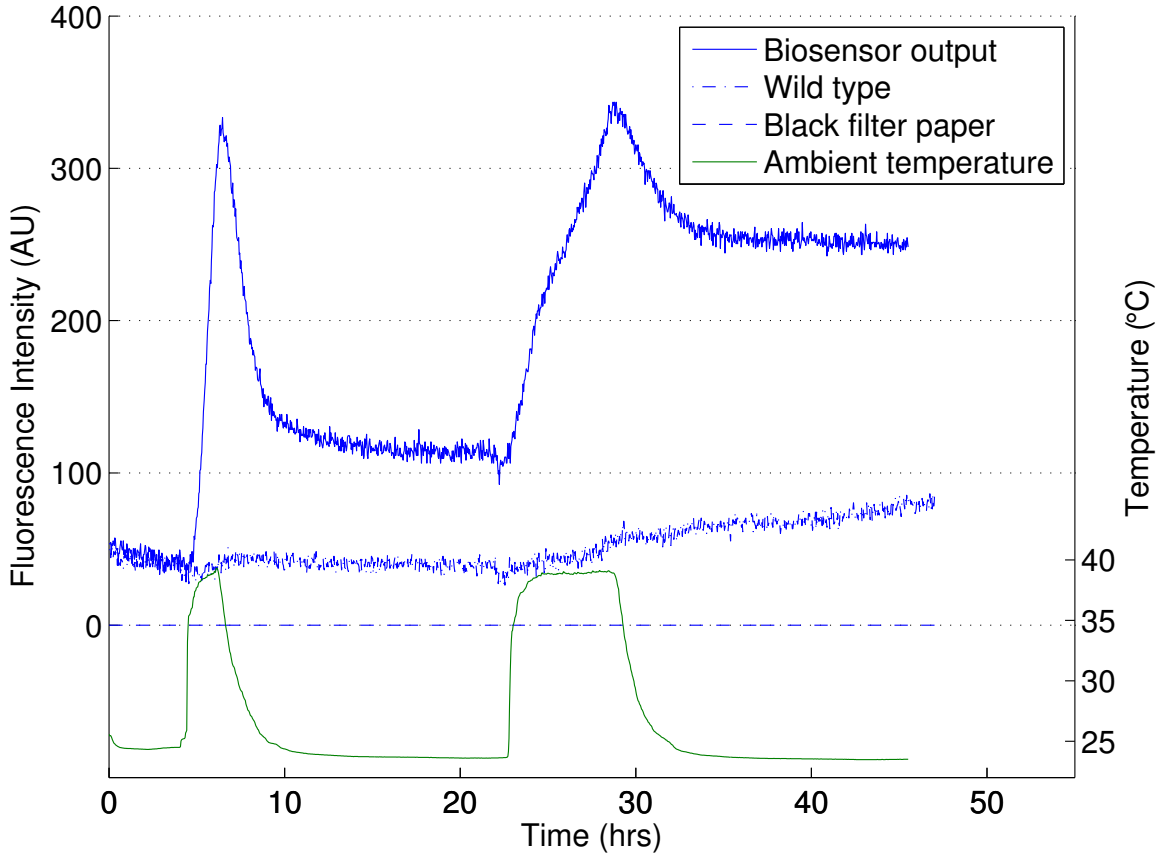


Figure 2.4: When a stressful heat profile is applied the biosensor's response is rapid and large. The strain of wild type that we use as a control exhibits yellowing in the leaves as it ages in the heat chamber, to which we attribute the upward drift in the fluorescence measurement from the leaves of the wild type.

Chapter 3

Experimental Setups

To apply the various heat stress profiles that we use to characterize our biosensor system we deploy two different temperature-controlled environments. The first is a small benchtop heat chamber, and the second is a small greenhouse of our own design and construction. In both cases although the test plant is inside the temperature-controlled test environment it is important to abstract the environments as actuators whose purpose is to apply specified ambient temperature profiles $T_{ref}(t)$ to the plant, as in Figure 3.1.

Figure 3.1: Open-loop characterization of the biosensor system's response to ambient temperature profiles. The temperature profiles is defined by $T_{ref} = T_{ref}(t)$ and regulated by the (inner loop) temperature control system C_T.

3.1 Heat Chamber

The heat chamber is a Precision Thelco Model 2 (Figure 3.2) a small COTS benchtop oven that is commonly used in biological laboratories to store specimens or supplies at elevated temperatures. It has an on-off switch and a built-in temperature setpoint regulator referenced to the position of a rheostat. That built-in regulator is represented by the block C_T in Figure 3.1, and we do not have access to its structure or parameters. The temperature inside the heat chamber is monitored by a trio of Crossbow MTS-420 wireless motes, Figure 3.3, each of which includes a temperature sensor. The temperature measurements are logged

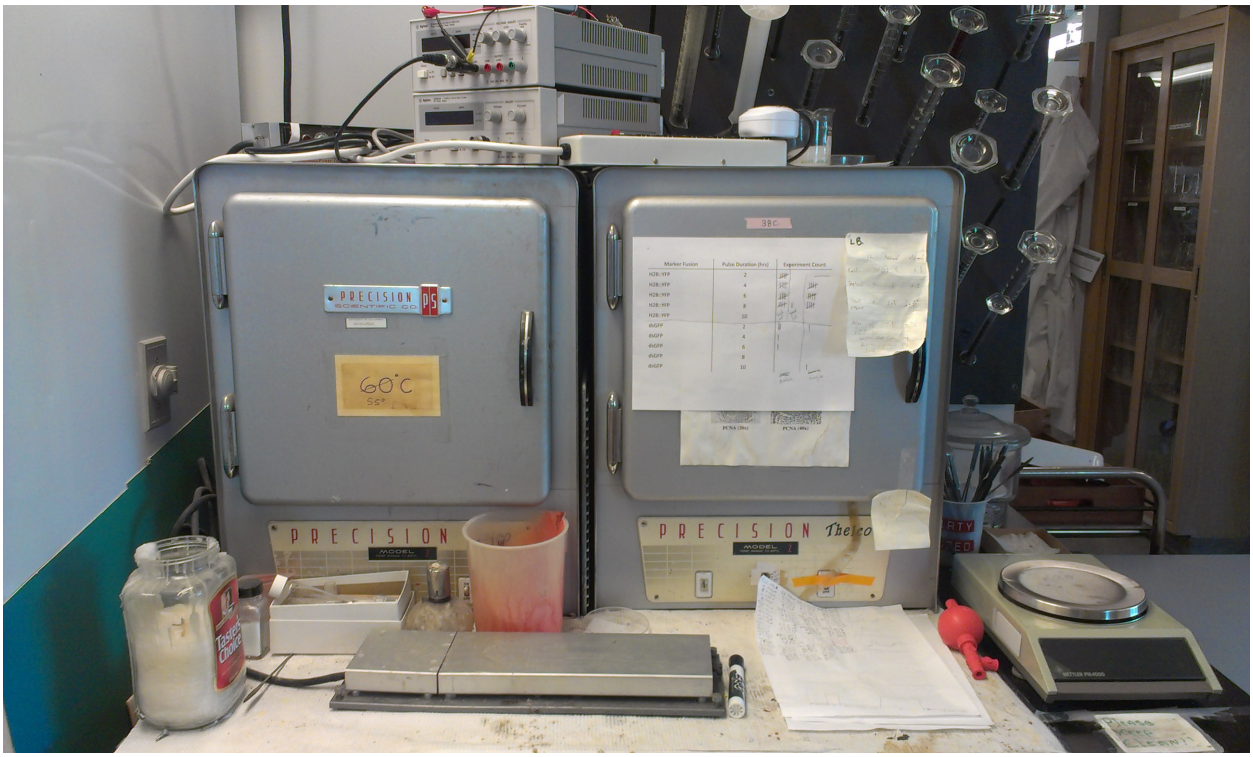


Figure 3.2: A pair of Precision Thelco Model 2 ovens. We use the oven on the right. The GFP Meter power supplies are visible on top.

to a database for later access. The heat chamber is practically limited to tracking on/off profiles, so going forward when data is presented showing open-loop temperature profiles that resemble pulse trains it is understood that those profiles were applied in the heat chamber. The greenhouse was reserved for more complicated open-loop characterization profiles, as well as for the closed-loop experiments.

The rheostat is fixed so that switching the oven on regulates the temperature to about 38°C. When the oven is switched off the interior cools passively to room temperature, a process that is accelerated by the through-hole in the top of the chamber (originally for a thermometer) through which the GFP Meter's fiber-optic cable passes. The full set of characterization temperature profiles is shown in Figures 3.4 and 3.5, with vertical dashed lines marking ON/OFF events. Notice the linear temperature response dynamics. The equal-width pulse trains shown in Figure 3.4 are designed to characterize the behavior of plants to the periodic temperature fluctuations that might occur during a diurnal cycle. When using the profiles shown in Figure 3.5, we are interested in how the response of varies after the second pulse (the one of varying width) since all the plants were "primed" with an identical leading pulse. The second "on" event of each pulse train follows sixteen hours after the first "off" event. The on/off events are triggered by a programmable power outlet with a somewhat unreliable clock, so the actual on/off times varied slightly.

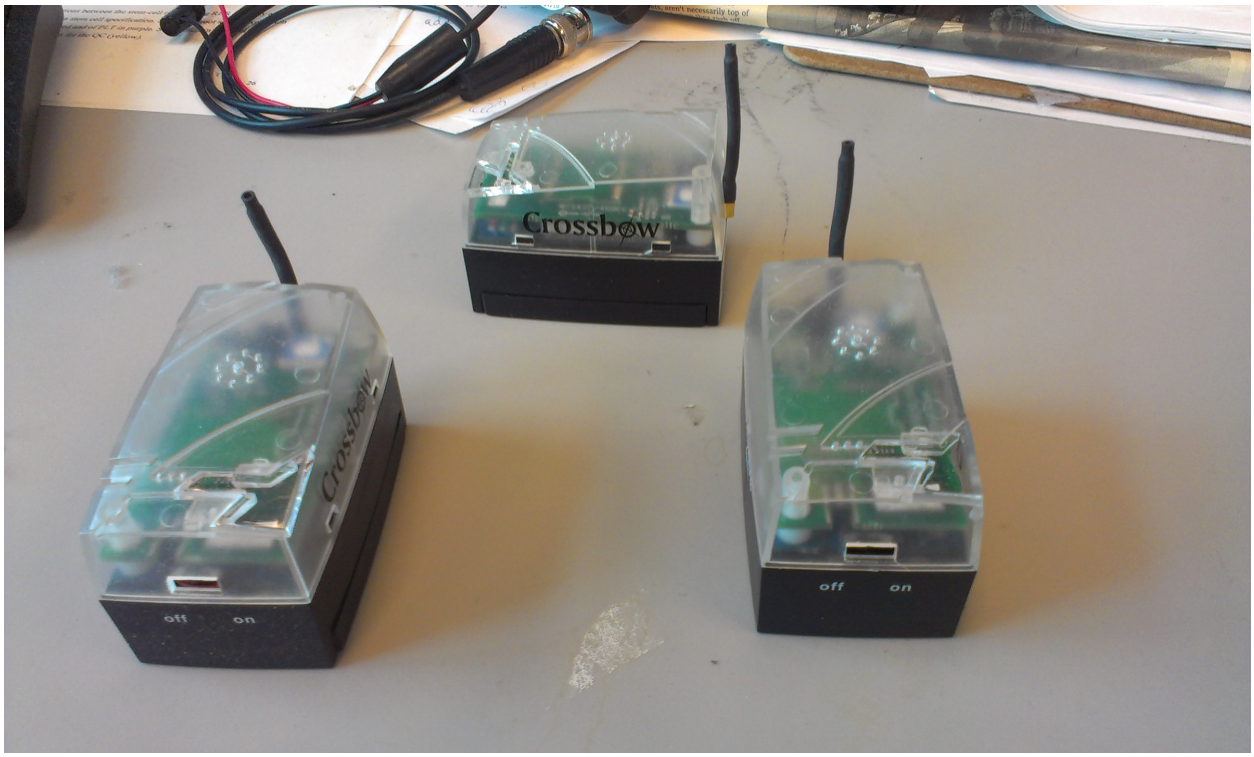


Figure 3.3: Three Crossbox MTS-420 wireless sensor motes.

3.2 Greenhouse

The greenhouse is a plastic structure divided into five bays, of which we have instrumented and actuated one. The greenhouse is shown in Figure 3.6, and a cross-section of the actuated bay is shown in Figure 3.7.Unlike the heat chamber, the greenhouse is cooled by force convection, so it can maintain temperatures slightly below room temperature, and can also transition from high temperature to low temperature very quickly. Above room temperature the temperature is controlled by two radiative heaters, a ceiling vent, and a sealable sidewall fan. Each heater has a fan behind it to facilitate temperature uniformity throughout the volume by maintaining air circulation. The temperature is monitored by a mesh of temperature-sensitive integrated circuits distributed throughout the volume. All interfaces are through a LabVIEW cRIO chassis, which enables precise tracking of time-parameterized temperature profiles using a controller (C_T in Figure 3.1) of our own design. Consequently in the greenhouse we are not limited to applying on/off pulse trains.

The greenhouse has a large enough volume that the fluorometer needs to be inside, exposed to the changing temperature profile. Although the GFP Meter does not show significant temperature-dependent fluctuations in the air-conditioned environment containing the heat chamber, it fluctuated drastically when exposed to the full 18°C temperature range of a heat stress profile in the greenhouse. For this reason the GFP Meter was never again

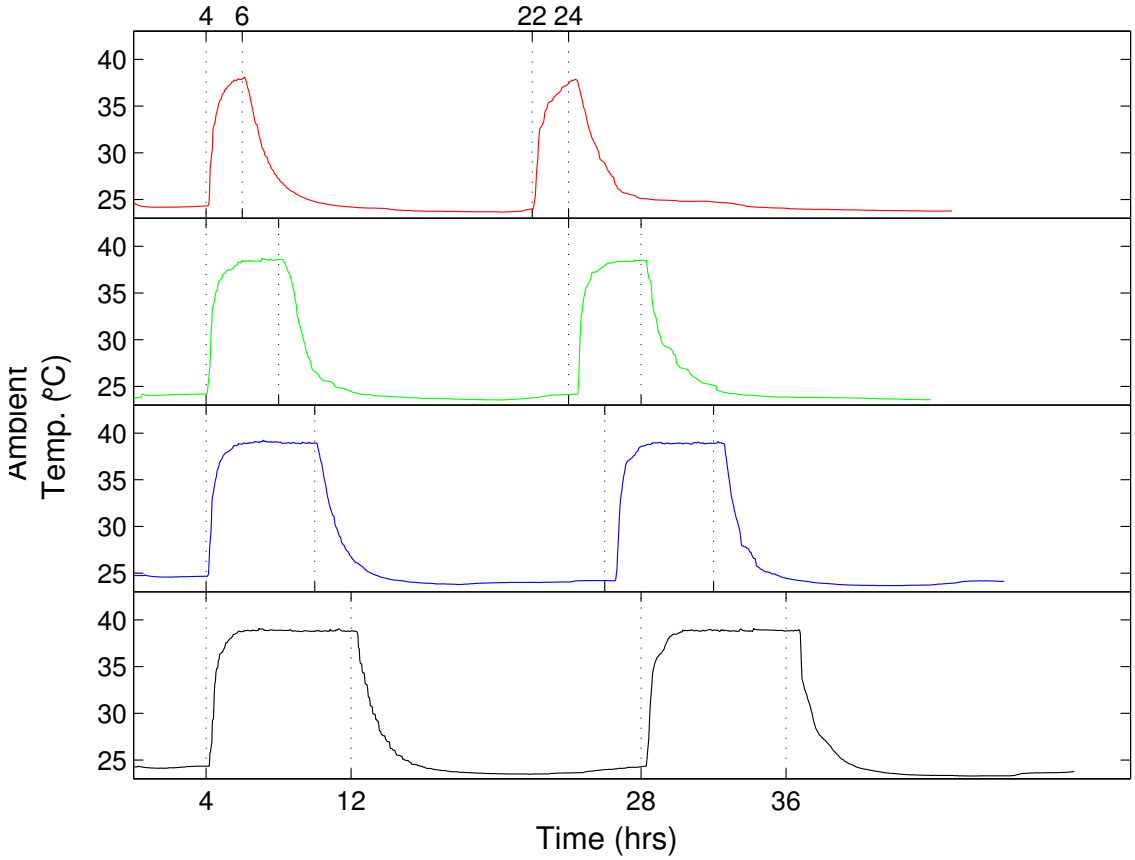


Figure 3.4: Two identical periods of exposure to heat stress.

used in the greenhouse. That role was filled by the GFP III Meter, which has internal temperature correction. The other functional difference between the GFP Meter and the GFP III Meter is the calibration method. The GFP Meter uses a two-point slope/bias calibration, while the GFP III Meter uses multi-point linear interpolation. In order to compare the devices quantitatively we (a) calibrate the GFP III Meter with interpolation points that are all higher than the signal range of our biosensor so that the calibration is slope/bias in the relevant range, and (b) present fluorescence measurements in terms of the *adjusted fold change* above baseline, a topic on which we will expound in section 4.2.

Both the GFP Meter and the GFP III Meter required some customization, accomplished very attentively by the Opti-Sciences engineering team. The GFP Meter was not able to execute repeated measurements out of the box, a problem that we addressed temporarily by building a rather entertaining Lego Mindstorms robot to push the "take data" button at regular intervals (Figure 3.8). Ultimately we had the meter customized to add this functionality. However, there were still two characteristics that demanded consideration

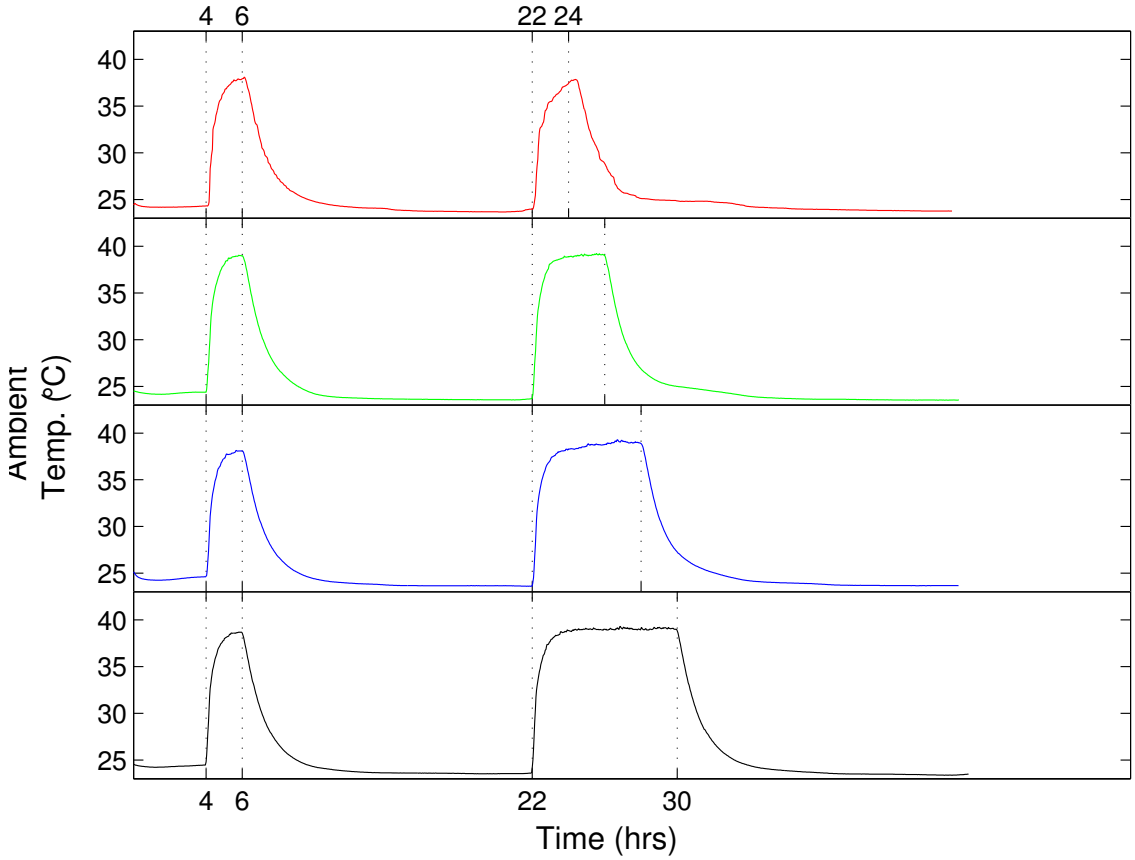


Figure 3.5: Varying exposure to heat stress after a two-hour priming stress.

when designing our experiments.

1. Each measurement is taken in a three-second cycle during which time the UV source is switched on and stabilized, the detector is queried, and the UV source is switched off.

2. The GFP Meter's onboard memory is quite limited and the firmware does not permit real-time transmission of data to a processing computer.

The cycle time constraint imposes one measurement/three seconds as a hard lower bound on our sampling rate. Ironically this lower bound is not achievable for long-running experiments because the YFP degrades under constant and prolonged exposure to UV light. Using an actual sample rate of one measurement/one minute did not exhibit adverse effects, but then we were constrained by the onboard memory size. Ultimately our experiments were standardized at a sample rate of one measurement/two minutes. This is still faster than

Figure 3.6: The test greenhouse, clearly showing five independently actuated bays. At the bottom of each bay is a sealable fan corresponding to the control input u_f in Figure 3.7. A the top of each bay, on the near pitch of the roof, is an actuated vent corresponding to the control input u_v in Figure 3.7. The cRIO chassis and interface computer shown at right.

the time scales of transcription and translation, enabling the detection of transients in the sensor output.

The GFP III Meter was released with repeated measurement as a built-in mode option, but the measurement cycle was altered so that when in that mode the UV source was on constantly. Opti-Sciences again made some custom revisions to return to the three-second measurement cycle, and also added computer-commanded data collection over USB. Those changes allow us to integrate the GFP III Meter into the LabVIEW interface.

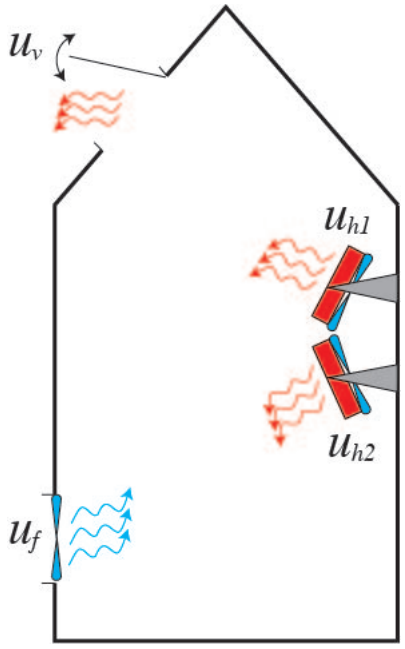


Figure 3.7: The actuation available in our test greenhouse. The commands u_{h1} and u_{h2} are PWM signals to the heaters, and whenever the PWM has nonzero width the associated fans are on. The greenhouse is cooled by opening the ceiling vent to angle u_v and/or uncovering and running the sidewall fan at speed u_f.

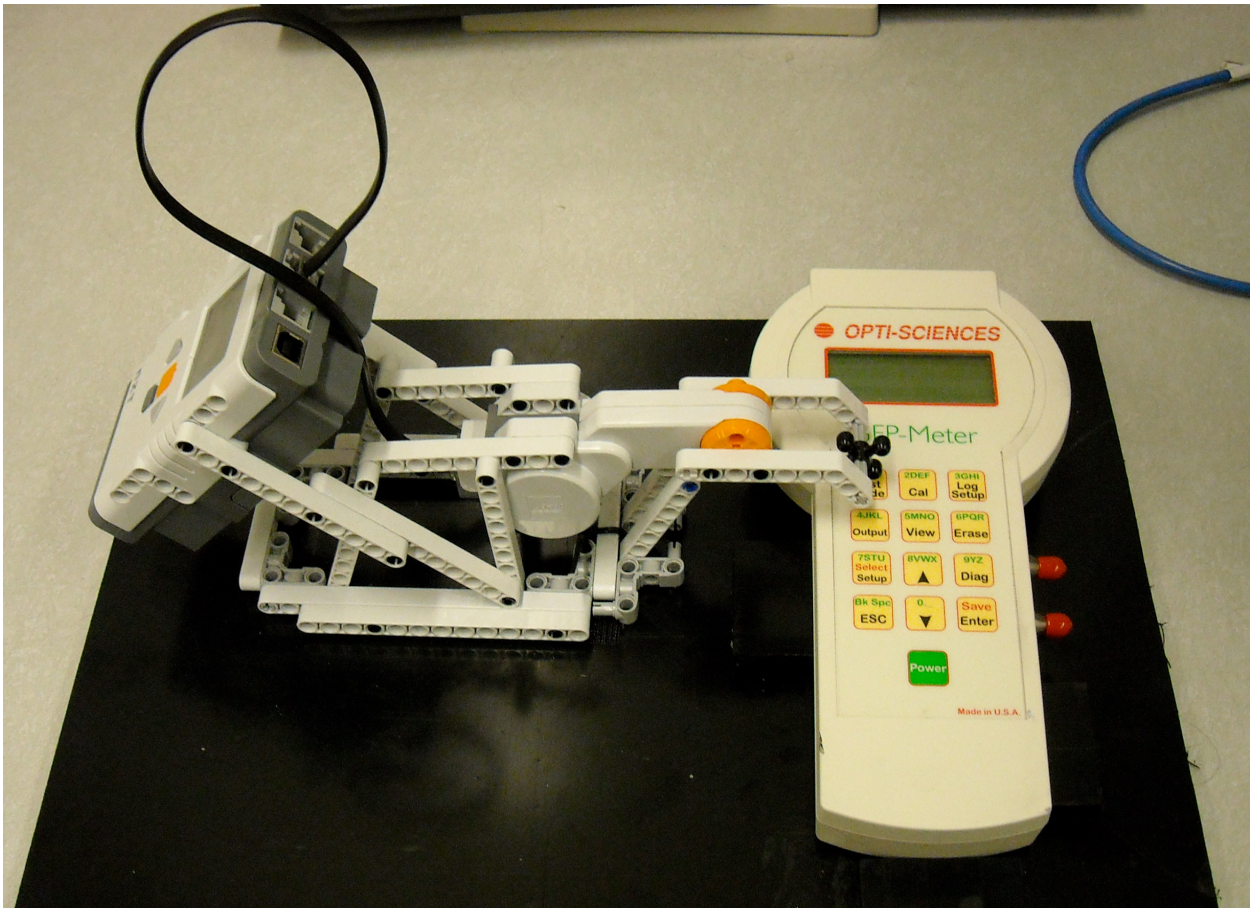


Figure 3.8: Prior to the firmware revisions to the GFP Meter, a small Lego Mindstorms robot was responsible for pushing the "take data" button at the specified sampling rate.

Chapter 4

Characterization

There is consensus in the literature surrounding *Arabidopsis* heat stress about temperatures that are relevant to pDREB2A [29, 35, 8]. It is not active below 30°C and it is active above about 33°C. Temperatures above 36°C are used to induce heat stress. This consensus is vague and implicit, but it will guide characterization of our biosensor.

The characterization will proceed in several steps. It was demonstrated in the previous chapter that the biosensor is functional in mature *Arabidopsis* plants. Going forward, we will first benchmark our biosensor system against two other existing methods of measuring gene pathway activation. Next, we will demonstrate that the transgene is stable across multiple generations, with similar temperature input profiles resulting in *quantitatively* similar biosensor output profiles. We will then exploit the high temporal resolution of our detection system to take a detailed look at the biosensor's response time, comparing our results to observations in the literature using chemical inducers to estimate the latency upstream of our (environmental-factor-induced) biosensor. We then present a brief study of the sensitivity of the biosensor to small temperature fluctuations, at temperatures both below and above the temperature range commonly accepted as the "threshold" for pDREB2A activation. Finally, we study the dynamic response of the biosensor system to cyclical temperature input profiles and observe a behavior resembling memory.

4.1 Benchmarking against existing techniques

There are several existing systems that, like our biosensor system, measure the transcription and subsequent expression of a fluorescent marker protein. Our system makes those measurements with uniquely high temporal resolution, but because it is a new arrival on the marker protein measurement scene it must be validated against existing techniques. We will compare our results to those obtained with optical microscopy and by the polymerase chain reaction (PCR). In the former case, our measurements are directly comparable. In the latter case comparison requires the acceptance of a simple dynamic model of mRNA translation.

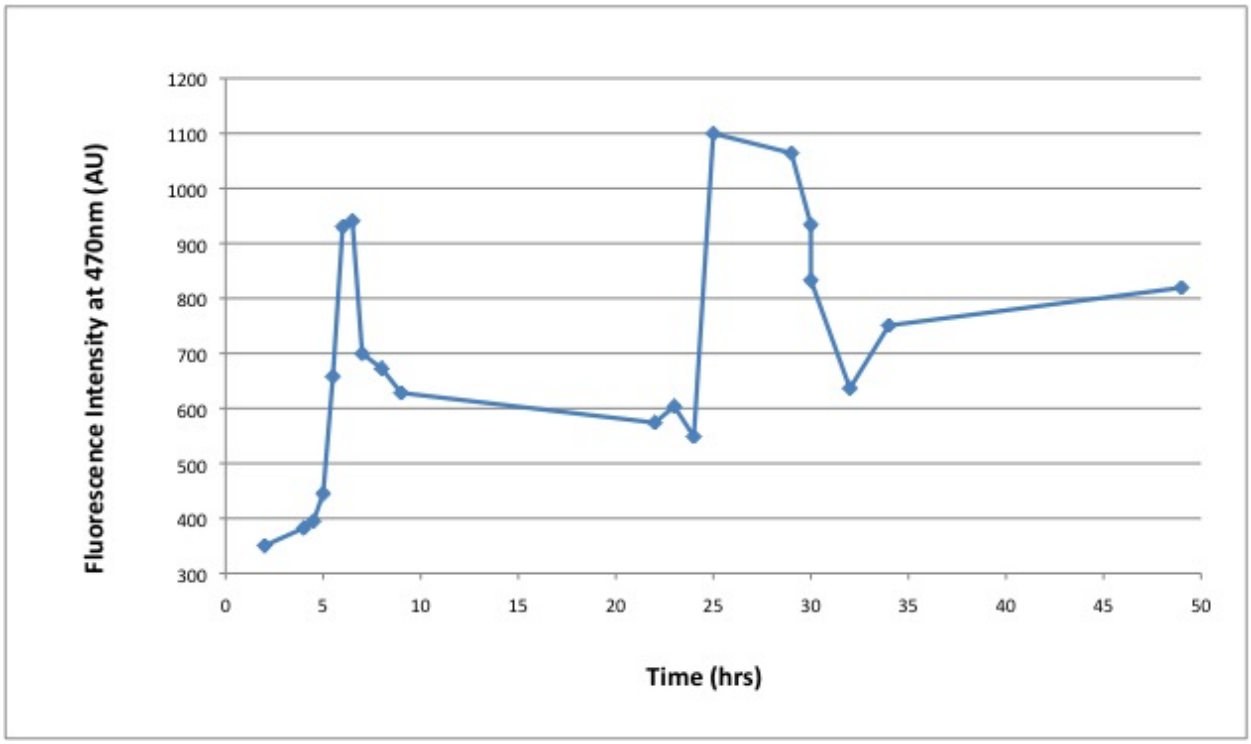


Figure 4.1: Time course measurements: optical microscope.

Optical microscopy

Researchers at UNC Greensboro carried out an extensive comparison of the GFP Meter to a commercially available spectrofluorometer [19]. They observed correlations between the GFP Meter output and the spectrofluorometer output of greater than $R^2 = 0.87$. We performed a similar experiment at a smaller scale, comparing the output of the GFP Meter to the fluorescence intensity computed by integrating the brightness of an image of the leaf tissue under an optical microscope. The experiment proceeded by placing two biosensor hosts in the heat chamber, with the GFP Meter clipped to a leaf of one of the plants. Single leaves were removed from the other plant at specific times through the duration of the temperature input profile and preserved for subsequent analysis with the microscope. The times at which the microscope samples were removed were chosen to capture the various features that we expected to see in the data, in particular the shapes of the responses to the two pulses and the steady-states that were attained after each return to room temperature.

Because the plants are *in situ* during the experiment, the acquisition of each microscope sample required human intervention. The data set was consequently much smaller than that presented in [19]. But even so, the data in Figures 4.1 and 4.2 clearly show a strong correlation ($R^2 = 0.85$) between the microscopy measurements and the GFP Meter. Most importantly, this experiment validates the GFP Meter's observation of persistent baseline

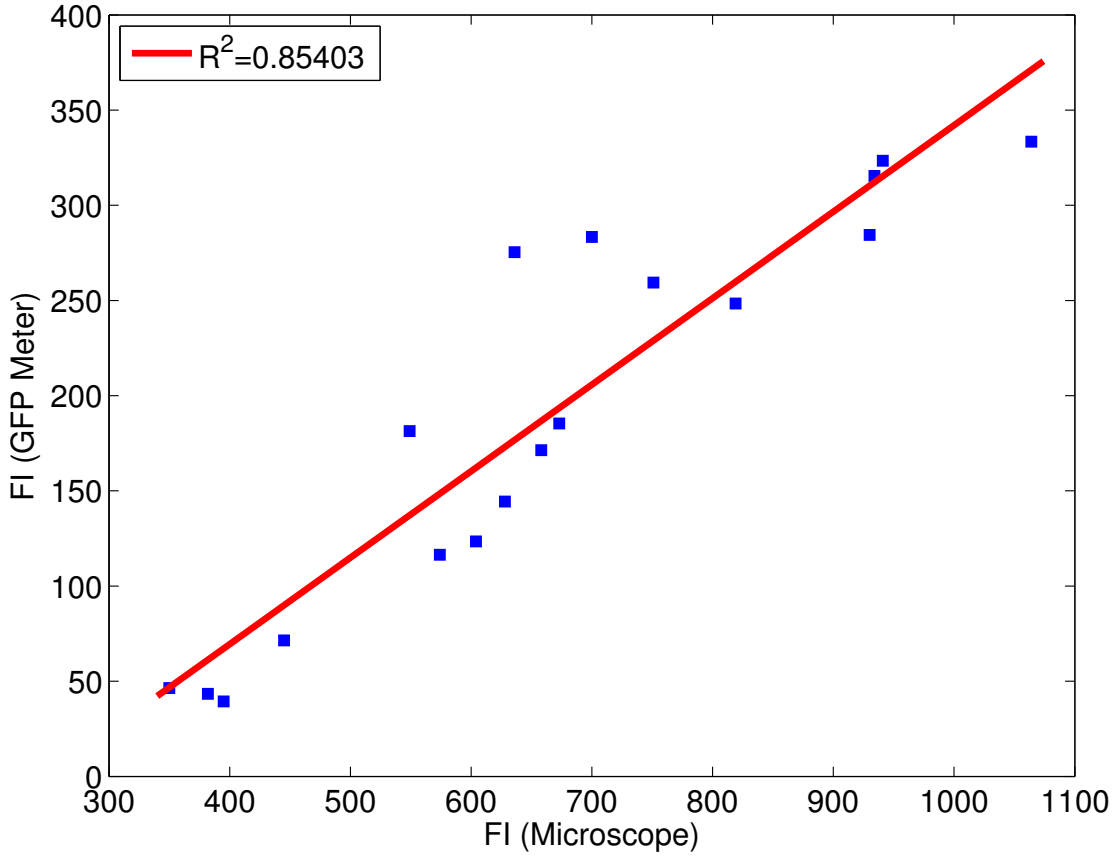


Figure 4.2: The GFP Meter and the optical microscope are well correlated. GFP Meter data shown in Figure 2.4.

elevation after the removal of heat stress.

PCR

The polymerase chain reaction (PCR) is a technique that enables concentration of a gene sequence in a sample solution. In order to validate the output of our biosensor system, the gene sequence occurrences to be counted are those of the mRNA molecules that code for YFP. Speaking very broadly, the concentration of functional protein molecules is directly related to the time integral of the concentration of coding mRNA molecules, and the measured fluorescence is directly related to the concentration of functional protein molecules.

To benchmark our biosensor system against PCR we took sequential samples from several plants that were simultaneously exposed to the same heat stress profile, in much the same manner as for the microscopy experiment. However, in this case more sample material was

necessary. In fact, one data point in the PCR benchmarking required the leaf matter from an entire plant. So the number of data points was limited to the number of plants that could reasonably fit in the benchtop heat chamber: seven (eight plants, but one was hooked up to the GFP Meter).

Figures 4.3 and 4.4 show the results of two different PCR benchmarking experiments. PCR measurements are naturally expressed in dimensionless fold changes above a basal level, so for consistency in these plots the GFP Meter output is also presented as dimensionless fold changes above the initial baseline level. In both experiments, the measured fluorescence intensity is qualitatively consistent with the theoretical relationship between mRNA transcript concentration and function protein concentration. That is, the densely sampled biosensor output (solid line) resembles the time integral of the sparsely sampled data PCR data (bars).

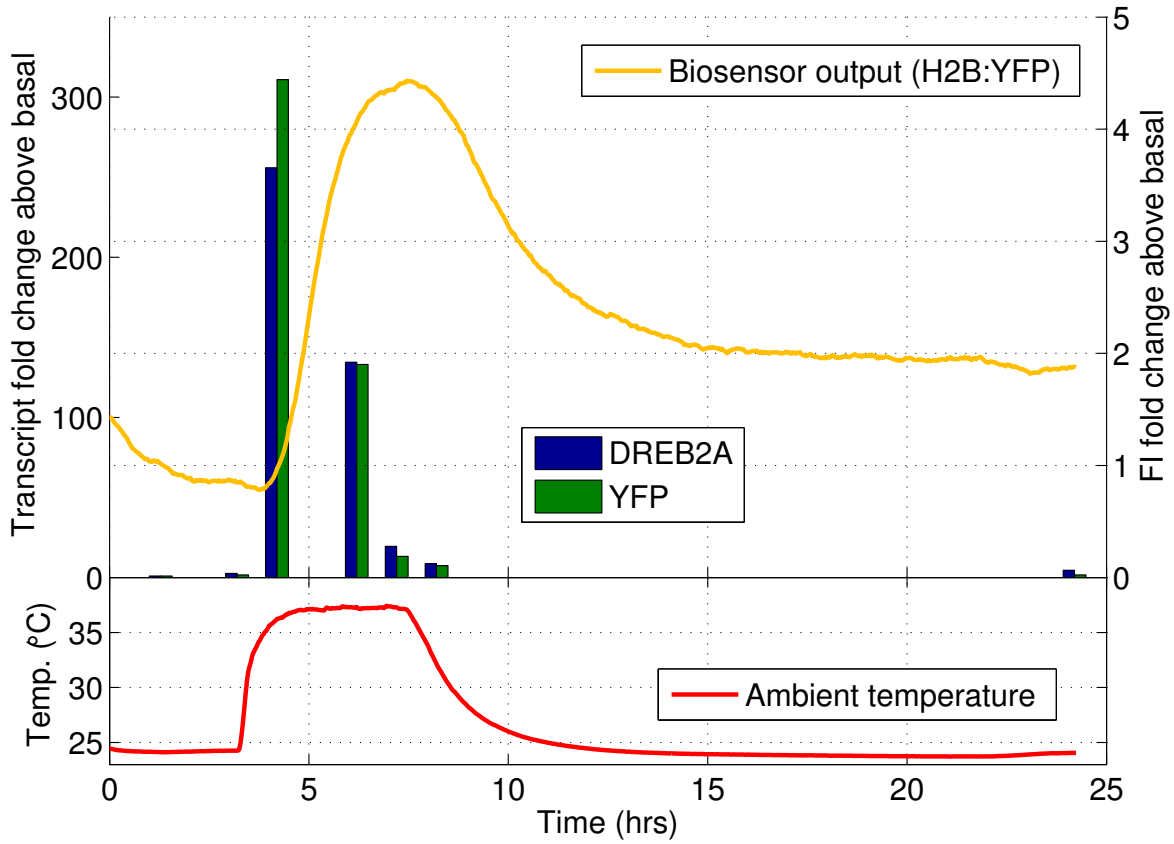


Figure 4.3: The peak in the PCR data at 4 hours is consistent with the fast increase in the biosensor output occurring immediately thereafter.

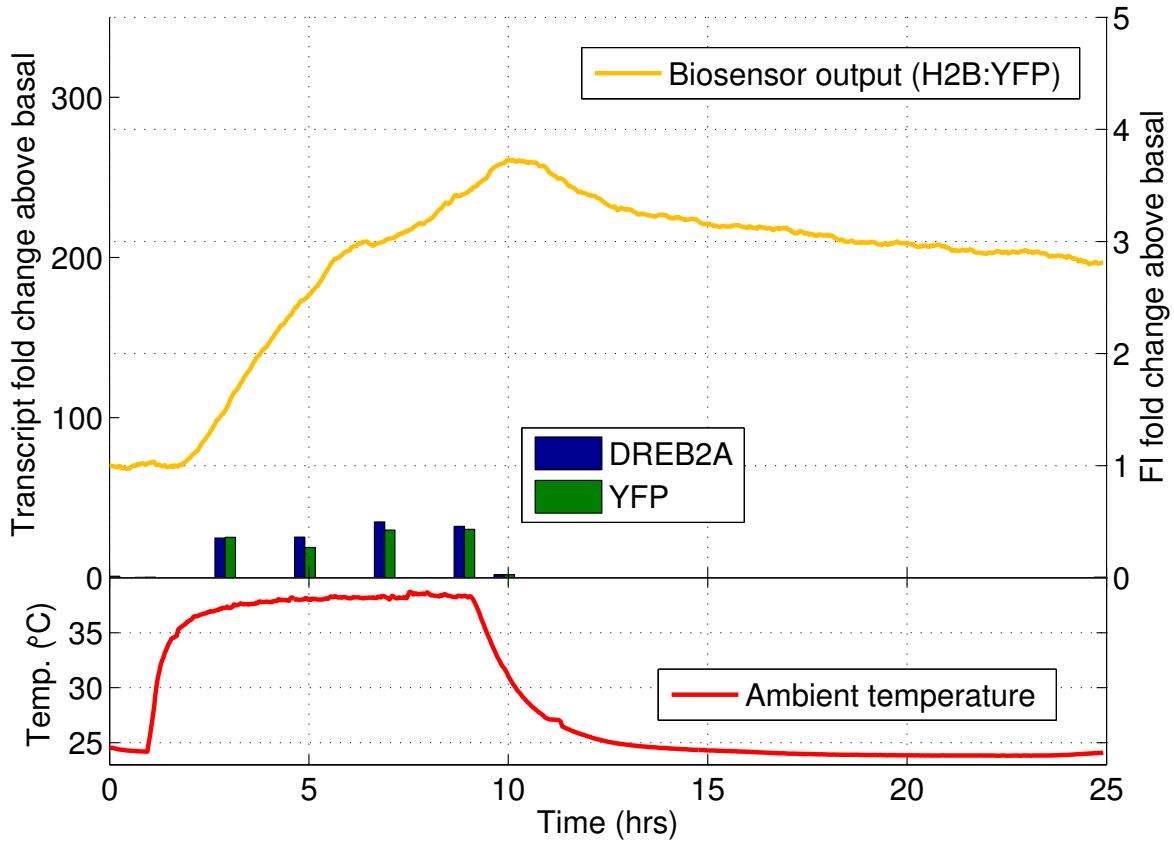


Figure 4.4: In this case theYFP expression level is uniformly lower, resulting in a uniformly slower rise in the biosensor output.

The correspondence is particularly pleasing because it holds true across the evident differences between these two plants. The plant in Figure 4.3 has a much more rapid initial response to the onset of heat stress, and that response is clear both in the PCR data and the biosensor output. In Figure 4.4 the plant responds more slowly to the onset of stress and again this fact is reflected in both the PCR data and the biosensor output.

In addition to the YFP transcript, Figures 4.3 and 4.4 also show PCR data for the DREB2A transcript. Recall that the DREB2A promoter is the transduction unit of our biosensor system, and that the mRNA coding for the YFP fusion is produced whenever pDREB2A is active. But since pDREB2A is an endogenous promoter in *Arabidopsis* the mRNA coding for the protein DREB2A is also produced, and indeed this is the protein that is physiologically relevant in the plant's heat stress response pathways. The *quantitative* similarity between the YFP and DREB2A transcript concentration is a very important fact in its own rite, indicating that the activation characteristics of our biosensor's signaling

unit are very similar to those of DREB2A. This supports the claim that our biosensor measurements are a proxy for the activity of the heat stress pathway in which DREB2A plays a role.

Finally, to place these data in the context of the literature, in [35], a PCR analysis is presented that includes the DREB2A transcript abundance during a period of prolonged exposure to elevated temperature (reproduced in Figure 4.5). In particular, the 37°C profile was applied to the plant sample for the entire 24-hour duration shown in the plot. The transient peak in the transcript abundance immediately after the onset of heat stress is very similar to the transient peak in Figure 4.3, as is the subsequent decrease over the first few hours of the heat stress event. But the measurements from [35] bear little resemblance, even qualitatively, to the data in Figure 4.4, highlighting the fact that there is significant biological variability in the response of these plants to heat stress, particularly in the nature of their transient response to the onset of stress.

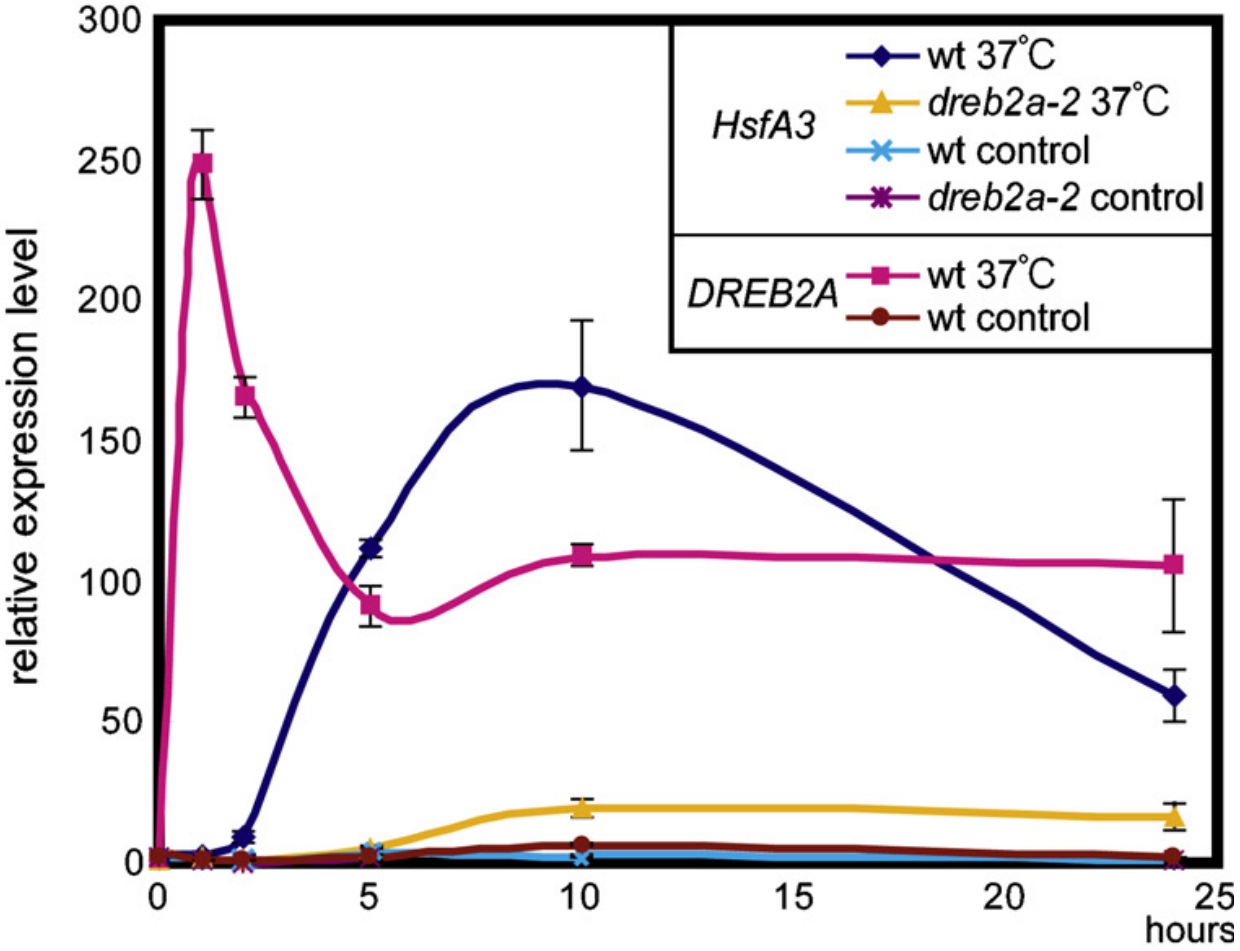


Figure 4.5: Taken from [35].

4.2 Stability of the transgene

Another important practical consideration is *stability* of the biosensor transgene, meaning that after several generations the inserted gene still exists in the plant's genome. This cannot be taken for granted. When a physiologically *necessary* gene is removed by random mutation, the affected plant will likely suffer deleterious effects or die, in any case removing itself from the gene pool and discouraging that random mutation, leading to the persistence, or stability, of necessary genes. However, when a gene that is not physiologically necessary (such as the YFP that we use as our signaling unit) is randomly mutated away the plant's reproduction will likely not be adversely affected and the mutation will persist, diluting the concentration of that gene in future generations. In our biosensor we have taken steps to encourage stability, both by using a physiologically relevant promoter as our transduction unit and by using Histone2B, an important protein involved in repairing damaged DNA, to effect the nuclear localization. By using building blocks that have survived thousands of generations of natural selection, we increase the likelihood that our biosensor system will persist.

We tracked the transformed line through five generations, shown in Figure 4.6 along with the naming conventions that we use to describe the lineage of each plant.

Aside: calibration, normalization, and filtering

In our experiments, the GFP Meter and GFP III Meter both use two-point linear interpolation calibration, i.e. if $\mathbf{x}$ is a time series of some intrinsic fluorescence level to be measured by a meter with calibration C_i, the measurement is $\mathbf{y}_i = a_i\mathbf{x}+b_i$, where a_i, b_i are scalar constants particular to calibration C_i. An issue raised in [19] about the GFP meter is unrepeatability of calibration, and that was observed by us as well. But we can undo the calibration using some internally consistent normalizations. For example, say we measure the same underlying $\mathbf{x}$ with two different meters, each with its own calibration C_1 or C_2. Denote by y_{ib} the baseline fluorescence measurement and by $\bar{y}_i$ another reference measurement chosen above the baseline. Then

$$\frac{\mathbf{y}_1 - y_{1b}}{\bar{y}_1 - y_{1b}} = \frac{\mathbf{x} - x_b}{\bar{x} - x_b} = \frac{\mathbf{y}_2 - y_{2b}}{\bar{y}_2 - y_{2b}}$$

For this reason the quantity

$$\frac{\mathbf{y} - y_b}{\bar{y} - y_b}$$

is *calibration-independent*, so we can compare this quantity across data sets, even when the data sets were gathered with various calibrations, for an accurate comparison of the intrinsic fluorescent protein expression.

The usefulness of this scheme relies on an appropriate choice of $\bar{y}$. Because we are actually comparing measurements of different underlying datasets $\mathbf{x}_i$, a good choice of $\bar{y}$ is one that corresponds to a quantity that is expected *a priori* to vary only slightly across the different $\mathbf{x}_i$. For the cross-generation comparison, since each plant saw the same 2-hr/6-hr heat stress

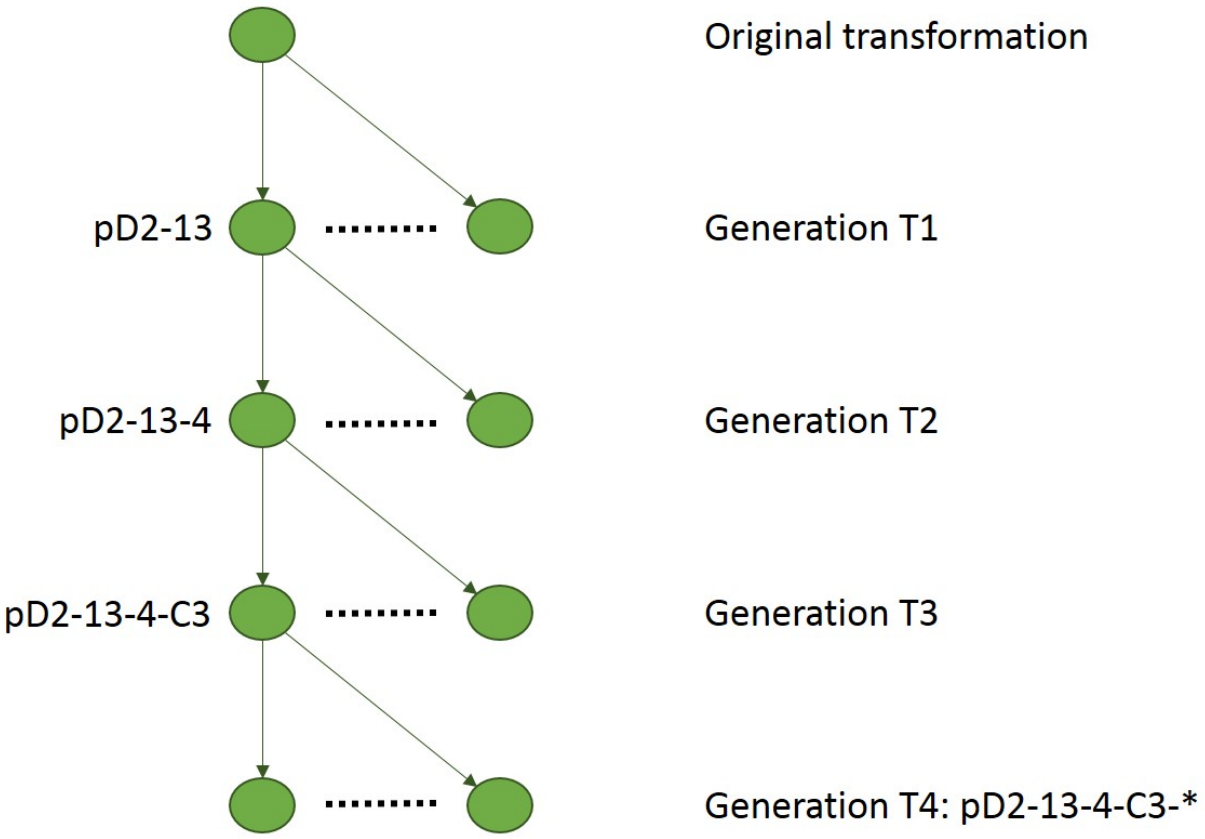


Figure 4.6: Family tree of the pD2:H2BYFP plants.

profile, we choose to set $\bar{y}$ to be the average steady-state value attained after the relaxation from the two-hour heat pulse. The basal measurement y_b is the average of the measurements from the first 4 hours of each experiment when the plant was resting at room temperature. See the annotations in Figure 4.7.

Figures 4.8 and 4.9 show this fold change adjustment in action. Figure 4.8 shows the raw data from two series of 2-hr/6-hr experiments on T4 generation plants, one performed with the GFP meter and other performed the GFP meter. The dark line and vignette ribbon represent the mean and standard error, respectively, of measurements from 5 different sibling plants.

Finally, the noise characteristics of the GFP meter are such that the absolute magnitude of the noise on the fluorescence signal is independent of the magnitude of the signal, resulting in a lower signal/noise ratio (SNR) from the GFP III Meter due to its calibration. However, the noise is at a sufficiently higher frequency than the genetic transients that prior to data processing the GFP III Meter data can be passed through the third-order roll-off filter without introducing appreciable delay. The filtering sequence is described pictorially in Figure 4.10;

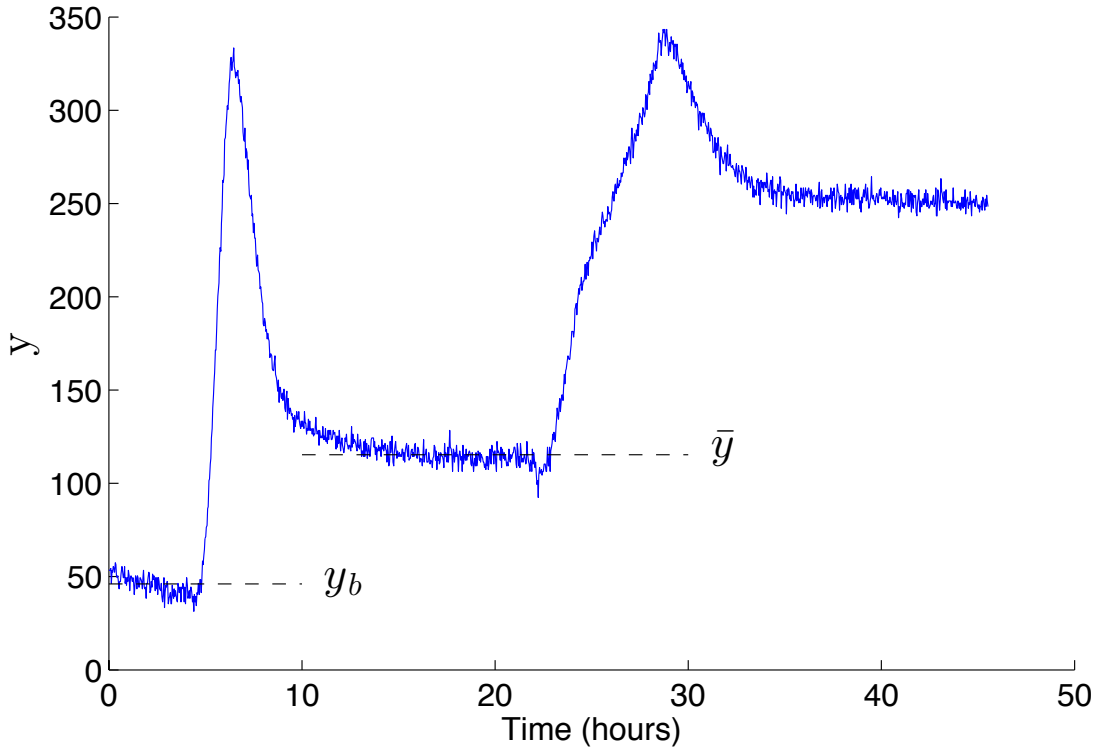


Figure 4.7: Time series features used to compute the adjusted fold change.

Cross-generational comparison

Bringing this section together, Figure 4.11 shows the adjusted fold change of the biosensor output in response to a 2-hr/6-hr heat stress profile, measured in hosts from generations T2-T4. As in Figure 4.9, the dark lines and matching vignette ribbons each represent the mean and standard error of 5 different sibling plants, so this plot represents 20 different biosensor hosts. (I stress this here to point out that a measurement is taken every two minutes, so this plot contains approximately 28,800 individual data points, gathered for the cost of 20 *Arabidopsis* specimens. This is essentially free.)

The *quantitative* similarity of the adjusted fold change across successive generations indicates that the biosensor transgene is stable. The mean and standard error of the biosensor output from all 20 specimens is shown in Figure 4.12. In Figure 4.11 it appears that the transient in response to the first heat pulse is where most of the variability between host plants is concentrated. However, in Figure 4.12 it is clear that it is rather the response to the second heat stress event that displays more variation between hosts.

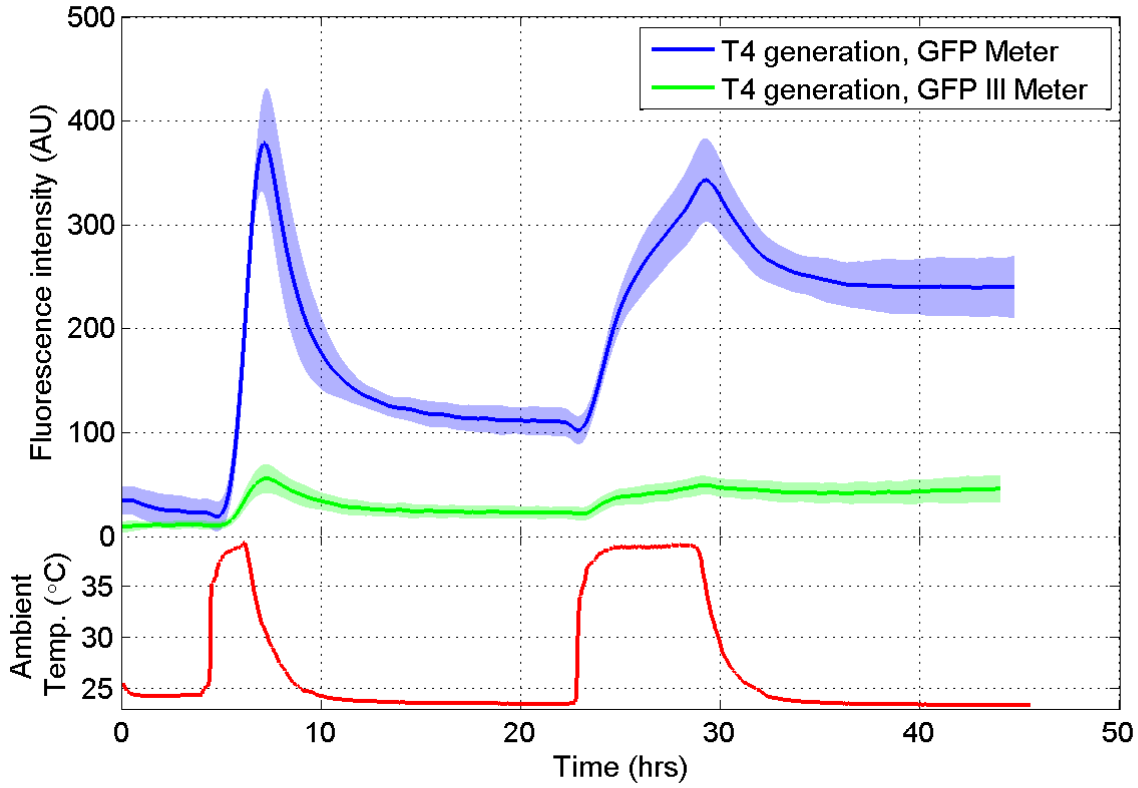


Figure 4.8: Because the calibrations of these two meters are completely different, the raw data taken from both in response to a identical stress profiles differs by an order of magnitude.

4.3 Response time

The response time of the biosensor is the time from activation to detectable change in output. This time is the sum of two shorter times:

1. the time between the onset of stress-inducing temperature and the first binding of the DREB2A promoter; and

2. the time between binding of pDREB2A and detection of YFP fluorescence.

For our biosensor system, the total time from the onset of stress to a detectable change in output is less than an hour.

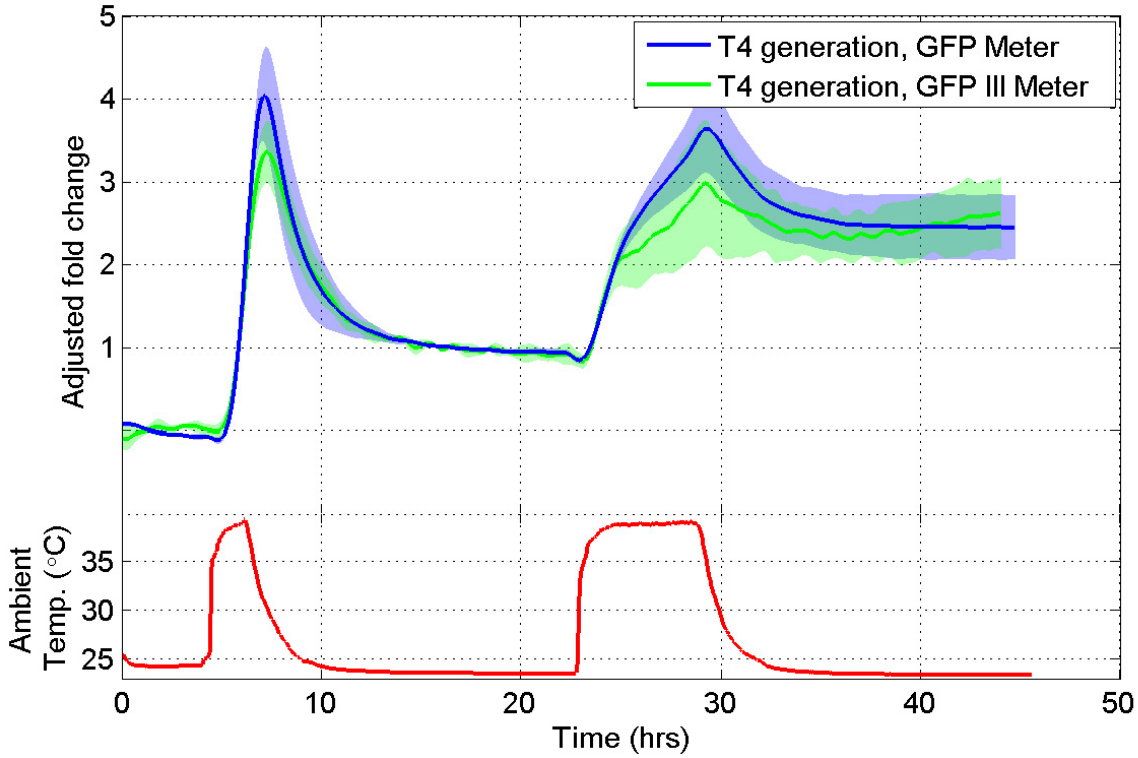


Figure 4.9: Fluorescence intensity: adjusted fold change

4.4 Sensitivity to small temperature changes

Staircase inputs

With respect to the conventional wisdom about the turn-on temperature of the DREB2A pathway, Figure 4.13 is a clear demonstration that relatively large changes occur in the biosensor's output somewhere between 30°C and 34°C.

For more detail, Figure 4.14 presents the results of a series of experiments with a smaller temperature step size, suggesting that the biosensor output can be distinguished from wild type background fluorescence around 32°C, and that some plants may exhibit transient heat stress response activity at temperatures as low at 28°C.

An interesting observation about these experiments is the quantitative consistency across the biosensor's output, even at these low temperatures. Of special note is that even when a plant has a large transient spike at very low temperature (as in the blue response trace in Figure 4.14) the response is actually caught and corrected, and the biosensor output returns to a level similar to that exhibited by plants that didn't exhibit the same aggressive transient.

4.5 Memory of past heat stress

Having quantitatively and qualitatively validated our measurement system against accepted techniques and determined that the range of activating temperatures is wider than previously thought, we now turn to characterizing our biosensor system in terms of its response to various heat stress profiles. Steady state response to a single heat shock is interesting, but we chose to go slightly farther and study the response of the biosensor to two-pulse trains (as have already been shown several times, starting with Figure 2.4). With enough rest time between responses, this allows us to determine single-pulse response characteristics as well as the "reset behavior" of the biosensor. Our first experiment series uses trains of two identical heat pulses, to 38°C. The experiments open with a four-hour measurement phase where the basal biosensor output is established, and the pulses are each followed by sixteen hours at room temperature. The long period following each pulse allows the observation of the decay dynamics of the biosensor, and any difference between the responses to the two pulses will point to ways in which heat stress acclimation alters subsequent stress response.

Twenty plants, grown in identical conditions from seed gathered from the same mother plant, were separated into four groups of five, and each group assigned a pulse duration of two, four, six, or eight hours. Each plant was then individually placed in the heat chamber and an input consisting of two pulses of the specified duration was applied as described above. In this manner five plants saw two two-hour pulses, five plants saw two four-hour pulses, etc. The duration of the heat pulse is measured from the time the heat chamber's temperature control system is turned on until the time it is turned off, not accounting for the dynamics of the temperature response. The results, shown in Figure 4.15, share two noteworthy features:

1. the transient responses to successive heat pulses differ;

2. after each stressful input is removed the biosensor output decreases to a new baseline that is elevated relative to the preceding unstressed baseline.

The data is presented as an adjusted fold change in order to render the results independent of meter calibration, but in this series of experiments there is no feature other than the basal biosensor output that can be expected *a priori* to be the same for plants exposed to the different profiles. For the presentation in Figure 4.15 the feature $\bar{y}$ is chosen to be the peak of the first transient. Even with this choice, though, the trend of longer heat stress exposure inducing larger subsequent baseline offsets is clear.

The persistent baseline offset is surprising, and poses an interpretive challenge. Conventional wisdom holds that exponential decay is a good description of protein degradation [1]. Indeed, the degradation of our marker fusion is very much exponential *relative to the adjusted baseline*. As an minimal example, consider Figure 4.16, which shows an exponential fit to the decay of the biosensor output after a single heat stress event, adjusted to the baseline offset. The fit is excellent, and the half-life of 2 hours is consistent with other studies in the literature for this YFP variant. To be slightly tongue-in-cheek, our YFP is exhibiting

expected decay *dynamics*, but very unexpected decay *statics*: the baseline offset persists far beyond the half- (or quarter- or eighth-) life of the YFP. We have conducted extended experiments where a plant was monitored at room temperature after the end of a heat stress event for four days, and the baseline offset did not change. These facts suggest that there is some mechanism actively keeping the H2B:YFP fusion active in the plant, and in so doing conferring on our biosensor a memory of the past heat stresses to which the plant has been exposed.

To better quantify the relationship between baseline adjustment and duration of past heat stress, we performed a separate series of experiments, this time with every plant exposed to a 2-hour "priming" heat stress, followed by 16 hours at room temperature and then a heat stress event varying from 2 hours to 8 hours. Five plants were exposed to each stress profile, and the results are shown in Figure 4.17. Here the choice of feature $\bar{y}$ to be the average steady-state value attained after recovery from the first pulse is meaningful, since every plant was exposed to the same 2-hour priming pulse.

The relationship between past heat stress exposure and baseline adjustment is significant in the sense that the mean adjusted fold changes are separated from each other by at least one standard error (impressively so in the case of the 2-hr/4-hr experiments, which thread the needle between the 2-hr/2-hr and 2-hr/6-hr experiments).

An observation that is consistent across all the experiments shown in Figures 4.15 and 4.17 pertains to the nature of the transient responses of the plants to the onset of initial and subsequent heat stress events. The initial rate of induction in response to the first heat stress event is much faster than in response to subsequent heat stress events, even though in both cases detectable induction occurs quickly.

4.6 Summary

The transgene that includes the transduction and signaling units of our biosensor is stable across multiple generations of biosensor host plants. Furthermore, the cost per data point of our biosensor system is very low relative to comparable quantification methods like PCR and optical microcopy. Because of this low cost and the fact that our biosensor measurement protocol is nondestructive and carried out *in situ*, we can run a large number of experiments, each generating a large volume of data. The large volume of data is possible because of a sampling rate that is fast enough to detect transients in a biosensor's response.

The high temporal resolution of our data enables a unique perspective on the biological variability among the biosensor host plants. Whereas many long-duration measurement techniques require significant interpolation between sparse data points (Figures 4.5 and 4.1 are good examples), our technique does not leave gaps large enough to alias any relevant features, so we can clearly see when, where, and how much the plants vary in terms of their biosensor output (Figure 4.12 and similar).

Rapid induction in response to both onsets of successive heat stress events indicates that this biosensor mechanism is resettable with respect to *changes* in the stress state of the

plant. This is as opposed to the steady state behavior, which is apparently not completely resettable after the stress has been removed.

We can conclude with certainty from our characterizations that our biosensor has a rapid response time and sufficient induction to be measured by a COTS fluorometer *in situ*. Furthermore, there is a good correlation between the steady-state biosensor output and the history of exposure to stress on time scales of a few days.

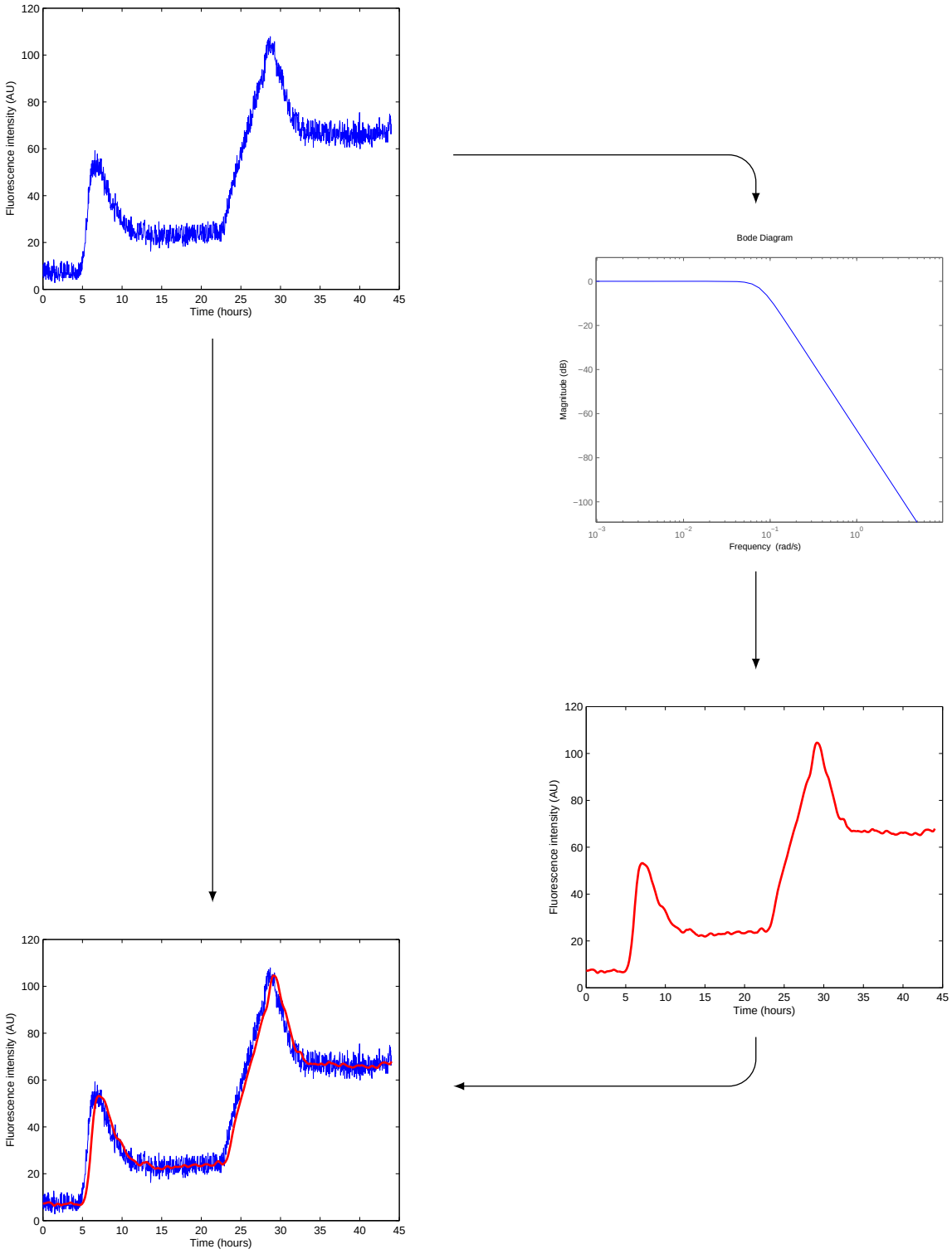


Figure 4.10: The sample period of the data is 120 seconds, corresponding to a sample frequency of $2\pi/120$ rad/sec. A rolloff filter with break frequency 0.075 rad/sec strikes a good balance between attenuating the noise and not introducing artificial measurement delay.

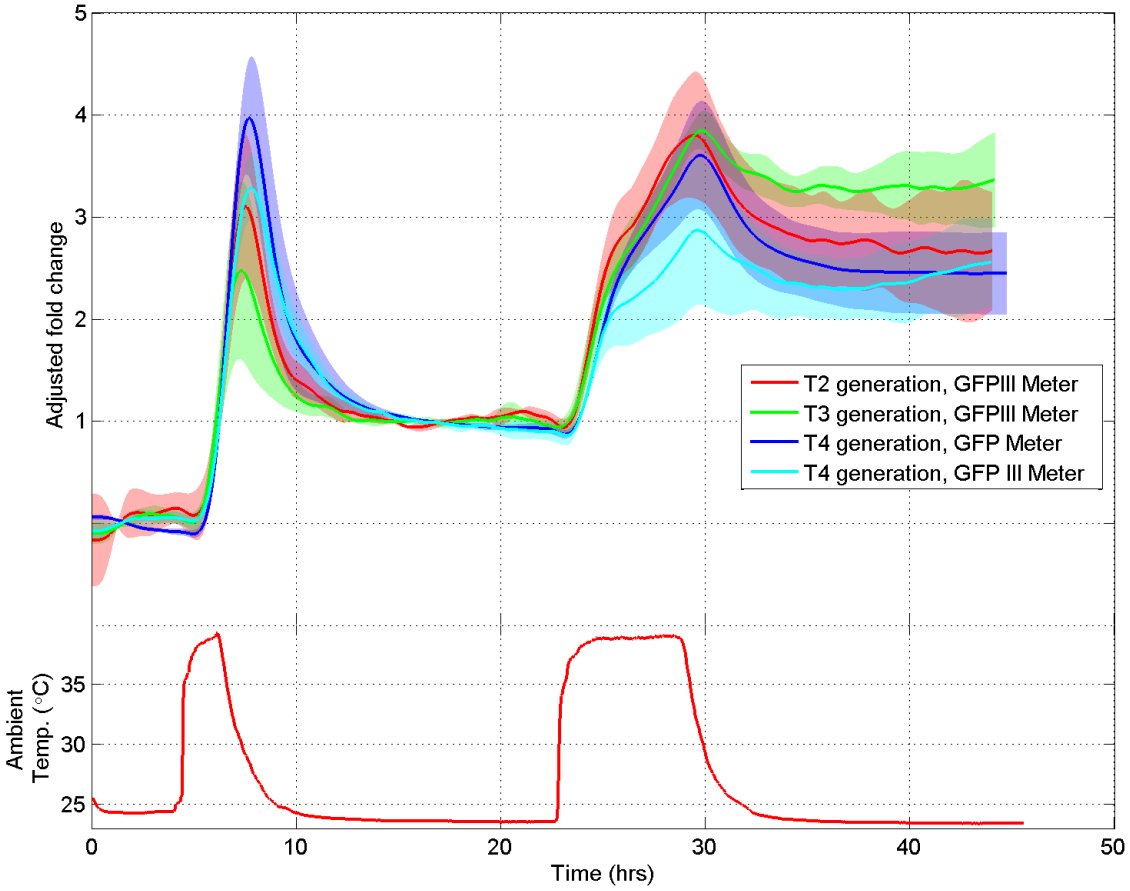


Figure 4.11: The transgene is stable with quantitatively similar expression .

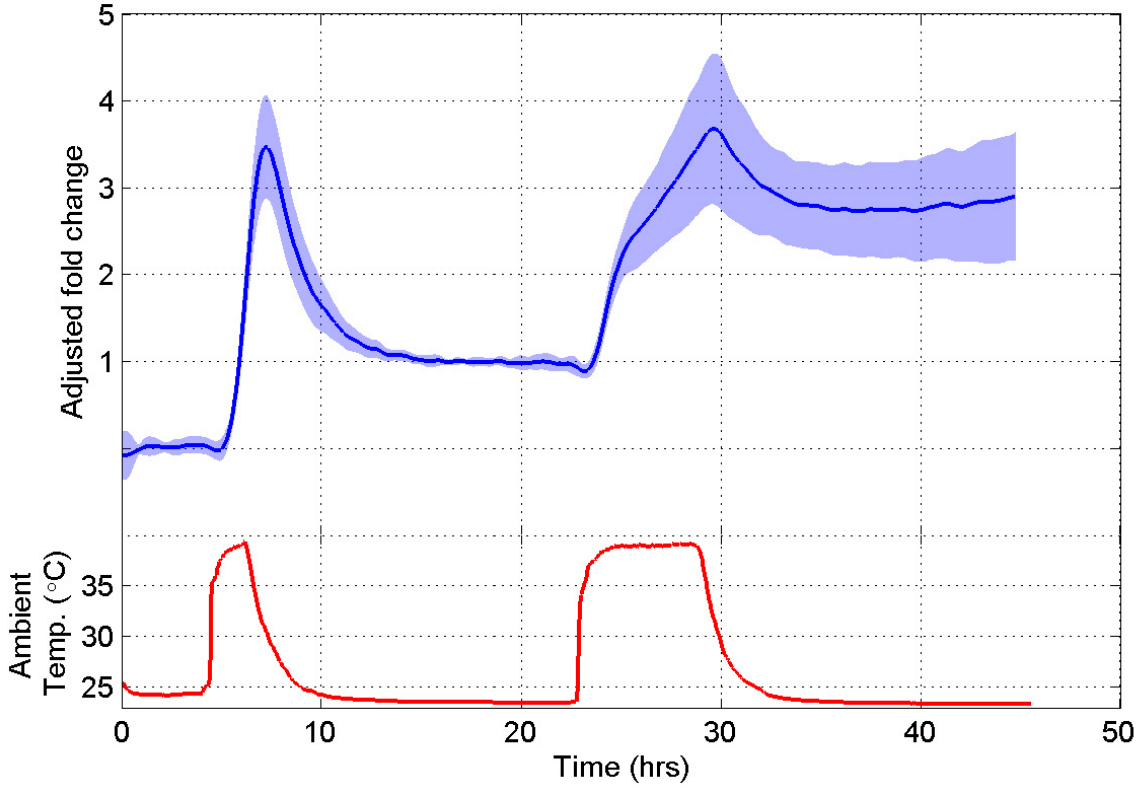


Figure 4.12: Average adjusted fold change of 20 biosensor hosts representing 3 generations of plants, all exposed to the same 2hr/6hr heat stress profile.

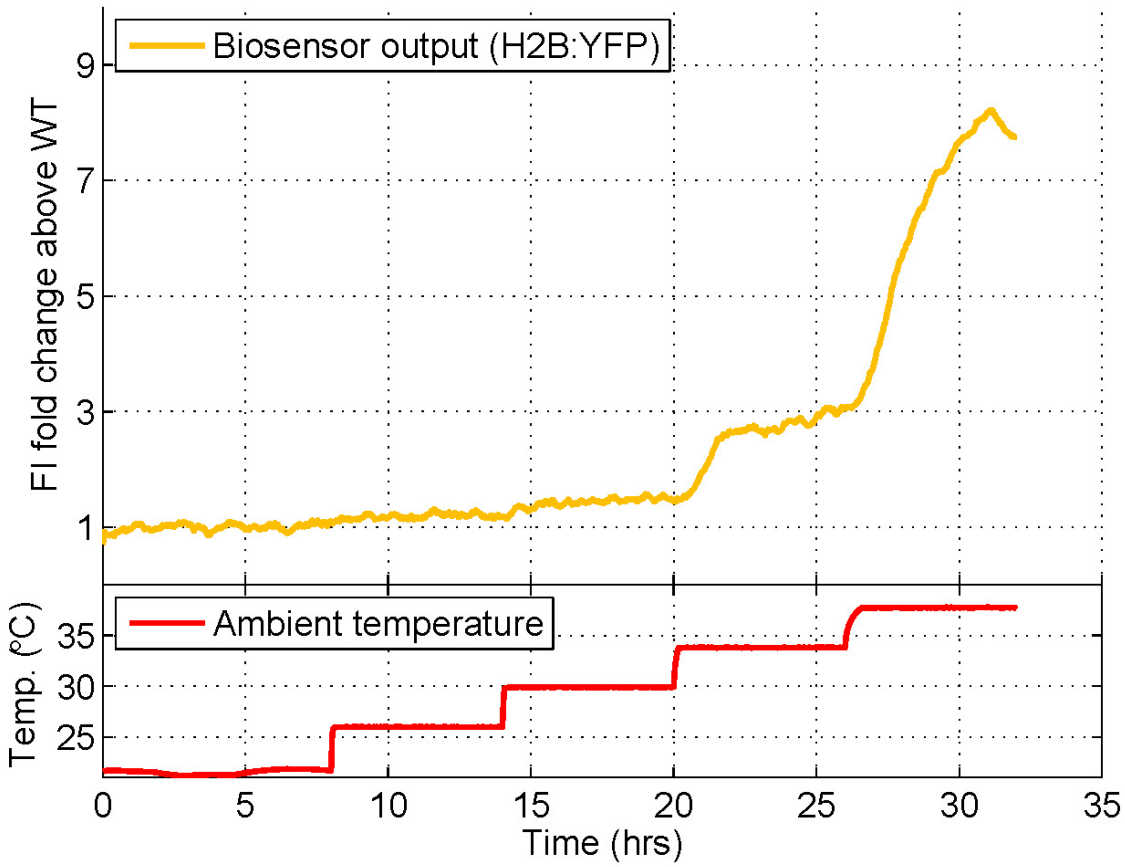


Figure 4.13: Biosensor response to a staircase heat stress profile with a step height of 4°C.

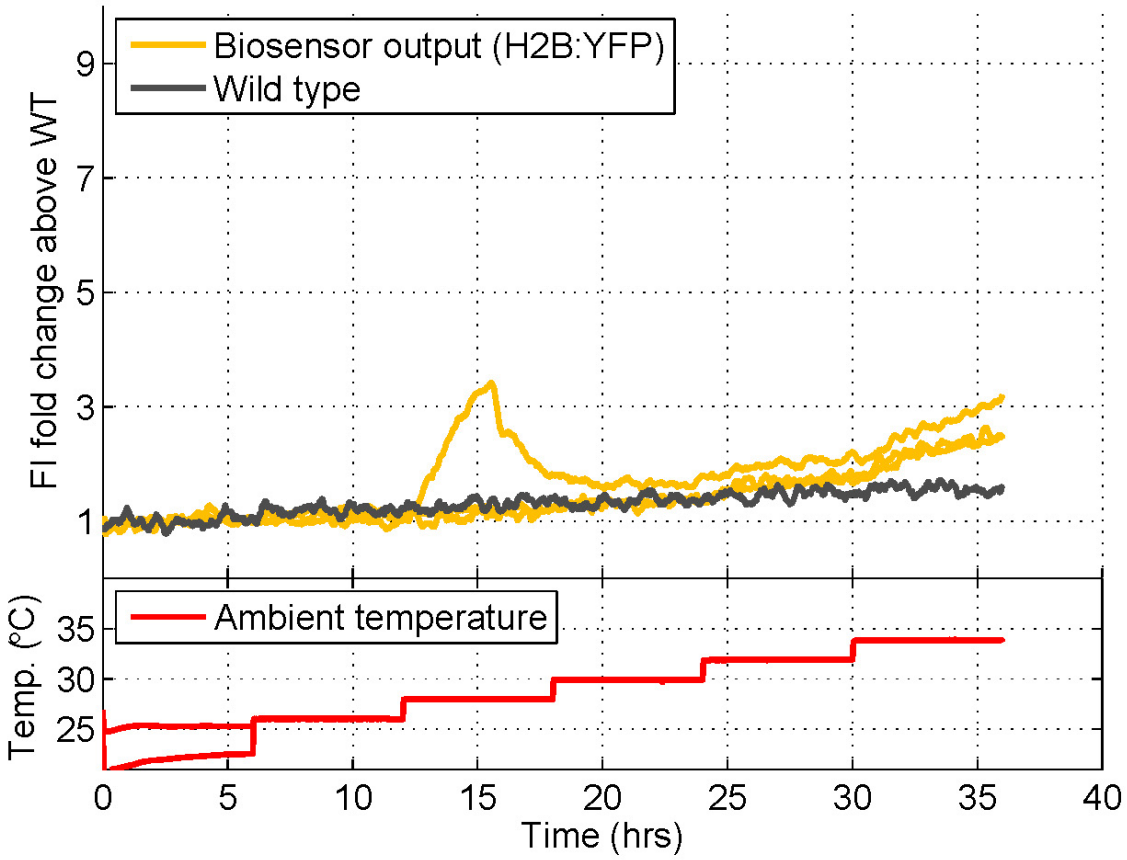


Figure 4.14: Staircase heat stress profile with 2°C step height. A wild type is shown for a clear view of when the biosensor signal is distinguishable. The y-axis scales are the same as in Figure 4.13 for easy comparison.

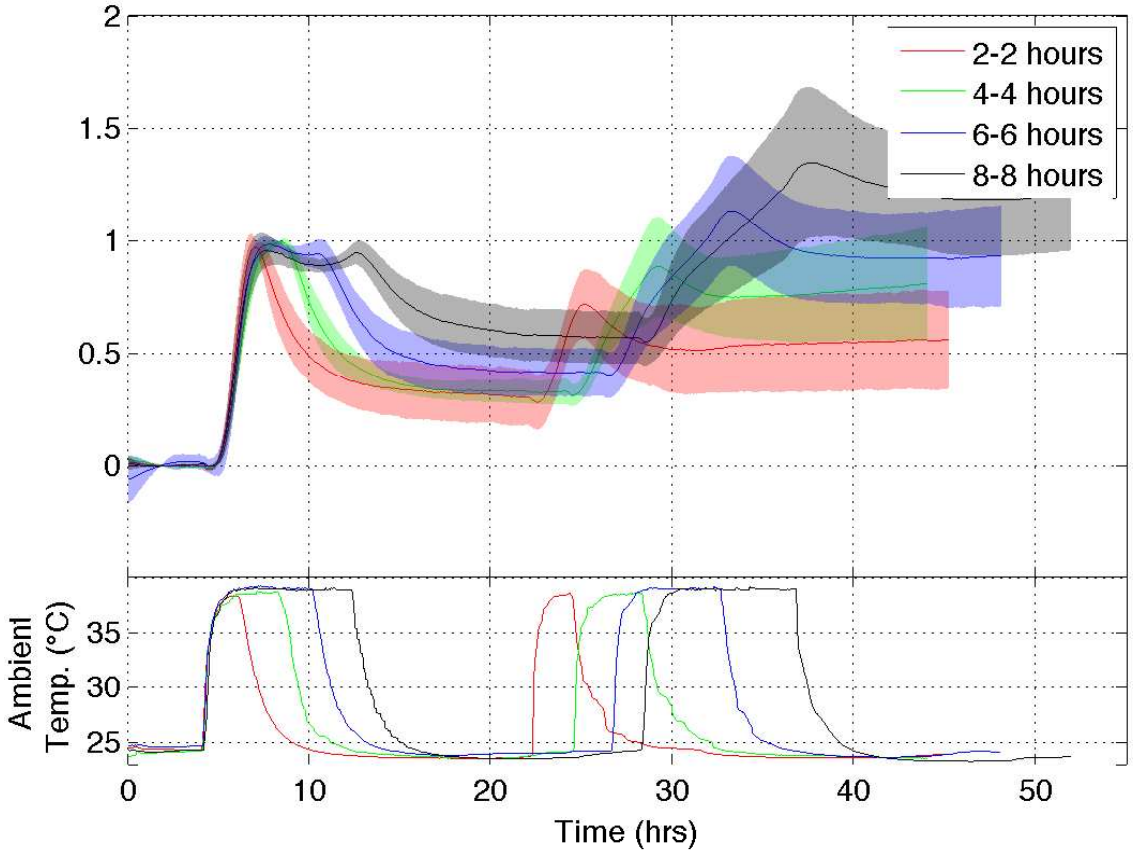


Figure 4.15: Pulse train heat stress profiles with equal-width heat pulses. Note that the fold change adjustment used here results in reduced standard error at the peak of the first transient, but much greater standard errors subsequently. This does not obscure the point of the figure, which is that the effects of the different pulse durations can be visibly distinguished by the corresponding elevations in steady-state biosensor output after removal of stress.

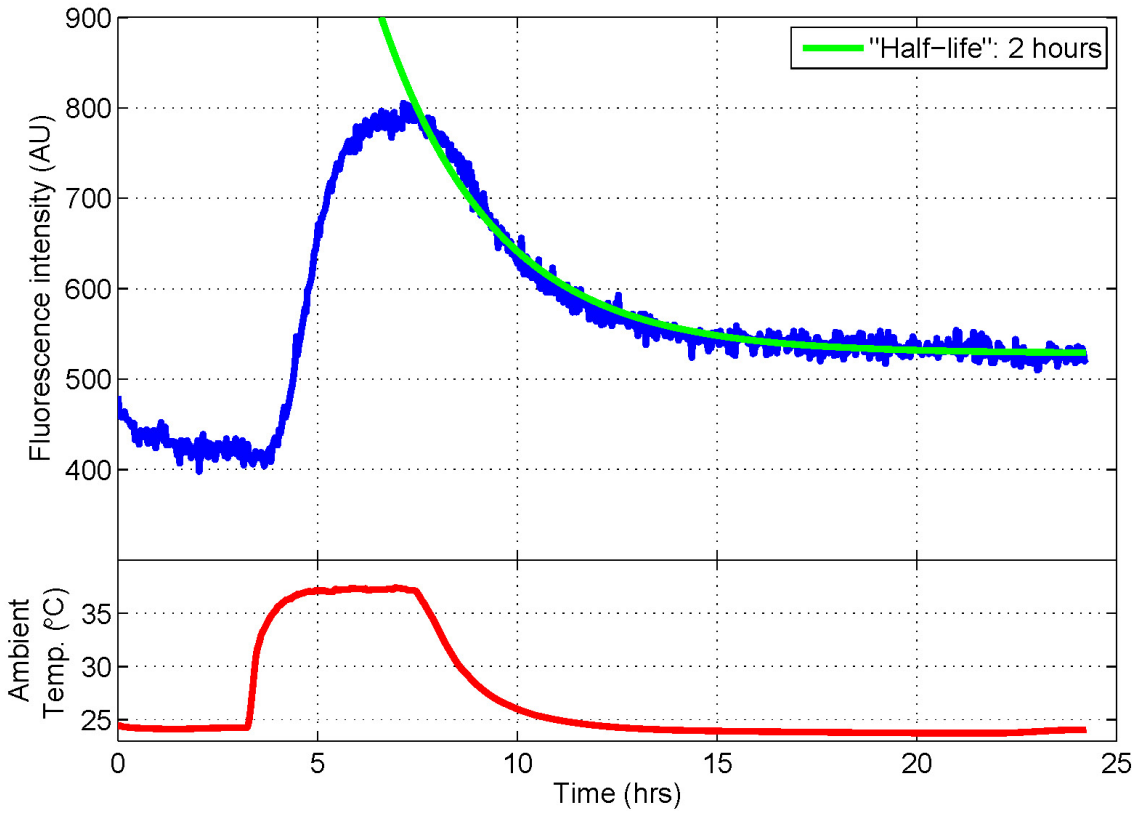


Figure 4.16: Underscoring the fact that the baseline elevation is not merely a slowly decaying transient, the decay of the biosensor output to that elevated baseline conforms to the standard exponential decay model of protein degradation.

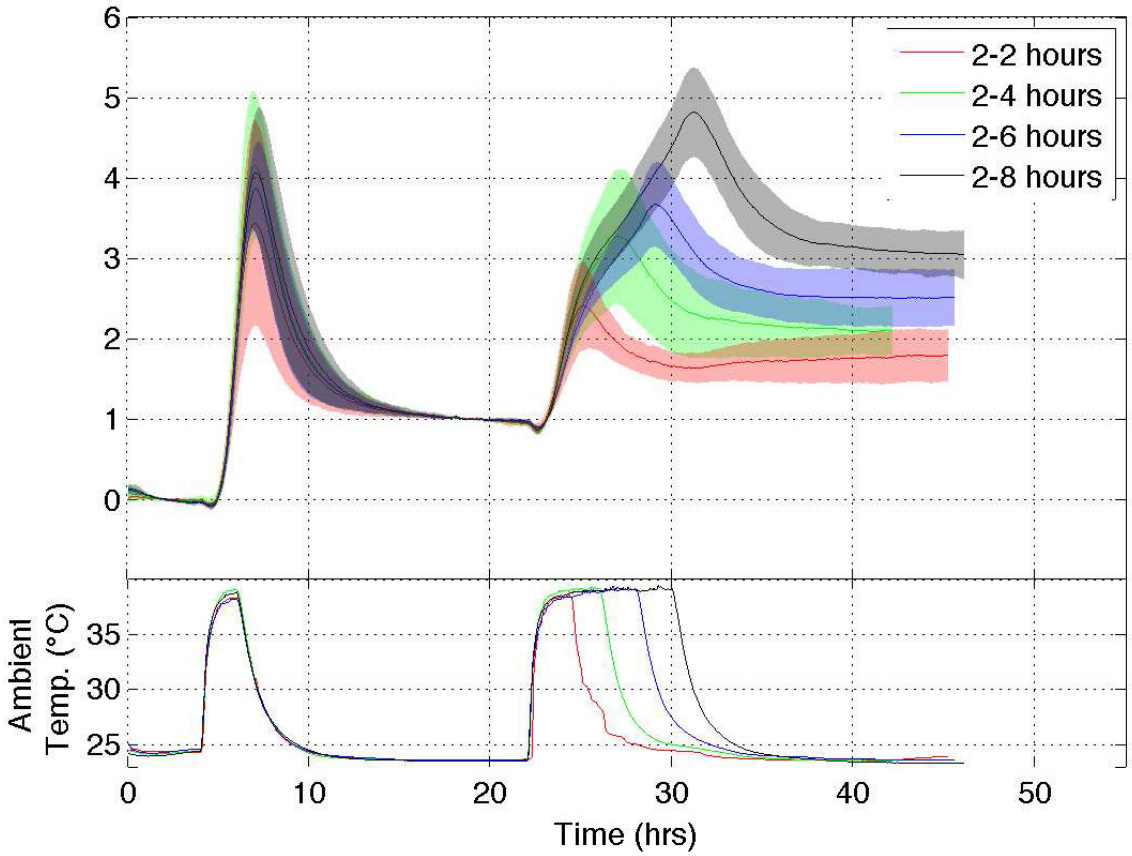


Figure 4.17: Now the small standard error is concentrated at the first post-stress steady state, and the post-2hr, -6hr, and -8hr steady states are significantly different.

Chapter 5

Characterization in Lieu of Modeling

With a thorough characterization of the biosensor's input/output behavior in hand, and with an eye toward the goal of closed-loop control of the biosensor output by means of temperature regulation, this chapter will briefly discuss the construction of a simple but functional input/output model of our biosensor system.

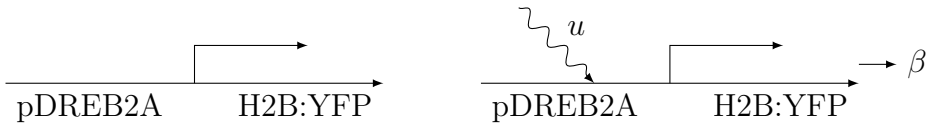
5.1 (In)applicability of mathematical models

There are standard mathematical models that describe the activity of gene networks. The simplest of them is the one-step activation/transcription/translation model, relevant to Figure 1.1. Denote by x the concentration of the protein coded by gene sequence G, and by a^* the concentration of act^*. The dynamics of x are described by

$$\dot{x} = -px + \beta \frac{a^*}{K + a^*}, \tag{5.1}$$

where $p > 0$ is a decay constant representing degradation of the protein, $\beta > 0$ is the maximum expression level, and $K > 0$ is called the activation coefficient [1]. This model is predicated on the Michaelis-Menten model of enzyme kinetics, and would be sufficient if our control input were a^*, but in reality we do not control a^* directly. Rather, as in Figure 5.1, we are changing the ambient temperature u, which we assume mediates a transition to activity of the inducer. Staying with our convention, continue calling the inducer a. When a is inactive it does not bind to pDREB2A, but in the presence of heat a undergoes a conformational change to its active form a^*, enabling binding with pDREB2A. Setting $a_T = a + a^*$ to be the total concentration of inducer (in both inactive and active forms) we arrive at the simple biosensor model

$$\dot{x} = -px + \beta \frac{f(a_T, u)}{K + f(a_T, u)}. \tag{5.2}$$



(a) Unstressed (b) Stressed

Figure 5.1: Because the mechanism mediating the path from heat stress to activation of pDREB2A is unknown, we simply conceive of the heat itself as the activator of pDREB2A.

It is worth nothing that if u were a *chemical* mediator in the $a \to a^*$ transition, the function $f(a_T, u)$ would be a Hill function. But because of the nature of our input u, Hill kinetics are not necessarily relevant.

What *is* obviously present in the input/output dynamics of our biosensor is integral-like behavior corresponding to the "memory" of the heat stress history, manifested in the quantitatively consistent baseline adjustment that increases as the total exposure to heat stress increases. This can be crudely modeled by adding a second state to (5.2) and defining the output fluorescence to be the first state:

$$\begin{aligned} \dot{x}_1 &= -p(x_1 - x_2) + \beta_1 \frac{f(a_T, u)}{K + f(a_T, u)} \\ \dot{x}_2 &= \beta_2 u \\ y &= x_1 \end{aligned} \tag{5.3}$$

With this modification, when the input u is removed ($u = 0$ in a the mathematical sense) the biosensor output (which is the state x_1) decays only as far as the current value of the state x_2, which will in turn increase whenever u is applied. To give a sense of how the model described by (5.3) behaves, Figure 5.2 shows the results of a cartoon simulation mimicking an 8-hr/8-hr experiment. Here u turns on and off by smoothly transitioning between 0 and 1, the output is initialized to 0, the parameters in (5.3) are

$$\begin{aligned} p &= 0.05 \\ \beta_1 &= 0.2 \\ K &= 5 \\ \beta_2 &= 0.002 \end{aligned}$$

and the activator transition function is defined to be $f(a_T, u) := u$. For reference, compare this cartoon to the results, shown in Figure 5.3, of a very early characterization experiment with a T1-generation plant. The main qualitative characteristics all match:

- The induction rate is high at the onset of heat stress and then slows;
- The decay after removal of heat stress is to an elevated baseline;

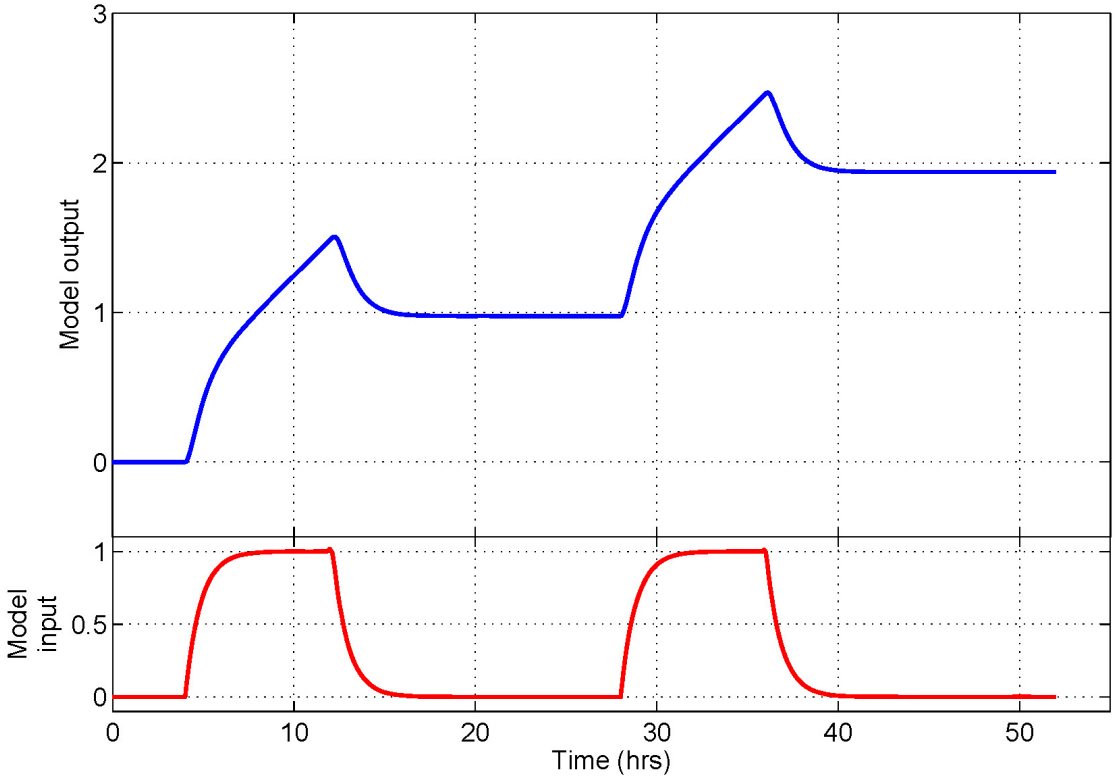


Figure 5.2: Simulation of a simple model designed to capture the qualitative features of the observed experiment results.

- The elevation of the baseline increases with the second heat stress event.

(Note that this particular plant shows extremely similar activation dynamics in response to both pulses, in contrast to all of the T4-generation plants used in the preceding characterization. In retrospect it is evident that the rapid reaction to the onset of the first stress event is a train that was selected in the propagated lines shown in Figure 4.6.)

The main points of this brief discussion are twofold. First, the biochemical processes that mediate the pathway from heat stress onset to induction of the biosensor are unknown and likely involve mechanisms like protein denaturation that do not have mathematical models that are both accurate and useful for control design. Second, the main qualitative features of the behavior observed during thorough characterization can be captured by a very simple model that is motivated by known mechanistic properties but does not directly reflect the underlying processes (a "gray-box" model).

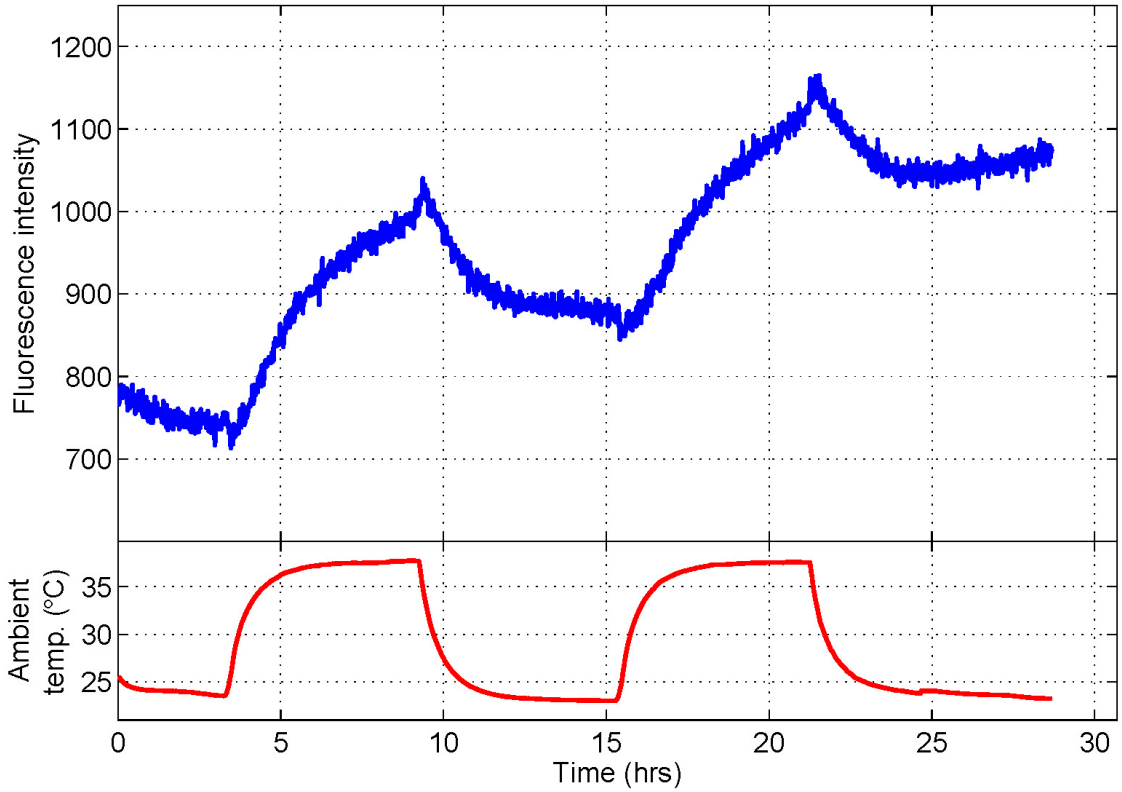


Figure 5.3: Early experiment data (raw fluorescence intensity) showing qualitative consistency with the behavior of the toy model, even with respect to the transient responses.

5.2 Interpretation

Speaking in control systems terms, the steady-state behavior of (5.3), which explicitly incorporates a model the rising baseline adjustment, resembles an open-loop integrator. This construction has three interesting interpretations:

1. The baseline cannot continue adjusting upward indefinitely; a physiological limit will be reached. In our plants, we see the biosensor output saturate after about 12-15 hours of continuous heat stress (it varies from plant to plant). The fact that open-loop integrators eventually saturate is inherent in real (as opposed to theoretically perfect) systems, and is intuitive.

2. It is a simple fact of classical control theory that "negative feedback around an integrator is stable." Because of the integral input-output behavior of (5.3), we expect that it could be stabilized by a very simple feedback control system. Since the correspondence

between (5.3) and the actual plant is good, we further expect to be able to achieve biosensor output regulation with a simple control system. This is as opposed to the (gratuitously?) complicated MPC scheme implemented in [17].

5.3 A qualitative model

We can remove ourselves even more completely from mathematical models by describing the biosensor's input/output map completely in terms of the qualitative characteristics of the characterization data:

P_1: At T_{high} the biosensor output is increasing;

P_2: At T_{low} the biosensor output is non-increasing;

P_3: Successive steady states are non-decreasing.

Taken together, these properties are a reference model $\mathbf{P}$ of the biosensor host plant ("Biosensor" in Figures 3.1 and 6.1). Note that there are transient discrepancies from this model, as in Figure 4.15 where the biosensor output decreases temporarily during the first high-temperature period for pulse durations of more than four hours. What is pleasing about this model is that it suggests a simple switching control strategy, and in the next chapter I will show that $\mathbf{P}$ has great success in enabling feedback control. This success provides confirmation for the assertion that a thorough characterization can be an adequate substitute for a mechanistic model for the purposes of feedback control. Furthermore, the simplicity of the system supports the intuition derived from the observation that the biosensor's input/output behavior resembles an integrator.

Chapter 6

Closing the Loop

Now the discussion shifts to use cases of a stress biosensor. These cases are classified by the actuation available with which to alter the growing environment. The "open-loop" case is one in which a grower has no control over the environment, for instance because the land was recently purchased and has not had nutrient-delivery systems installed. In this case the biosensor serves as a characterization tool, alerting the grower to the presence and severity of abiotic stresses. This information can be used to determine the crop that is best suited to the land, and it can be incorporated in the specification of actuation systems to be installed.

Here we will consider the other main class of applications: using a biosensor to enable closed-loop control of the growing environment, which assumes that actuation is available with with to effect changes in the environment factor that induces the biosensor. For practical closed-loop applications a model is needed to relate the biosensor output to the yield objectives of the crop. This model would function as a supervisory outer loop, determining desired stress states that an inner-loop controller can then maintain with environment actuation. Although an accurate supervisory model does not yet exist, we have implemented the inner-loop system, demonstrating closed-loop control of the biosensor's output using temperature as the actuated environment factor.

The system is implemented in our test greenhouse. A nonlinear regulator compares biosensor fluorescence with a reference (in a practical application this reference would be computed by the supervisory model) and computes temperature setpoints which are attained and maintained by the greenhouse's temperature control system. In the attainable range of temperatures, the greenhouse's temperature control system is well modeled by a first-order response, shown in Figure 6.1, emphasizing the fact that our actual objective is to regulate the output of the biosensor. The greenhouse is, in practical terms, an actuator.

6.1 Control strategy

Our task is to design the controller C_F so that the biosensor output F_{meas} tracks the reference F_{ref}. What exactly "track" means in this context depends on the structure of C_F. The

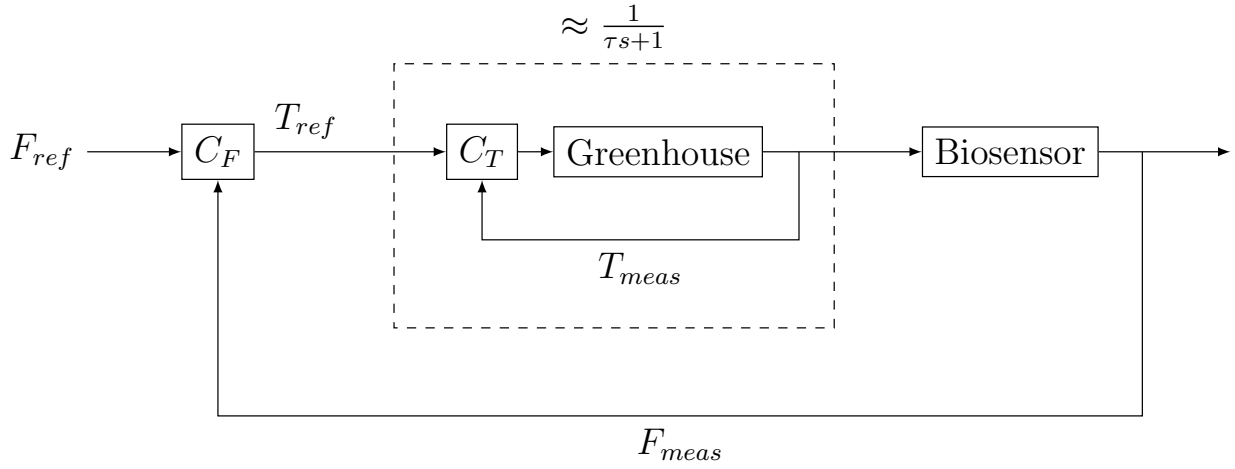


Figure 6.1: Stress setpoint regulation control loop.

simplicity of the qualitative model **P** proposed in the previous chapter suggests that C_F can have a similarly simple strategy. The simplest of those strategies is the hysteresis relay shown in Figure 6.2. As such, the system C_F has inputs F_{ref} and F_{meas}, output T_{ref}, and is parameterized by F_{th}, T_{low}, and T_{high}. We define "good tracking" as when the value of F_{meas} is persistently in the range $[F_{ref} - F_{th}, F_{ref} + F_{th}]$.

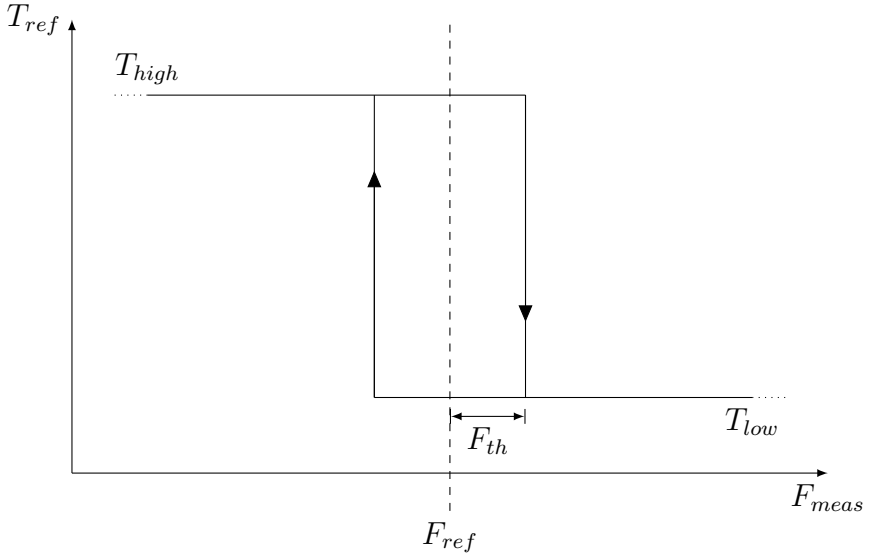


Figure 6.2: A graphical representation of the hysteresis relay used to set T_{ref} in the experiments shown in Figures 6.4-6.8. T_{low}, T_{high}, F_{ref}, and F_{th} are parameters chosen at the start of the experiment. F_{ref} and F_{th} and are defined as fold changes above baseline fluorescence.

To test this control strategy in simulation before implementing it in the greenhouse we

do have to describe **P** with differential equations. We described a candidate model in (5.3). To make this example clearer to look at we choose the slightly different parameter set

$$\begin{aligned} p &= 0.04 \\ \beta_1 &= 0.2 \\ K &= 5 \\ \beta_2 &= 0.002 \end{aligned}$$

and represent the temperature dynamics of the greenhouse by the linear system

$$\frac{1}{4s+1}.$$

These differential equations cannot capture the transient differences we observe between the biosensor's response to initial and subsequent heat stress events. The real system has a much more rapid response to the first heat stress event than to subsequent events.

Simulating this model closed-loop system gives a qualitative idea of how we can expect our actual biosensor to behave in feedback with a hysteresis relay. The results, shown in Figure 6.3, are encouraging. The main observation is that the closed-loop temperature reference should be a pulse train, and eventually the accumulated exposure to T_{high} will increase the baseline adjustment until finally the biosensor output will reach a steady state inside the hysteresis band. Furthermore, this qualitative model predicts that the spacing between successive pulses will increase as the plant adjusts closer to its final steady state.

6.2 Experiment conditions

The greenhouse is located inside a temperature controlled building with large skylights letting in abundant natural light. The ambient temperature in the building fluctuates diurnally between about 20°C and 28°C depending on the season. The upper extent of these fluctuations is not sufficient to induce measurable biosensor output, but because the greenhouse control system to cooling with forced convection of ambient air, the diurnal temperature fluctuations do set a lower bound on the achievable temperature inside the greenhouse.

6.3 Results

In practice the controller input F_{ref} and parameter F_{th} are actually defined in terms of scalar multiples of the background fluorescence F_b:

$$F_{ref} = s_r F_b, \quad F_{th} = s_t F_b$$

The experiments in the greenhouse with the biosensor host in the loop proceed in three phases:

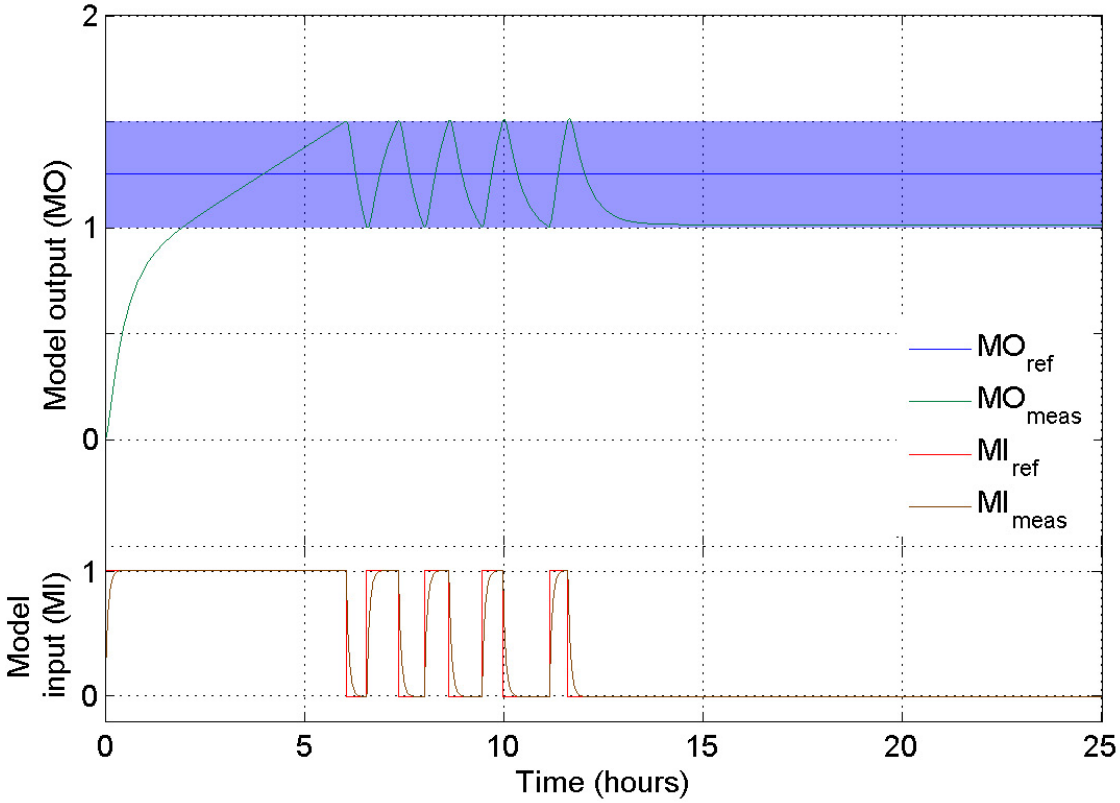


Figure 6.3: A realization of **P** with ordinary differential equations as in (5.3), in feedback with a hysteresis relay.

1. Scalars s_r, s_t are specified
2. The fluorescence is measured for four hours at room temperature to determine F_b
3. Closed loop control is initiated with $F_{ref} = s_r F_b$ and $F_{th} = s_t F_b$

The results of an illustrative closed-loop experiment are shown in Figure 6.4. I point out first that the discrepancy between the initial response of the actual biosensor host and our model is immediately apparent in the very short first pulse, after which the response rate slows, leading to a longer second pulse. After recovering from that first rapid induction, the behavior of the real biosensor host in closed loop is strikingly similar to that of the qualitative model. In particular, note the decreasing frequency of the pulses before the baseline is adjusted all the way into the hysteresis band.

Another experiment, using most of the same controller parameters, is shown in Figure 6.5. The plant used in this experiment has a considerably different initial transient peak

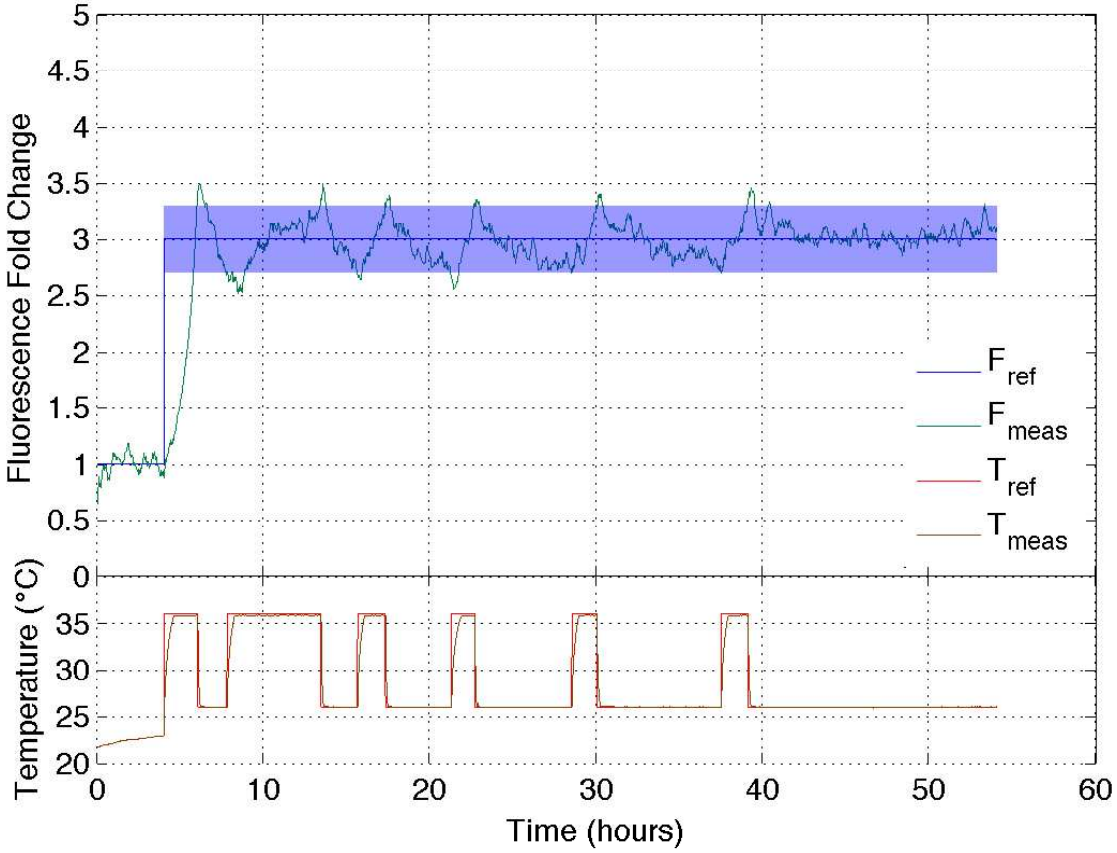


Figure 6.4: $F_{ref} = 3F_b$, $F_{th} = 0.3F_b$, $T_{low} = 26°C$, $T_{high} = 36°C$.

that does not exceed the upper boundary of the hysteresis band, so the closed loop control signal remains high. A consequence of the baseline adjustment accumulation during the long initial heat pulse is the small number of followup pulses that are required to bring the steady state to rest within the hysteresis band.

Next, in Figure 6.6, we see that when the hysteresis band F_{th} is increased there are correspondingly fewer switches in T_{ref}. This plant also has a relatively low transient peak.

Figure 6.7 reflects the same controller parameters as in Figure 6.6, but this plant has a higher transient peak so we see a brief interruption in the high-temperature pulse after the initial spike.

Finally, we demonstrate in Figure 6.8 that the hysteresis relay can track a time-varying reference. In this experiment s_r was increased from 3 to 4 after 20 hours. The hysteresis band was purposely reduced ($s_t = 0.2$) to highlight the difference between the reference bands.

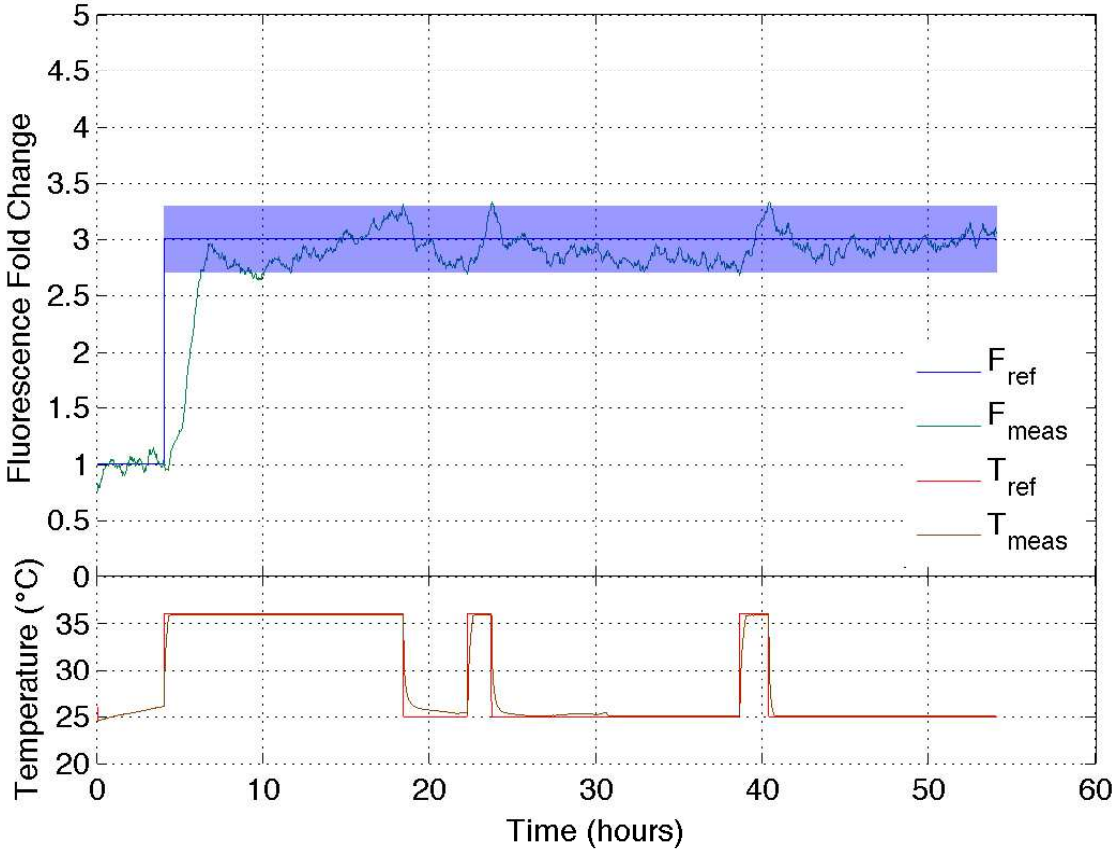


Figure 6.5: $F_{ref} = 3F_b$, $F_{th} = 0.3F_b$, $T_{low} = 25°C$, $T_{high} = 36°C$.

Observe through all these experiments that this very simple control strategy based on a qualitative model of the dynamics is successful in tracking constant and time-varying references, and is able to do so for a system that is uncertain (almost by definition of the model $\mathbf{P}$), but in particular that varies greatly in the qualitative nature of the response to the first onset of heat stress.

Finally, I reiterate that the biosensor host plants in these experiments are living, mature, potted specimens of *Arabidopsis*, with the biosensor measurement taken *in situ* every two minutes and fed back continuously in real time to the greenhouse climate controller.

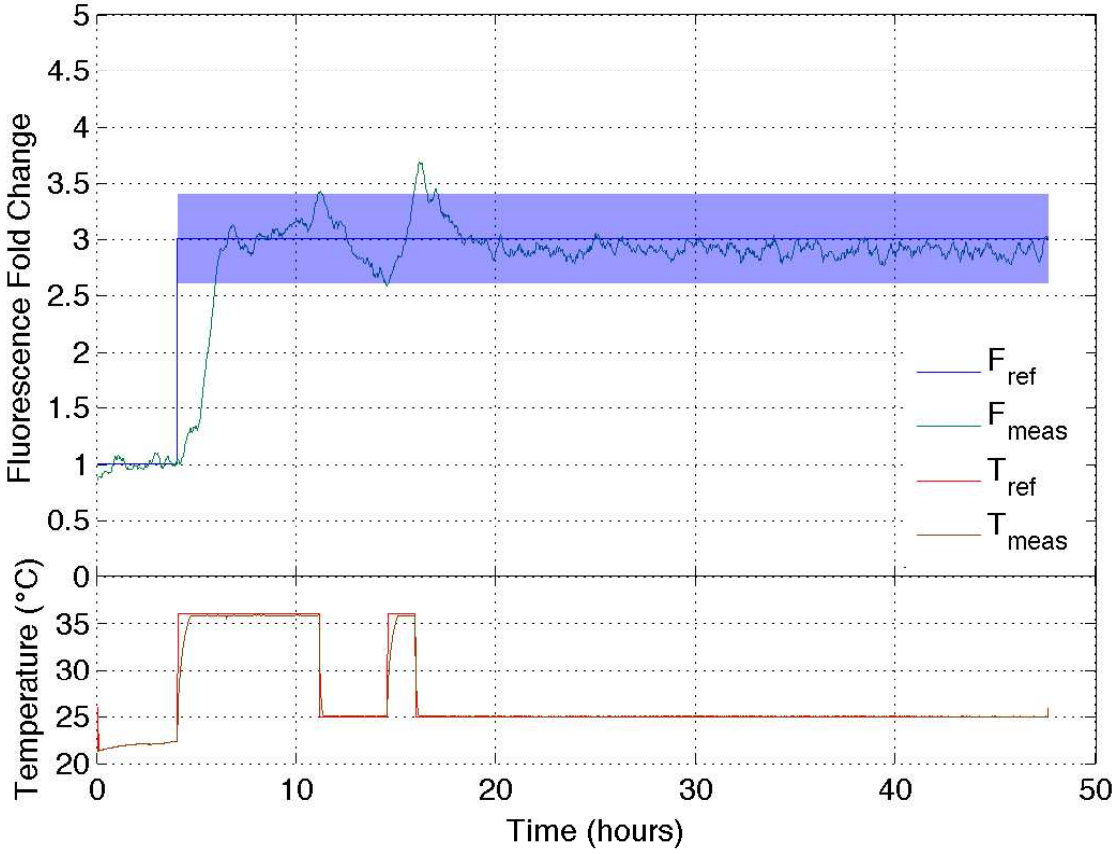


Figure 6.6: $F_{ref} = 3F_b$, $F_{th} = 0.4F_b$, $T_{low} = 25°C$, $T_{high} = 36°C$.

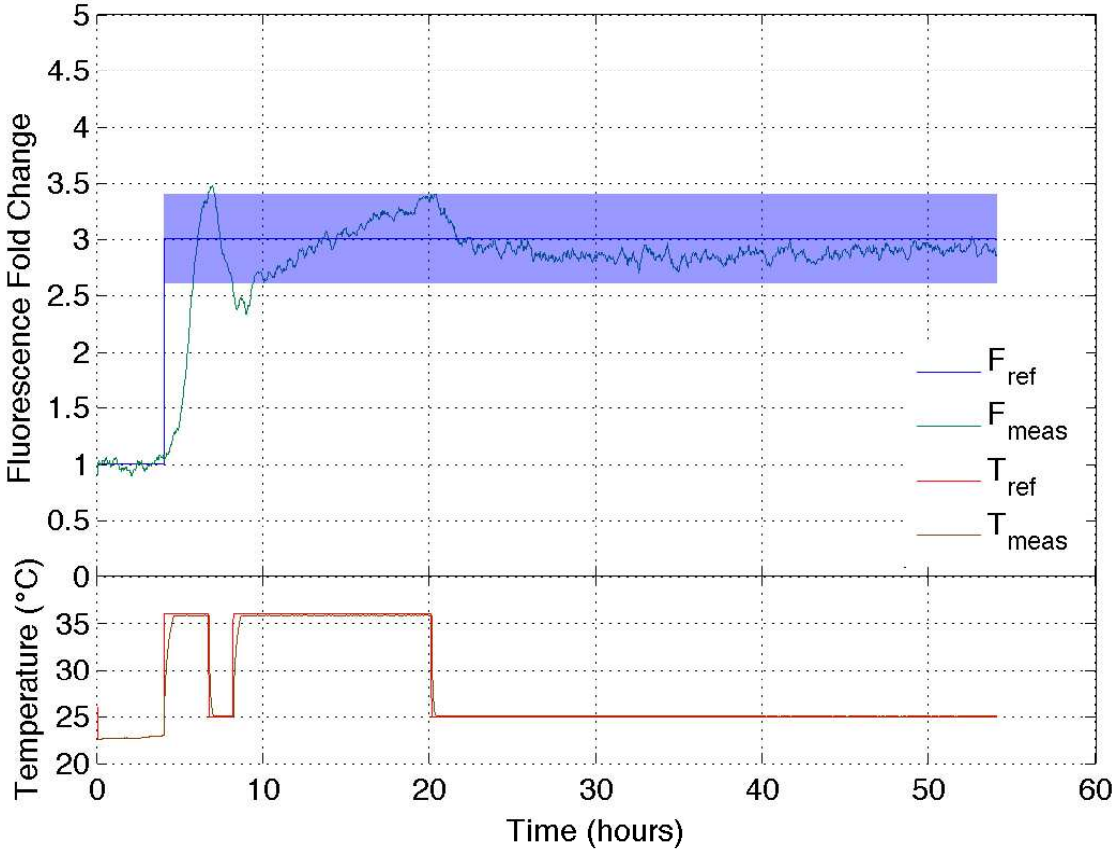


Figure 6.7: $F_{ref} = 3F_b$, $F_{th} = 0.4F_b$, $T_{low} = 25°C$, $T_{high} = 36°C$.

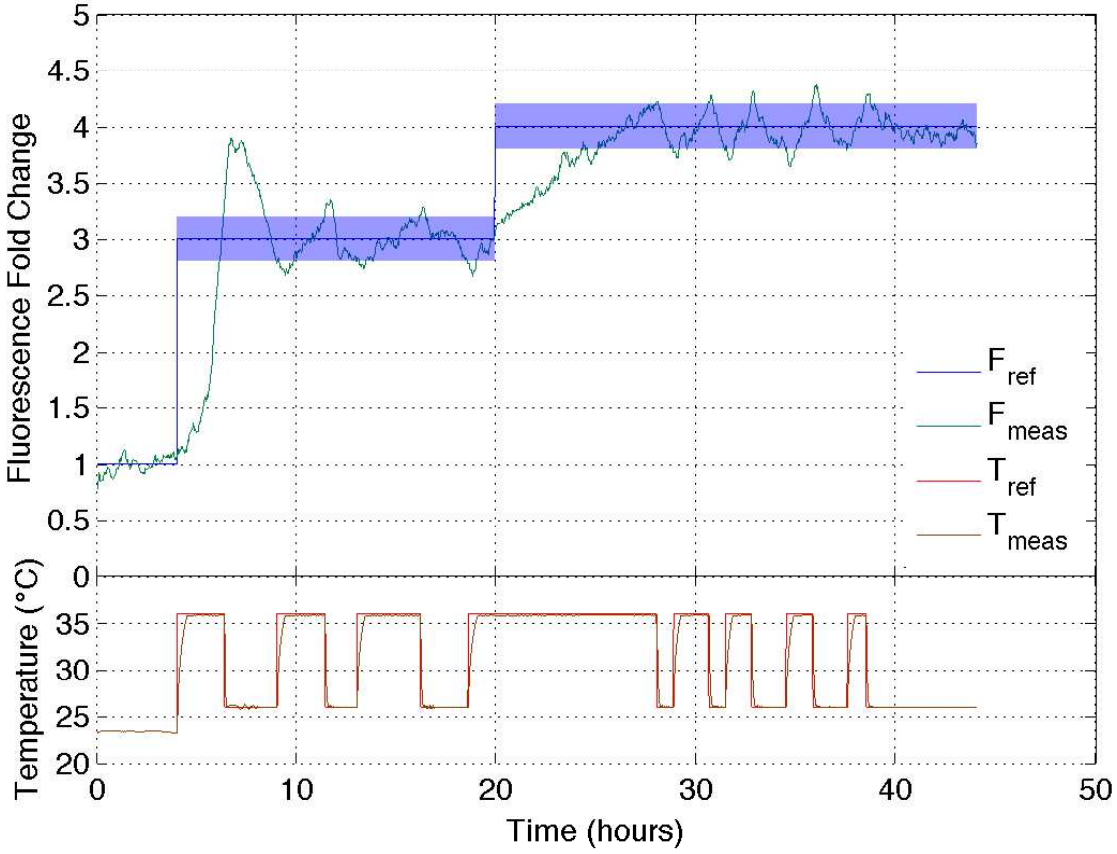


Figure 6.8: $F_{ref} = \{3b, 4b\}$, $F_{th} = 0.2b$, $T_{low} = 26°C$, $T_{high} = 36°C$.

Chapter 7

Summary, Future Work, Conclusions

Closed-loop setpoint regulation of the output of a stress-responsive biosensor is possible, with a real-time *in situ* sensing mechanism that has a very low marginal cost of data. In the process of showing this we have raised many questions that are individually interesting. In this closing chapter I will describe a few of these open doors.

7.1 Memory of stress

The characterization certainly shows a behavior similar to memory in the continued upward adjustment of baseline fluorescence attained after cessation of successive heat stress events. We do not know what mechanism is conferring this behavior, whether it is naturally a characteristic of the DREB2A pathway, or whether it is entirely a figment of the biosensor system. Either explanation is an interesting topic of future study. If the memory is a characteristic of the pathway, it would suggest that DREB2A's mechanistic role in heat stress response deserves more attention that is currently paid to it in the literature. If the memory is a figment of the biosensor, it likely affords a quantitative view of the role of the protein Histone2B in preventing denaturation of DNA.

7.2 Biosensor development

The success of our particular heat-stress responsive biosensor system motivates the development of biosensor systems that response to a variety of environment factors with a variety of dynamic characteristics. Specifically, if focus remains on heat-stress responsive biosensors, it might be possible to develop a sensor without the memory behavior. This would ideally quantify the instantaneous stress state, resettable to zero upon cessation of heat stress events.

Another development focus could be on speeding up the biosensor induction after onset of stress. Because our (genetically encoded) biosensor is only "1 transcription long," the time from activation of pDREB2A to biosensor induction can't get any shorter. But there

are other processes upstream of the activation of pDREB2A that it might be possible to accelerate or bypass with better knowledge of their mechanisms.

Other stresses

We chose to use heat stress for the work so far because it is easy to modulate in lab settings. However in agricultural practice the environment temperature is only controllable by the grower in actuated greenhouses. An obvious next step is to measure the activity of pathways that respond to different types of stresses that are controllable even in field settings. Water stress is a particularly good example, especially respecting the prevalence of drip irrigation, which allows very precise control of the flow rate, sometime even to the level of individual crop rows.

We propose the cultivation of transgenic biosensor plants that are responsive to many environmental stresses, trusting that companion research in plant stress physiology will uncover biochemical processes with particularly strong connections to yield objectives. Together, these developments will lead to biosensors that quantify those processes and thereby the link between stress measurement and economic yield.

In contrast to our very simple biosensor with one strong inducer stress, it may be beneficial to report the activity of a pathway that is induced by multiple stresses. Assessing the observability of such a system is a research opportunity for biologically minded control theorists. As we have observed and demonstrated, the biological variability and nonlinearity of even a simple stress response pathway cause qualitative differences in open- and closed-loop behavior. We are working to improve our understanding of these differences with existing feature-based uncertainty quantification strategies. A concrete understanding of the biological variability has important consequences with respect to field implementation of biosensors.

Tunable gain

Our work has been limited to copies of endogenous promoters that directly induce transcription of the biosensor FP. The "gain" of this biosensor is related to the transcription rate, which is specific to the promoter. Consequently the biosensor's gain is fixed by the choice of promoter. It is easy to imagine an application that demands measuring the activity of a weak promoter, which may not produce a sufficient SNR. One way of overcoming this limitation could be with a *transcription activator-like effector*, or TALE. In the form presented in CITATION the TALE is incorporated in positive feedback, but this system of theoretical infinite gain can interfere with normal cellular functions. We are currently testing an *open loop* TALE. This is schematically represented in Figure 7.1. The endogenous promoter activates transcription of a gene encoding the synthetic activator *TALEe* which is designed to bind the matching synthetic promoter *pTALE*, activating transcription of the signaling unit *FP*. The strength of the promoter *pG* cannot be tuned, but the strength of the promoter *pTALE* can, for example by including multiple binding sites for several copies of *TALEe*.

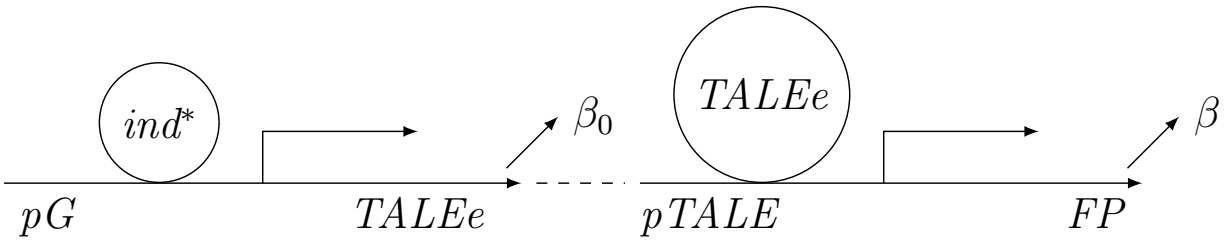


Figure 7.1: Open-loop TALE.

7.3 Proposed use cases

In the first use case we envision, biosensor host plants would be distributed throughout a cultivated field and their stress state fed back constantly to a central monitoring station. In this case the biological variability of the sensor hosts would determine the appropriate number of biosensors to deploy so that the aggregate biosensor stress state accurately represents that of the entire crop. With that confidence, the stress state would be used to inform and advise the grower about present conditions and appropriate interventions to avert crop-loss events. Assuming the grower is vested with oversight of resource allocation, investment in such a system carries little risk but high potential reward, especially in the case of flavor crops where a sustained stress level at particular phases of the growth cycle is advantageous.

Alternatively, as growers are branching out to marginal land, a palette of transgenic biosensors could be used as an agricultural "canary in the coal mine" to determine the stress profile of a tract of land before planting a crop.

Bibliography

[1] U. Alon. *An introduction to systems biology: design principles of biological circuits*. Chapman & Hall/CRC, 2007.

[2] R.M. Amasino. "Vernalization and flowering time". In: *Curr. Opinion in Biotech.* 16 (2 2005), pp. 154–158.

[3] M. S. Antunes et al. "Programmable ligand detection system in plants through a synthetic signal transduction pathway". In: *PLoS One* 6 (1 2011).

[4] S.J. Clough and A.F. Bent. "Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*". In: *The Plant Journal* 16.6 (1998), pp. 735–743.

[5] J. Cockram et al. "Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity". In: *Journal of Experimental Botany* 58.6 (2007), pp. 1231–1244.

[6] M. Fosbrink et al. "Visualization of JNK activity dynamics with a genetically encoded fluorescent biosensor". In: *PNAS* 107.12 (2010), pp. 5459–5464.

[7] I. Golding and E.C. Cox. "Spatiotemporal dynamics in bacterial cells: real-time studies with single-event resolution". In: *Methods in Cell Biology* 89 (2008), pp. 223–251.

[8] J. Hua. "From freezing to scorching, transcriptional responses to temperature variations in plants". In: *Curr. Opinion in Plant Bio.* 12 (2009), pp. 568–573.

[9] D.I. Jackson and P.B. Lombard. "Environmental and management practices affecting grape composition and wine quality - a review". In: *Am. J. Enol. Vitic.* 44.4 (1993), pp. 409–430.

[10] K. Jiang et al. "Expression and characterization of a redox-sensing green fluorescent protein (reduction-oxidation-sensitive green fluorescent protein) in Arabidopsis". In: *Plant Physiology* 141.2 (2006), pp. 397–403.

[11] B.N. Kholodenko. "Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades". In: *Eur. J. Biochem.* 267 (2000), pp. 1583–1588.

[12] D.A. Laurie. "Comparative genetics of flowering time". In: *Plant Molecular Biology* 35 (1997), pp. 167–177.

[13] O. W. Liew et al. “Signature optical cues: emerging technologies for monitoring plant health”. In: *Sensors* 8 (2008), pp. 3205–3239.

[14] S. Lindquist and E.A. Craig. “The heat shock proteins”. In: *Annu. Rev. Genet.* 22 (1988), pp. 631–677.

[15] M.M. Mahfouz et al. “De novo-engineered transcription activator-like effector (TALE) hybrid neclease with novel DNA binding specificity creates double-strang breaks”. In: *PNAS Early Edition* (2010).

[16] J. Martin, A.L. Horwich, and F.U. Hartl. “Prevention of protein denaturation under heat stress by the chaperonin Hsp60”. In: *Science* 258.5084 (1992), pp. 995–998.

[17] A. Milias-Argeitis et al. “In silico feedback for in vivo regulation of a gene expression circuit”. In: *Nature Biotech.* 29 (2011), pp. 1114–1116.

[18] G. Miller and R. Mittler. “Could heat shock transcription factors function as hydrogen peroxide sensors in plants?” In: *Annals of Botany* 98 (2006), pp. 279–288.

[19] R.J. Millwood et al. “Instrumentation and methodology for quantifying GFP Fluorescence in Intact Plant Organs”. In: *BioTechniques* 34.3 (2003), pp. 638–643.

[20] R. Mittler. “Abiotic stress, the field environment and stress combination”. In: *Trends in Plant Sci.* 11.1 (), pp. 15–19.

[21] F.J. Molz. “Models of water transport in the soil-plant system: a review”. In: *Water Resources Research* 17.5 (1981), pp. 1245–1260.

[22] M.C. Morris. “Fluorescent biosensors of intracellular targets from genetically encoded reporters to modular polypeptide probes”. In: *Cell Biochem. Biophys.* 56 (2010), pp. 19–37.

[23] Opti-Sciences Inc. *GFP-Meter*. URL: `http://www.optisci.com/datasheet/gfp_meter.pdf`.

[24] A. Pellegrino et al. “Towards a simple indicator of water stress in grapevine (*Vitis vinifera* L.) based on the differential sensitivities of vegetable growth components”. In: *Australian J. Grape and Wine Research* 11 (2005), pp. 306–315.

[25] F. Qin et al. “Arabidopsis DREB2A-interacting proteins function as ring E3 ligases and negatively regulate plant drought stress-responsive gene expression”. In: *The Plant Cell* 20 (2008), pp. 1693–1707.

[26] L. Rizhsky et al. “When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress”. In: *Plant Physiology* 134 (2004), pp. 1683–1696.

[27] M. Ronen et al. “Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics”. In: *PNAS* 99.16 (2002), pp. 10555–10560.

[28] A. Sadanandom and R.M. Napier. "Biosensors in plants". In: *Curr. Opinion in Plant Bio.* 13 (2010), pp. 736–743.

[29] Y. Sakuma et al. "Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression". In: *PNAS* 103.49 (2006), pp. 18822–18827.

[30] Y. Sakuma et al. "Functional analysis of an arabidopsis transcription factor DREB2A, involved in drought-responsive gene expression". In: *The Plant Cell* 18 (2006), pp. 1292–1309.

[31] P.F. Scholander et al. "Sap pressure in vascular plants". In: *Science* 148.3668 (1965), pp. 339–346.

[32] A. Schwartz et al. "Formalization of the MESF unit of fluorescence intensity". In: *Cytometry B (Clinical Cytometry)* 57B (2004), pp. 1–6.

[33] G. van Straten, H. Challa, and F. Buwalda. "Towards user accepted optimal control of greenhouse climate". In: *Computers and Electronics in Agriculture* 20 (2000), pp. 221–238.

[34] R.Y. Tsien. "The green fluorescent protein". In: *Annu. Rev. Biochem.* 67 (1998), pp. 509–544.

[35] T. Yoshida et al. "Functional analysis of an arabidopsis heat-shock transcription factor HsfA3 in the transcriptional cascade downstream of the DREB2A stress-regulatory system". In: *Biochemical and Biophysical Research Communications* 368 (2008), pp. 515–521.