The Evolution of Color and Form in the Species Hemerocalis: A Study in Controlled Gene Expression

Terrence P. McGarty, Ph.D. [1]

Abstract

This paper presents an overview of the development of color in the genus Hemerocalis from the perspective of the underlying genetic mechanisms. This genus has several well known and distributed species and has been hybridized extensively over the past one hundred years. In this paper we review the genus and the hybrids and then we provide a review of the current state of research in color expression in flowers such as this genus. The review of the current research summarizes many of the most recent efforts in gene control and expression and the resulting enzyme controls in the anthocyanin pathways. We focus on the anthocyanidin pathways but also discuss recent work in the carotenoid and flavonol pathways as well. The questions still outstanding is; how do all of these pieces come together to create the hybrids as we now know them and a corollary which is; can we use what is known to develop more sophisticated hybrids.

 Dr. McGarty is Head of Plant Research at The Telmarc Gardens, a subsidiary of The Telmarc Group, LLC.

Contents

1 Introduction 4
1.1 Overview 4
1.2 Why This Genus 6
1.3 Prior Work 7
1.3.1 Genetic Structure of Hemerocalis 7
1.3.2 Gene Expression and Pathway Control 8
1.3.3 Modeling of Gene Expression: Analysis and Synthesis 8
2 HEMEROCALIS Species and Color 9
2.1 The Species 9
2.2 The Early and Latter Hybrids 16
3 Classic Genetics 19
3.1 Mendellian Analysis 19
3.2 Breeding and Hybridizing 23
3.3 Summary of Mendellian Approach 24
4 Flower Color Expression 25
4.1 Pathways and Enzymes 25
4.2 Anthocyanins 26
4.3 Other Color Elements 30
4.4 Carotenoids 31
4.5 Flavones 32
4.6 Pathways 33
4.6.1 Anthocyanin Pathway 34
4.6.2 Carotenoid Pathway 38
4.6.3 Flavonol Pathway 39
5 Genetic Influence 39
5.1 Plant Cells 40
5.2 Plant DNA 41
5.3 Plant Gene Processes 43
5.4 Activators 45
5.5 Repressors 48
5.6 Summary of Actions 49
6 Expression Analysis and Implications 49
6.1 Approach: Engineering versus Science 50
6.2 A Control Paradigm 50
6.3 Cell Signalling: Intra and Inter Cell 51
6.4 A Model for Secondary Production 53
6.4.1 Linear Models 54
6.4.2 Second Order Models 55
6.5 Pattern Initiation 57
6.6 Methods for Verification and Validation 62
7 A Set of Examples 64
7.1 Example 1- Known Expression Model, Known Secondary Model, Unknown Color 64
The Additive Primary Colors: 68

The Subtractive Primary Colors 69
7.2 Example 2: Unknown Expression Model, Known Secondary Model, Known Colors from Many 69
7.2.1 Case 1: Fixed Controlling Gene Variable Suppressor/Activator 70
7.2.2 Case 2: Constant Controlling and Suppressor/Activator Genes 72
7.2.3 Case 3: Fixed Expression Gene and Present or Absent Suppressor/Activator Genes 75
7.2.4 Case 4: Noise 76
7.3 Example 3: To Be Determined Expression Model, Known Secondary Model, Desired Color 77
7.3.1 Case 2 Expression 77
7.3.2 Case 3: Present/Absent Suppressor/Activator Genes 78
8 Conclusions 79
9 Appendix A: An Example 81
9.1 The Genetic Structure 81
9.2 Color Model 83
9.3 The Model 85
9.4 Experiment and Data 87
9.5 Optimization 88
9.5.1 Modification for on/off A/R Genes 92
9.6 Design 93
9.7 Some Issues on Estimates 94
9.7.1 Averages of Functions 94
9.7.2 Conditional Estimates 95
10 Appendix B: Estimation 96
10.1 The Model 96
10.2 Color Model 98
10.3 The Estimator Model 100
10.4 Optimization 102
10.4.1 Modification for on/off A/R Genes 104
10.5 Model Variants 106
11 References 108
12 Web Sites 113

1 INTRODUCTION

The development of color in hybrids of plants can be viewed in many ways. One, the Mendellian approach, is that there is some element, called a gene, which is on a chromosome, and there is some mixing set of rules, dominant and recessive, which when applied allow for the control of the color development. The recent point of view is that there are genetic pathways which are controlled by enzymes, proteins, and that understanding the mechanisms of the control of these pathways is key to understanding the process.

However, the questions we raise herein, and then seek answers in the current literature, and finally propose possible paths of inquiry are as follows:

1. Given a dozen or more species plants which are relatively stable and consistent in the wiled, how does the variation in color in hybrids arise. Namely, what is the cellular basis of color, and moreover what is the genetic set of mechanisms which control this.

2. Given the complexity of color, form and variegation in the hybrids, what is the genetic basis for the control mechanisms intracell and intercell. For example, how are such colorations as eyezones formed and what is the intercellular communications mechanisms which effect this.

3. Given what now appears to be a set of well understood pathways control mechanisms by enzymes produced within the cell and the gene control mechanisms for expression of these proteins, how are these combined to produce intra cellular coloration and what are the inter cellular communications which spread the colors out over the inflorescence.

4. Given that we can answer the above, can we generate a mathematical control modem for gene expression and control and using the model approach the coloration problem as a problem of system identification or inversion.

5. Given that we could solve the above problem, then how could we invert the inversion and apply positive control to coloration and produce whatever color and for we would so desire.

We attempt in this paper to address these issues and set forth a combined understanding of what appears to be at this time a fragmented set of research efforts.

1.1 Overview

Our approach in this paper is fairly straightforward. We focus on a specific genus, Hemerocalis, and on a specific part of the plant in that genus, the inflorescence. The questions we ask are; (i) what is the cause of the colors we see in the flowers given what was in the original species, (ii) what are then pathways that generate the substances which produce the colors and what enzymes

control the pathways, (iii) how can we develop a system level model for this process, (iv) can we, using the system model, develop methods to develop desired colors.

Thus we proceed as follows:

1. *Review of the Genus:* The first step is to review the genus. Hemerocalis has about a dozen species, most of which can be interbred with one another. Some are self sterile and many can be bred. Hemerocalis has been hybridized for the past hundred years and many records of their ancestry exists. Thus it impossible using this genus to track many of the genetic linkages. The twelve species are all consistent within the species, there is some local variation and some geographical variation but it is possible to develop classification keys which generally are predictable and stable. The different species have a similar form but the colors vary between species, and there is even some color variation within species. In addition some of the current phenotypic species may also be variants of another species. We do not get into these arguments since the ultimate determinator will be a genetic classification, much of which is already under way. We then proceed to show how in the one hindered years we have been able to introduce significant variation in this genus. We then use this as a basis for developing a discussion as to why and how can it be controlled.

2. *Classic Genetics:* We review the classic Mendellian genetic approach and attempt to apply it to the species. We see that the concept of a gene as used by Mendel falls apart quickly in Hemerocalis and its hybrids. One would have to posit thousands of genes just for color and form and variegation, not to mention the other factors. The Mendel approach may work well for peas with limited characteristics but it has no places in this analysis.

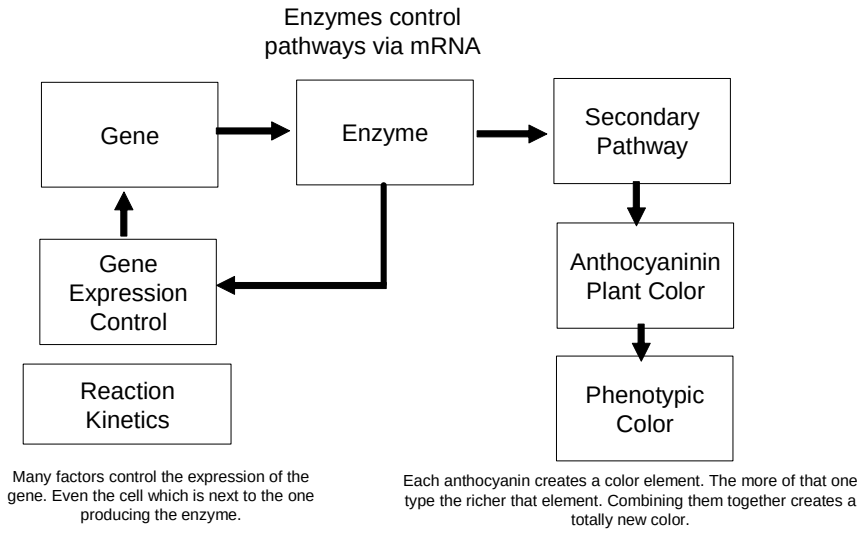
3. *Flower Color:* We then proceed to present an overview of the process of developing color in flowers. We present an overview of the anthocyanins, flavonols, and carotenoids. We review their pathways and summarize recent research which had identified the enzymes on each link of the pathway and the genes controlling those enzymes. This has been accomplished over the past few years and is critical to the understanding of the overall system approach.

4. *Cell Genetics:* We then provide a details overview of cell genetics and how activators and repressors are key elements in the overall expression of enzymes and in turn the development of color. We present a review of the cell elements and especially the process of gene expression. We discuss activators and repressors and the mechanisms of their actions. Their existence results from the work of Monod and Jacob in the early 1960s.

5. *System Models for Gene Expression:* There has been a recent development in the biological community of applying system models to biological systems. We build on that effort and develop medals for the expression of flower colors. Simply put, we recognize that color is a result of a mixture of secondary plant products such as anthocyanins. We can from the color of a flower determine what the mixture of each anthocyanin is. The concentration of an anthocyanin is a result of the concentration of the enzymes in the pathway which produces the anthocyanin, and typically the lowest enzyme concentration is the dominant factor. We also know that the concentrations of the enzymes are a result of activators and repressors, proteins also generated in a cell, which turn on or turn off the enzyme controlling the

pathway. Combining these ideas we can develop a top down system model for color. The output or observation equation is the color, and the system equation is a dynamic process wherein the states are the protein concentrations from a large enough set of gene expressions, wherein genes are allowed to control other genes via an nth order dynamic process. We also allow for uncertainty by adding a "noise" process which converts the overall system model into a linear dynamic stochastic system with observables. We then extend that model from a single cell to a matrix of interconnected cells. This then allows us to explore the processes one sees in the development of eyezones and other sharp transitions of color in flowers. We use models which have been previously studied for color variation and apply those to the flower.

In particular we will focus on each of the biochemical elements shown below:



1.2 Why This Genus

One of the first questions which can be asked is why this genus? There are many reasons for using this genus to stud the process of gene expression. The following are a few reasons:

1. The genus has been hybridized for just the last one hundred years. Thus there is a wealth of hybridization cross information to be able to asses what the genetic makeup is of the novel hybrids.
2. A great deal of recent research has provided detailed explanations for the control of color pathways and these apply directly to the genus.

3. The hybrids have been able to express color and form variants which are quite striking and allow for a clear identification of both pathways and gene expression mechanisms.
4. The genus is composed of s finite set of stable species. The underlying species of the genus Hemerocalis is generally well circumscribed and is currently under extensive study.
5. The genus does not appear to have significant transposon effects or viral effects. Unlike tulips and others species where viral changes are the generally more reflective cause of phenotypic change or in corn where transposons have a significant impact on phenotype, Hemerocalis appears to e dominated by gene expression changes.
6. The genus has multiple hybridizers making multiple changes per years. The American Hemerocalis Society lists over 50,000 hybrids and there are well over 500 active hybridizers in the US alone.[2] A typical hybridizer may make anywhere from 200 to 5,000 crosses per year and keep 1% of the crosses for registration, the remaining 99% going into a possible general pool of hybridizers "road kill".

For these and many other reasons Hemerocalis is an attractive genus.

1.3 Prior Work

The key prior works fall into three categories; (i) underlying genetic studies and understandings of the genus, (ii) detailed elucidation of the control of pathways and the effecting gene sequences, and (iii) the development and application of models for the analysis and synthesis of gene expression.

1.3.1 Genetic Structure of Hemerocalis

Various recent works by Chung in Korea and by Noguchi, Kang, De-yuan have provided detailed genetic analyses of the genus. Chung (2000) has provided a detailed study of three populations of the species hakuensis and has shown that there is a significant within species variation. This has been know for many years. This was a problem for many plant sytematicists who had few examples of species available and used this limited number to describe the species. It is necessary to perform extensive field work to fully understand the within species variation.

Noguchi and De-yuan (2004) have studied the species citrina. Their work included a detailed analysis of certain exons and an understanding of the evolution of this species. They have begun to establish a bases for genetic analysis of within species characteristics.

Kang and Chung (2000) have performed a detailed study of multiple species and included one from many cross breedings.

[2] The author is one of those hybridizers have introduced over fifty cultivars in the past twenty years. The author believes that it is essential in any science that one must have hands on experience with the subject matter at hand, either in the micro or macro, of optimally, both.

Tomkins et al (2001) have stated a detailed analysis of the full genetic variation in the genus using AFLP markers. They have also extended this to include many of the current common hybrids. Their key observation is that in the recent hybrids that they studied the genetic similarity has increased by approximately 10%. This demonstrates a rather interesting effect. Namely if the genetic diversity is decreasing and the phenotypic changes are increasing then it must be clearly via expression.

There is an excellent summary report by Lensaw and Ghabrial (2000) which discusses the impact of viruses and the tulip phenomenon in the Netherlands in the seventeenth century. This is a useful study since it let's us disregard the viral effects.

1.3.2 Gene Expression and Pathway Control

Within the last five to ten years there has been considerable growth in the understanding of the control of the pathways which provide for color. The recent paper by Winkel-Shirley (2001) provides a superb summary of this work. The author reviews prior efforts and puts the entire pathway management into perspective. She provides all controlling proteins and their causing genes. She does this for anthocyanins and flavones and isoflavonoids. The relationship to the abundance of effecting enzymes and anthocyanin expression as well as flavone expression is clearly demonstrated. This gives us a key in the development for latter model.

The work by Mol et al (1998) predates Winkel-Shirley and is the seminal paper on genetic control of pathways. Mol and his colleagues have continued to publish their results on further clarification of the pathway management by gene expression.

Holton and Cornis (1995) were the first to publish the full pathway. Their work is seminal in the area.

Work on carotenoid pathways has been completed by Naik et al (2003) and Bartel and Matsuda (2003).

1.3.3 Modeling of Gene Expression: Analysis and Synthesis

The development of a systems modeling for gene expression has most recently be exceptionally well articulated in the work by Szallasi et al (2006). In this work a collect of authors who are a mixture of systems experts and micro biologists present an up to data summary of all key works in this area.

The work by Hatzimanikatis and Lee (1999) is also an excellent modeling tool which applies a more Boolean approach to modeling the expression of genes. However the authors also extend their models to include mRNA and other pathways in a linear time varying system model as well. This latter approach coincides with the recent directions portrayed by Szallasi and is consistent with the approach taken herein.

Vohradsky (2001) has provided a neural network approach to the understanding of gene expression. Although highly flexible this model is at best amenable for limited simulation analysis.

Perkins et al (2004) have provided a detailed systems model for expression using classic dynamic systems models.

Chen et al (1999) have also provided a detailed dynamic model using their "differential equations" approach,

We see that Perkins et al and Chen et al have a great many similarities, as does the collection of authors in Szallasi but they all seem at best to just becoming aware of the wealth of well understood theory in the control and estimation area, see McGarty (1974).

2 HEMEROCALIS SPECIES AND COLOR

The species Hemerocalis is indigenous to Asia, specifically China, Korea, Japan, and Eastern Russia. It is a mountainous plant and is generally quite hardy. Since the late 19th century there has been a great deal of hybridizing of the plant. Thus for just over one hundred years, hybridizers have been cross the species and their descendents to create a wide variety of new and innovative hybrids. From the species which is predominantly yellow, orange and a brownish red color, comes a wide variety of forms and color. Bright reds, purples, shades of gold, doubles, plants with eyzones and plants with spider like form and shape. In this section we review the genus and its associated species and then we look at some of the hybrids.

2.1 The Species

The Genus Hemerocalis has a dozen or more species.[3] The identification of the species is still somewhat in flux. One of the earliest classifications was done by Stout in the late 30's and still stands with some modifications. There are many others who have proposed alternative classifications but when one looks at the literature one seem many differences and a few commonalities. We will in this paper not focus on a definitive classification but use several of the more well defined species to make the point.

Below we have shown several species and their variation. One must recognize several factors even in a species;

1. Species are geographically clustered. Thus citrina in one place will look like citrina in another but there may well be differences.
2. There has been some work on the genetic diversity within and between species. There is still a great deal more to be done.

[3] See the papers by Schabell. They are an excellent historical collection of the original works characterizing the species. The work by Stout still remains per-eminent. The work by Erhardt is somewhat useful but I have found inconsistencies and in addition it is extremely difficult to see an overall structure.

3. Many species are self sterile, such as citrina, but can be crossed with other species to create hybrids.
4. Some species like fulva Europa are triploid and are sterile and propagate via a vegetative process.
5. Variability exists within species and within the same geographic area, and one sample of a species may not look exactly like another from the same location, however the variation is a micro variation, one could still identify the species from the collection of phenotypic characteristics.
6. None of the species expresses the characteristic we see in many of the newer hybrids, and that will be a question key to this analysis.

Note when looking at the flower colors we see yellow, reds, some darker brownish reds, and orange. There is some variation of color.[4]

H. altissima

H. aurantiaca

H. dumortieri

H. citrina

If we look at the above four species we see four distinct colors, shapes and color patterns. The dumortierii is an early spring flowering plant whereas the altissima is late summer and early fall.

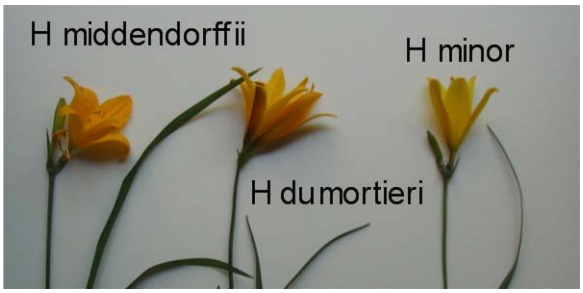
[4] The recent paper by Tompkins is useful since it uses the AFLP approach to determine a broad base of cross species variability as well as geographical variability.

Citrina and aurantiaca are mid summer plants but citrina is night blooming and is self sterile. However we have successfully crossed citrina and aurantiaca.

We can continue with the species as follows:

H. minor

H. middendorfii



Comparison

H. dumortierii

The above also shows the branching habit of the flower. The three shown are all early blooms.

The coreana species is show below. We have two plants obtained from two sources. Note the difference in color.

H. coreana variation; The left one above has been growing for several years, Right one is recent acquisition. Both have bracts, large ones which show it to be coreana. Why the color difference? Is it a variety, geographically different part of species, early color to change latter?

The final selection of species plants is shown as follows:

H. minor

H. middendorfii

H. coreana

H. fulva

H. hakuensis

H. multiflora

The geographic distribution of the species in their original locations in China are shown below.[5] Note the many locations and that most of them are mountainous. There are other locations in Russia, Korea and Japan. Generally they are found in mountainous regions.[6]

Origin

Species
H. altissima **BLACK**
H. citrina **GREEN**
H. dumortieri **PINK**
H. esculenta **PURPLE**
H. flava **BLUE**
H. forrestii **DARK GREEN**
H. fulva **DARK YELLOW**
H. lilioasphodelus **YELLOW**
H. middendorffii **CYAN**
H. minor **RED**
H. nana **SAND YELLOW**
H. plicata **LIGHT RED**

We have also created a data base for the phenotypic characteristics of the plants in our current collection. This is shown below. One could use this data to create a clade analysis and then also create a key. We have done that in another report and we feel that the result should be considered as preliminary.

[5] This has been prepared by Yong Wen, of MIT, one of my graduate students.

[6] We have found that our best results are on the lower slopes of the White Mountains in New Hampshire. The climate is similar to the areas where the species come from originally. It is cool, and in Zone 4b with lowest temperatures at -35 F and is sand soil with 52" to 58" of rain per year.

Name	*Data*	*Length Leaf (in)*	*Width Leaf (in)*	*Height of Scape (in.)*	*Diameter Flower (in.)*	*Petal Width (in)*	*Petal Length (in)*	*Sepal Width (in)*	*Sepal Length (in)*	*Branches per Scape*	*Buds per Scape*	*Petal Color*	*Sepal Color*	*Fragrance*
middendorfii	5/19	10	0.75	10.00	2.00	0.38	2.20	0.75	2.20	1	2	orange	orange	n
minor	5/19	24	0.38	18.00	3.00	1.00	2.50	0.63	2.50	1	3	yellow	yellow	n
dumortieri	5/28	18	0.50	21.00	3.00	0.38	2.20	0.25	2.20	1	3	orange	orange	n
middendorfii	5/22	23	0.50	15.00	3.00	1.00	2.50	0.75	2.50	1	3	orange	orange	y
dumortieri	5/22	23	0.50	14.00	1.75	0.75	2.20	0.50	2.00	1	3	orange	orange	y
hakuensis	7/5	29	0.50	32.00	4.50	1.15	3.50	1.00	3.00	2	4	orange	orange	y
aurataniaca	7/7	42	0.75	40.00	6.00	1.00	4.50	0.50	4.50	2	4	orange	orange	n
Kwanso	7/7	34	1.00	40.00	6.00	1.50	3.00	1.00	3.00	2	5	orange	orange	n
fulva	7/7	53	1.00	46.00	7.00	1.50	4.00	0.75	3.50	2	5	orange	orange	n
coreana	7/21	42	1.50	42.00	5.00	1.50	4.00	0.75	4.00	2	3	orange	orange	y
citrina	7/21	30	0.75	36.00	3.00	0.50	4.00	0.25	4.00	7	2	yellow	yellow	y
multiflora	7/21	25	0.75	34.00	2.00	0.75	3.00	0.50	3.00	5	3	orange	orange	n
dumortieri	7/22	35	0.75	30.00	4.50	1.25	3.50	0.80	3.50	8	2	orange	orange	y
altissima	7/29	48	0.75	57.00	5.25	1.00	3.50	0.50	3.00	5	3	orange	orange	y
multiflora	8/9	30	0.75	44.00	3.00	0.80	2.20	0.50	2.20	7	3	orange	orange	y

If we were to take the various authors who have attempted to characterize the species we obtain the chart shown below. One thing evident in the chart is the lack of agreement. Again we believe that agreement can only be obtained after a detailed genetic analysis.

Name	*Stout (1934)*	*Erhardt (1992)*	*Plodeck (2003)*	*Munson*	*Hortus Third (1976)*	*Peat & Petit (2004)*	*Grenfell (1998)*	*Petit & Peat (2000)*	*PFAF (2000)*	*McGarty*
H altissima		X	X	X	X	X	X	X	X	X
H aurantiaca	X	X	X	X	X	X	X	X	X	X
H citrina	X	X	X	X	X	X	X	X	X	X
H coreana		X	X	X		X	X	X	X	X
H darrowinina						X	X		X	
H dumortieri	X	X	X No Picture	X	X also calls it H sieboldii	X	X	X	X	X
H esculenta		X	also he calls dumortierii and middendorffi v esculenta in this species	X		X	X	X		
H exaltata	X	X	X	X			X	X	X	
H forestii	X	X	X No Picture	X	X	X	X	X	X	
H fulva	X	X	X		X	X	X	X	X	X
H graminea				X	X also H dumortieri and H minor		X	X	X	
H hakuensis		X	X	X		X		X	X	X
H honngdoensis						X				
H lilioasphodelus	X H flava	X	Also H flava	X Calls it H flava	X	X	X	X uses H flava	X	X
H littorea				X				X	X	
H longituba				X						
H micrantha		X	X no picture	X		X			X	
H middendorffii	X	X	Also H dumortierii v middendorffii	X	X	X	X	X	X	X
H minor	X	X	X	X	X	X	X	X	X	X
H multiflora	X	X	X	X	X	X	X	X	X	X
H nana	X	X	X No Picture	X	X	X	X	X		
H pedicellata		X	X No Picture	X		X			X	
H plicata	X	X	X No picture	X	X	X	X	X	X	
H taeanensis						X				
H thunbergii	X	X	X also H serotina and H sulphurea	X	X but calls it H serotina	X	X	X	X	X
H yezoensis		X	X also H flava v yezoensis			X	X		X	

Some authors have placed these species in groups. We have shown this in the following Table. Erhardt seems to be setting the standard but there are several inconsistencies in his approach, Peat and Petit appear to be repeating Erhardt. There is no true well established and accepted classification, however.

Name	*Stout*	*Erhardt*	*Plodeck*	*Munson*	*Hortus Third*	*Peat & Petit*	*Grenfell*	*Petit & Peat*	*PFAF*	*McGarty*
Groups										
	1. Forked; Scapes < Leaves (nana, plicata, forestii)	Fulva (aurantiaca, fulva)				Fulva (aurantiaca, fulva, hondoensis, taeanensi)				
	1. Forked; Scapes > Leaves (flava, minor, thunbergerii, citrina, fulva, aurantiaca, exaltata, multiflora)	Citrina (altissima, citrina, coreana, lilioasphodelus, minor, pedicellata, thurbergii, yezoensis)				Citrina (altissima, citrina, coreana, lilioasphodelus, minor, pedicellata, thurbergii, yezoensis)				
	2. Unforked, Unbranched (dumortierii, middendorfii)	Middendorffii (dumortieri, esculenta, exaltata, hakuensis, middendorffii				Middendorffii (dumortieri, esculenta, hakuensis, middendorffii)				
		Nana (forestii, nana)				Nana (darrowinia, forestii, nana)				
		Multiflora (micrantha, multiflora, plicata)				Multiflora (micrantha, multiflora, plicata)				

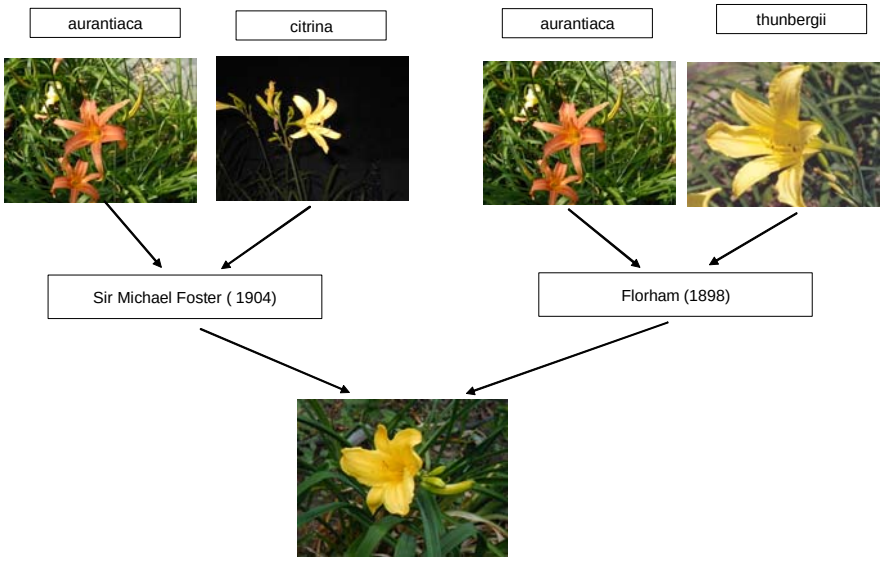
2.2 The Early and Latter Hybrids

The hybridizing of this plant has been documented by several other authors and will be not be detailed herein.[7] However we do want to present an overview of the changes which have occurred as a basis for the questions which have been presented earlier.

The following Table depicts six hybrids dating from 1924 through 1999. The change is startling. The Hyperion Hybrid was just one generation from the species, it was the offspring of two separate species crosses. It bears no resemblance to any of its ancestors. In many was it represents one of the first true hybrids.

The Hyperion cross is shown below:

[7] There are many good works on the hybrids. The classic is Munson albeit a bit dated. The recent work by Peat and Petit is excellent.



The following Table depicts seventy five years of hybridizing from Hyperion in 1924 through 1999. Potentate was the first truly red flower and became a benchmark which holds even today. Prairie Blue Eyes was an attempt to obtain a blue, close but not totally there. There have been attempts at whites as well with considerable success. The other more recent flowers show increasing complexity. The 1999 flower, Now and Zen, shows an eyezone, a colored or tinted edge to the sepals and petals and a well demarcated throat region as well.

Hyperion (1924)

Potentate (1943)

Prairie Blue Eyes (1970)

Outrageous (1978)

Wings of Chance (1985)

Now and Zen (1999)

With the above development there are several questions which we can ask:

1. Hyperion was the F2 development of species crosses, or at least that is the way it appears from the early literature. Hyperion has a substantially different form from any of the F1 plants or the species. What accounts for this change? Hyperion has a color which is a stronger yellow than citrina and does not reflect any of the variegation of the other F0 parents. What, has a set of genes been suppressed, if so which ones.
2. Potentate has what is called a “throat” a gold region inside the flower. Throats like this do appear in the species. What controls the throat characteristic and this throat is quite clear. The

transition between throat color and the predominant color of the flower is very abrupt, what genetic switch allows for this abrupt change?

3. Prairie Blue Eyes is one of the early attempts to get a blue. One would assume it would be possible. Only recently has genetic engineering produced a blue rose. The question then is can we breed a blue daylily or does it require some form of genetic engineering?
4. The Outrageous flower starts to show the dramatic change in form as well as color. It is a recurved flower with significant color variation. In this case form as well as color is being changed. What are the genetic linkages between them?
5. Wings of Chance and Now and Zen show how quickly genetic variation can proceed. The throat becomes an eyezone, a region from yellow to red to yellow. And in Now and Zen we see edging colors appear on the end of the petals and sepals. Again, what is the gene expression control mechanism which effects this unstable change. We have many examples from dynamic systems, can we apply them here?

What we then ask is; knowing the species and assuming the species have some steady set of genes, and that the genes in the species express themselves so as to generate the colors we see on the species, and furthermore given that we have not introduced any new genes nor have we mutated any of these plants, how, through hybridizing alone have we managed to allows new combinations of genes to be expressed and to have existing genes expressed at new and greater or lesser rates than the species plants.

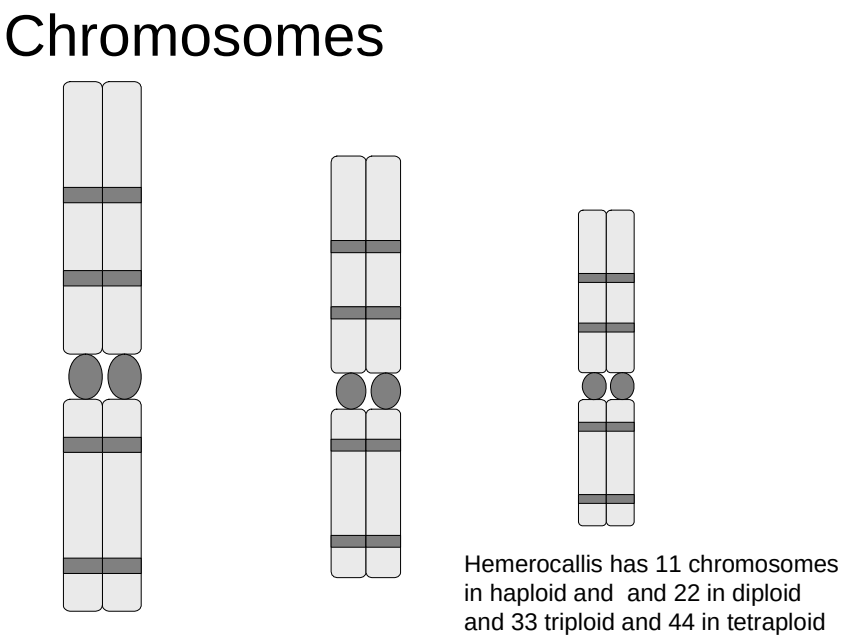
3 CLASSIC GENETICS

In this section we present an overview of the classic Mendellian analysis.[8] The Mendellian analysis makes classic assumptions which prevailed until the advent of the Watson and Crick model, and even slightly beyond. In fact many breeding programs build upon a Mendellian approach. We argue that such an approach is partially correct but lacks most of the key elements which must be considered.

3.1 Mendellian Analysis

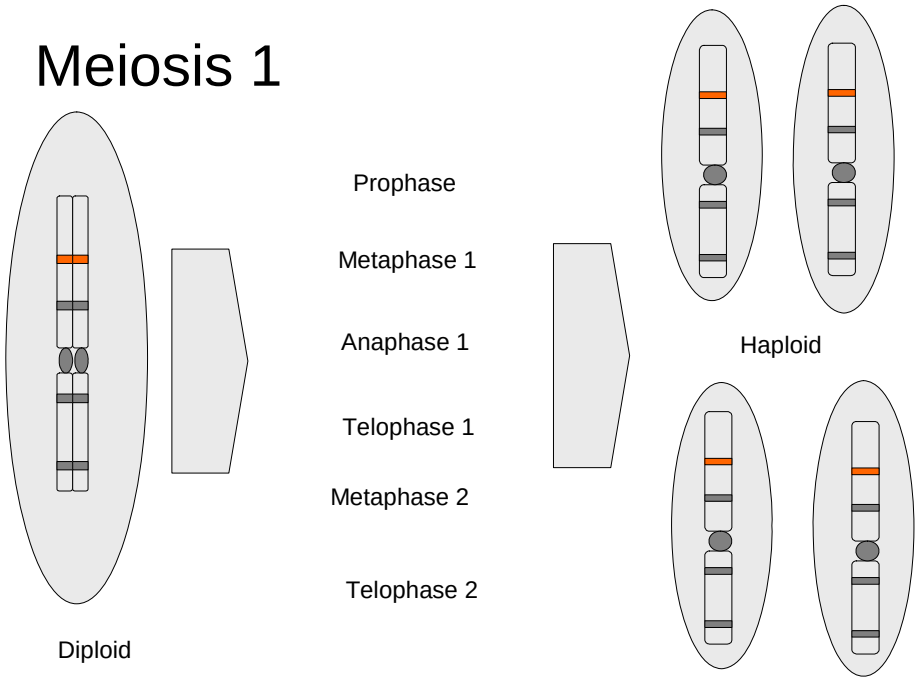
In Mendellian analysis we assume that there exist a gene on a chromosome which provides some characteristic, say yellow petals. Hemerocalis has 11 chromosomes with a diploid being the normal configuration. There is no sex chromosome as in humans. There are triploids with 33 chromosomes and tetraploids with 44 chromosomes. But the species has 22 in each cell. We characterize this as shown below.

[8] See Griffiths. This is an excellent overview of genetic analysis.



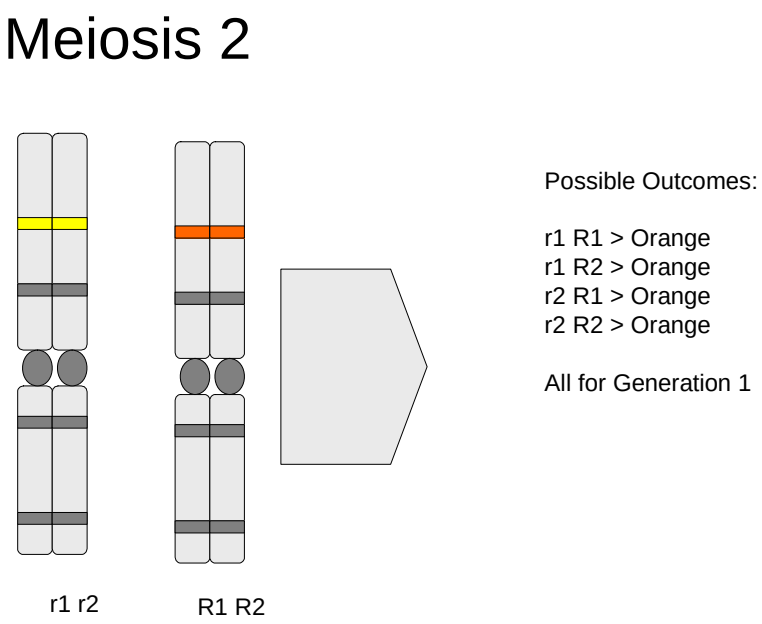
Now each chromosome with some gene segment has some controlling characteristic, such as a gene for the color orange. In the process of meiosis in the sex cells the chromosome pairs split, some even cross over, and a mixing and matching of chromosomes and genes are made. Our intent is not to provide a detailed summary of Mendellian analysis but to focus on the key points which will be used to continue our analysis.

When a plant creates a pollen grain or a female oocyte they are products or meiosis. And these cells are haploid, namely only one copy of the chromosome. Thus is we have two orange genes, one on each chromosome, then in meiosis we end up with a male and female haploid cell each having one orange gene.

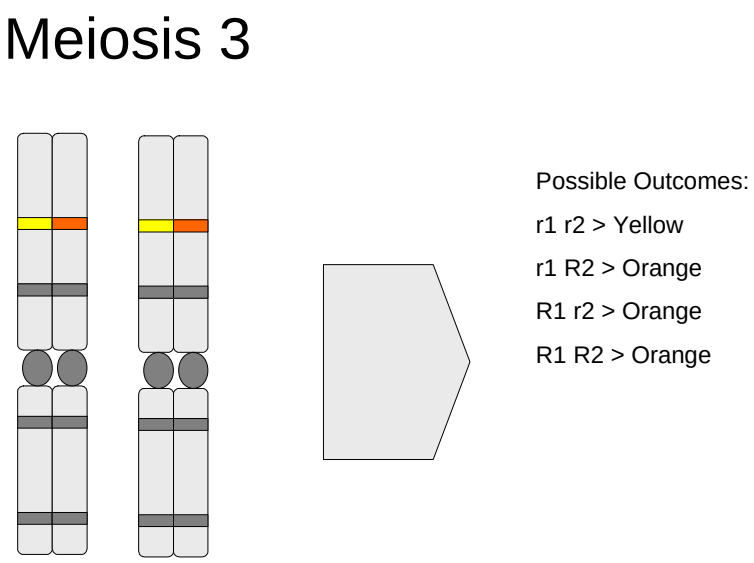


If, however, we take a yellow plant, allow it to create the haploid cells via meiosis and take a pure orange plant, let it create haploid cells via meiosis and then cross these plants we get what we see below.

Namely there are four possibilities, each equally likely; we have a yellow with an orange in any one of the ways shown below. Thus in what is called the F1 generation we only get orange plants since the orange gene is dominant and each of the F1 plants have the same genetic makeup, a haploid with a yellow and a haploid with an orange.



Now we go to the F2 generation. This is the offspring of the F1. Remember that all F1 have same gene structure, a yellow and an orange gene. These break apart in meiosis and combine again when the plants are fertilized. The net result in the off spring in F1 is a set of chromosomes with a yellow and orange chromosome. When they split there is a possibility of the off spring of the off spring in the F2 to have two yellows which means yellow or one of each yielding orange or a pure orange. Thus with one gene we find that a dominant gene will give $1/4^{th}$ with the recessive and $3/4^{th}$ with the dominant color. We show this below.



3.2 Breeding and Hybridizing

Now what does this tell us about hybridizing daylilies. Frankly very little. Mendel had peas, and he was looking at peas all one color, one gene one phenotype. There was no mixing, no complicated gene control. There could be a simple control of a gene and a phenotypic characteristic.

For example, if we had a black daylily with an eyezone, and black was dominant as was an eyezone then the table below predicts the result. This is the classic Mendel analysis.

Mendel's Genetics

		Male Gametes			
		BA	Ba	bA	ba
Female Gametes	BA	BABA Black, Eye			
	Ba				
	bA				
	ba				baba White, no eye

However, this is not the case. Go back and look at the species and then look at the hybrids. How does one go from here to there? That is a key question. Genes are being expressed differentially in various was and the control of those expressions varies across the sepal and petal. That is an issue we wish to explore.

3.3 Summary of Mendellian Approach

We can summarize the world view of a Mendellian:

1. Genes exists and are parts of a chromosome.
2. There is a one to one relationship between a gene and some phenotypic characteristic. The gene control that characteristic.
3. A gene may be dominant or recessive, namely there may be a stronger effecting gene than another.
4. To get a characteristic the plant must have a gene which expresses that characteristic.
5. Some genes are sex related or may have some effect on other genes but that is not a significant factor.
6. The gene is the operative entity and there is not accounting for pathways, expression, activation or suppression.
7. Mendel's approach fails to account for DNA and the underlying pathways.

The message to take away from the Mendellian analysis is simply; in hybridizing there is no simple one to one relationship between gene and phenotypic characteristic. What we see is a

complicated system of variable gene expression; over and under expression, and the release of the gene products related thereto. We look at this in the next section.

4 FLOWER COLOR EXPRESSION

We have just shown that there are a wide variety of coloration in the daylily. In a little over a hundred years we have taken the dozen or so species and intermixed them and as a result have created a very complex set of flowers with characteristics which differ dramatically from the species.[9] The species have managed to maintain their separate identities over thousands of years but in a small fraction of time we have been able to introduce multiple forms and colors. To understand this process we first have to understand where the colors come from. How do we get purple from a plant which is red, yellow, orange and possibly even brown? How are the colors made and how do we get from there to where we are today.

The first step in understanding that process is to understand the pathways that lead to color production in a single cell. Then we can address the issue of multiple cells and finally how the cells communicate. How do we get an eyezone for example. Why if a cell is whit do we go so abruptly to a purple eyezone. What is the mechanism for this process? We begin the exploration of this issue with a analysis of the underlying pathways.

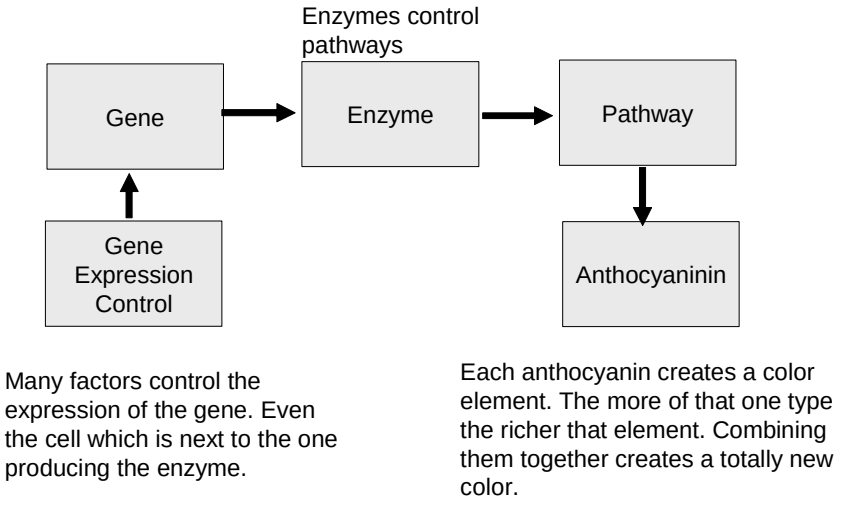
4.1 Pathways and Enzymes

Pathways are nothing more than a set of chemical reactions which get us from some primitive chemical to a more complex but useful chemical structure.[10] In fact the pathways may be just a set of processes going from any one chemical structure to another independent of the nature of the starting and starting chemical. Some pathways are linear going from a beginning to and end and some are circular taking us from the beginning and back again; the Krebs cycle is an example. What makes the pathway work? Just three elements are required: (i) the underlying chemical constituents, (ii) some form of energy, (iii) generally some form of facilitation such a catalyst and in our analyses this is an enzyme.

The general flow structure we look at is shown below. In our view, not the only such view but one convenient for the development of our argument, we have the pathway but it facilitated by an enzyme, a protein. The protein is generated by a gene. And the gene is activated by some other element, generally another protein. In our case shown below the output is some anthocyanin. The more of the enzyme, namely the more the gene expresses itself the more anthocyanin we get. Thus if we can get the gene to express then we get more of that specific anthocyanin, more pelargonidin for example. We defer to the next section how we get this gene to express so strongly.

[9] See Lensaw and Ghabrial for an excellent discussion of the tulip. In contrast to the daylily, the tulip craze of the seventeenth century was a dramatic bubble, and the irony was that most of the color variations were induced by viruses.

[10] See Taiz for an excellent overview. Dey is also a superb and current reference. The older references by Goodwin are useful but they fail to account for the genetic effects.



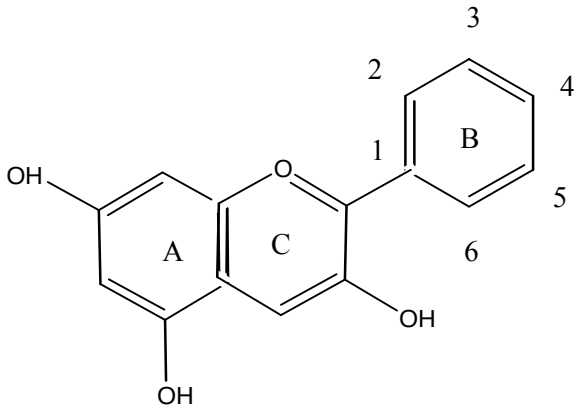
The opposite is also true. Namely if we can suppress the gene then we can get less and even possibly no anthocyanin from the pathway. This is the first step in the development of an overall system model.

4.2 Anthocyanins

Let us consider our first pathway. This is the pathway which creates anthocyanins.[11] The anthocyanin molecules is shown below. Note on the B ring we have six sites to which we can attach differing molecular chains. This will be an important element when we see the different configurations and their implications.

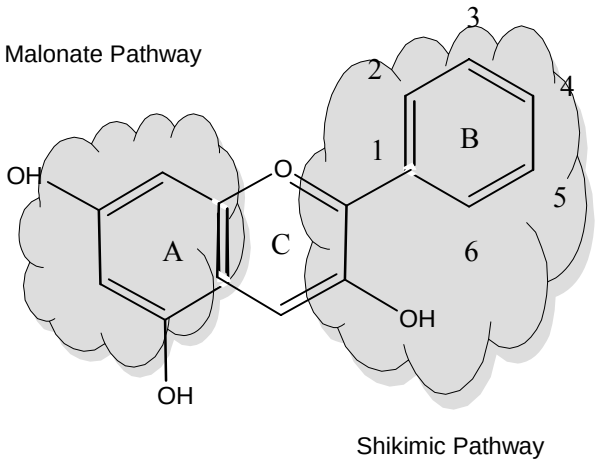
[11] See the papers by Mol and also by Winkel-Shirley. They are excellent in the characterization of the pathways. Also the papers by Holton and the one by Jaakola are quite useful here as well.

Anthocyanidin



The anthocyanin or anthocyanidin molecules comes from two different pathways. In the figure below we have taken the basic resulting molecule and have shown that there are two elements; one is from the shikimic pathway and the other from the malonate pathway. This means that we have to understand both pathways to understand the ultimate abundance of the product.

Anthocyanidin



Before continuing we want to look at what the results would look like if we have different substitutes on the B ring. In the Table below we show that the terminations on the 3, 4 or 5

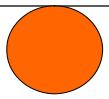
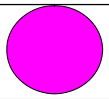
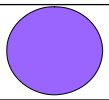
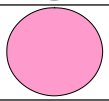
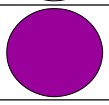
elements yield different results. The results give pelargonadin, cyanidin, delphinidin, peonidin, and petunidin. Each obviously named after their related flower and each resulting an anthocyanin of a different color.

Colors

Anthocyanidin	*Substituents*	*Color*
Pelargonadin	4'-OH	orange-red
Cyanidin	3'-OH, 4'-OH	purplish red
Delphinidin	3'-OH, 4'-OH, 5'-OH	bluish purple
Peonidin	3'-OCH_3, 4'-OH	rosy red
Petunidin	3'-OCH_3, 4'-OH, 5'-OCH_3	purple

In the Table below we have shown the colors of each of these as well as the weighting of a red, green ad blue combination which best matches the color. Thus one can in an 8 bit color schemes, as one would find in any PC color scheme, get the resulting anthocyanin colors by blending the R, B, G elements to yield what we are seeking. This relating the colors back to RBG is critical since it get reflected in the ultimate flower color.

Colors (R, G, B)

Pelargonadin (255, 102, 0)	
Cyanidin (255, 0, 255)	
Delphinidin (153, 102, 255)	
Peonidin (255, 153, 204)	
Petunidin (153, 0, 153)	

Now if we assume we have only anthocyanins for color, and that we have the above combinations available, we ask how do we combine these colors in a weighted manner to obtain the desired color. This approach is critical to the overall understanding. First we show by a weighted RBG we get the color we seek or the color which is presented. The we assume that if we can then do the same for each anthocyanin, then we can create any desired color from a weighted collection of anthocyanins. This means that we can then determine what the relative percents of expression of any anthocyanin is and this lets us then go back to how strongly the gene for that anthocyanin is expressed. The model we presented earlier will be a key element in this overall process.

No let us start with a simple expression. For any color we have by definition:

$$\text{Color} = \alpha\,\text{Red} + \beta\,\text{Blue} + \delta\,\text{Green}$$

For example, we may have a (0,0,255), or a (128, 128, 128). Or any other set of combinations.

Likewise we could state this by means of some combination of anthocyanins. Namely:

$$\text{Color} = \text{a}\,[\text{Pelargonidin}] + \text{b}\,[\text{Cyanidin}] + \text{c}\,[\text{Delphinidin}]$$

but we can relate the anthocyanins to the basic colors or red, blue and green as:

$$[\text{Pelargonidin}] = \alpha_P\,\text{Re}\,d + \beta_P Blue + \delta_P Green$$

$$[\text{Cyanidin}] = \alpha_C\,\text{Re}\,d + \beta_C Blue + \delta_C Green$$

$$[\text{Delphinidin}] = \alpha_D \,\text{Re}\, d + \beta_D Blue + \delta_D Green$$

If we define a color vector of Red, Blue and Green as:

$$M = \begin{bmatrix} \text{Re}\, d \\ Blue \\ Green \end{bmatrix}$$

Thus if we define the mix vector as **m** then we have:

$$Color = m^T M$$

or:

$$m = \begin{bmatrix} \alpha \\ \beta \\ \delta \end{bmatrix}$$

But we have the following matrix:

$$\Lambda = \begin{bmatrix} \alpha_P & \beta_P & \delta_P \\ \alpha_C & \beta_C & \delta_C \\ \alpha_D & \beta_D & \delta_D \end{bmatrix}$$

which yields:

$$Color = m^T \Lambda M$$

The above analysis shows us that we can analytically determine the expression of the anthocyanins from the color of the cell by means of the above formulas. These are relative expressions but by benchmarking any one element we can make them all absolute in the cell as well.

4.3 Other Color Elements

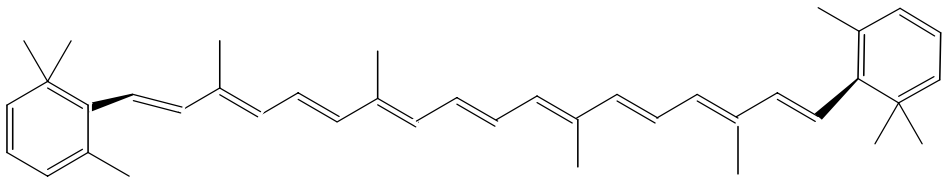
Anthocyanins are not the only elements which are secondary products which produce color. There are three classes of chemicals which give rise to color; anthocyanins, flavones or flavonols, and carotenoids. The Table below depicts the different elements and their colors. The approach we took above for the anthocyanins can be take for the flavones and carotenoids as well. It should be noted that there may not be a unique solution here but there are several possible but they can be narrowed down by actual determination of one to three elements as baseline.

Class	*Agent*	*Color*[12]
Anthocyanidin		
	Pelargonidin	orange-red
	Cyanidin	purplish-red
	Delphinidin	bluish-purple
	Peonidin	rosy red
	Petunidin	purple
	Malvinidin	
Flavonol		
	Kaempferol	ivory cream
	Quercetin	cream
	Myricetin	cream
	Isorhamnetin	
	Larycitrin	
	Syringetin	
	Luteolin	yellowish
	Agipenin	Cream
Carotenoids		
	Carotene	orange
	Lycopene	Orange-red

We now summarize the other element classes.

4.4 Carotenoids

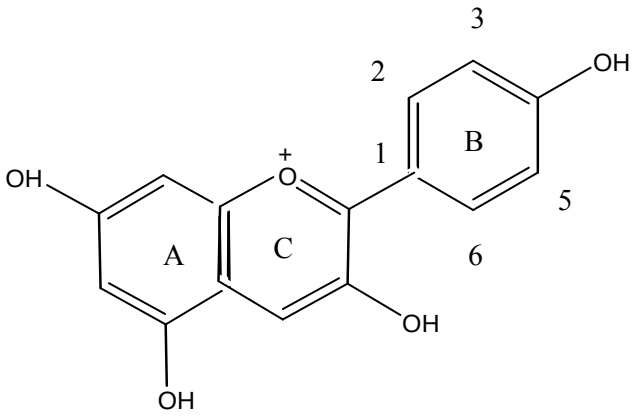
Carotenoids are what is quite common in the carrot, the orange hew we see in that root. Its molecular structure is shown below, this is beta carotene.



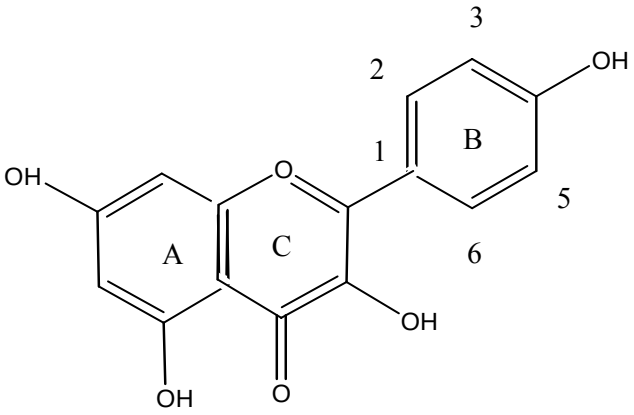
[12] See Taiz p. 334 for the anthocyanidin color and Bernhardt for the flavonol and carotene.

4.5 Flavones

The flavonols, or flavones are quite similar to anthocyanin. Their structure is shown below. Note that we have compared it to that of anthocyanin.



Anthocyanidin



Flavonol

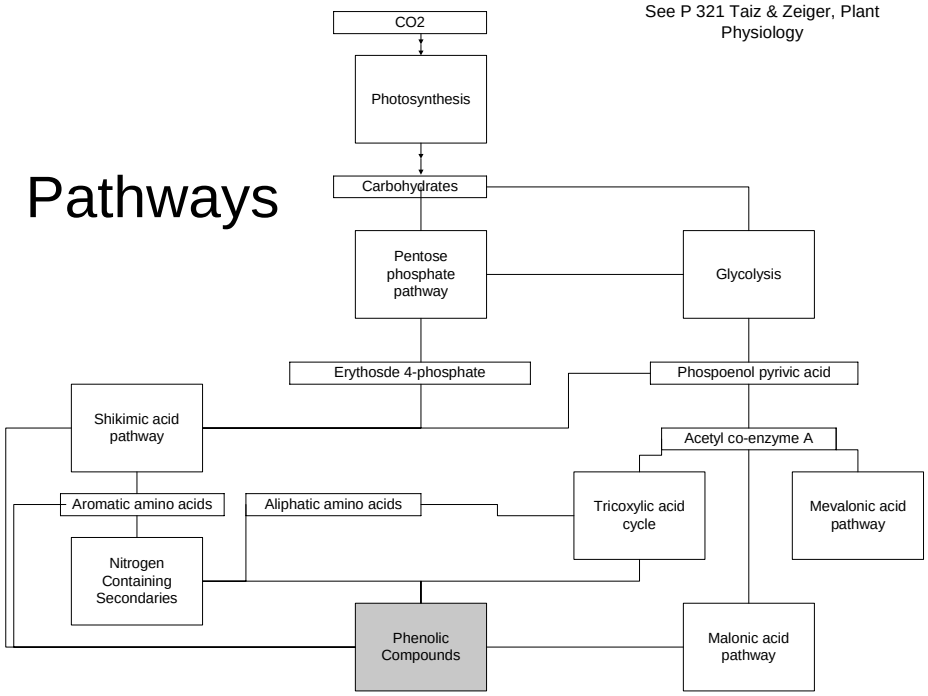
We can also show how closely they relate in substitutions and colors. This is shown in the Table below.

Flavonol	Anthocyanidin	Substitution	
		3’	5’
Kaempferol	Pelargonidin	H	H
Quercetin	Cyanidin	OH	H
Myricetin	Delphinidin	OH	OH
Isorhamnetin	Peonidin	OCH_3	H
Larycitrin	Petunidin	OCH_3	OH
Syringetin	Malvinidin	OCH_3	OCH_3

4.6 Pathways

In this section we present the pathways for the three classes we have described above. We first present an overview of the pathway and then we present the details of the pathway and the enzymes used in each step. The key observation is that we must have enzymes to go from step to step in the pathways and that if any one enzyme is missing we cannot proceed on that path, and further the path with the small amount of enzyme becomes the limiting path. Thus, we do not have a one to one map here. The production of any one anthocyanin, for example, if limited by the lowest produced enzyme, and the other enzymes may be present in abundance.

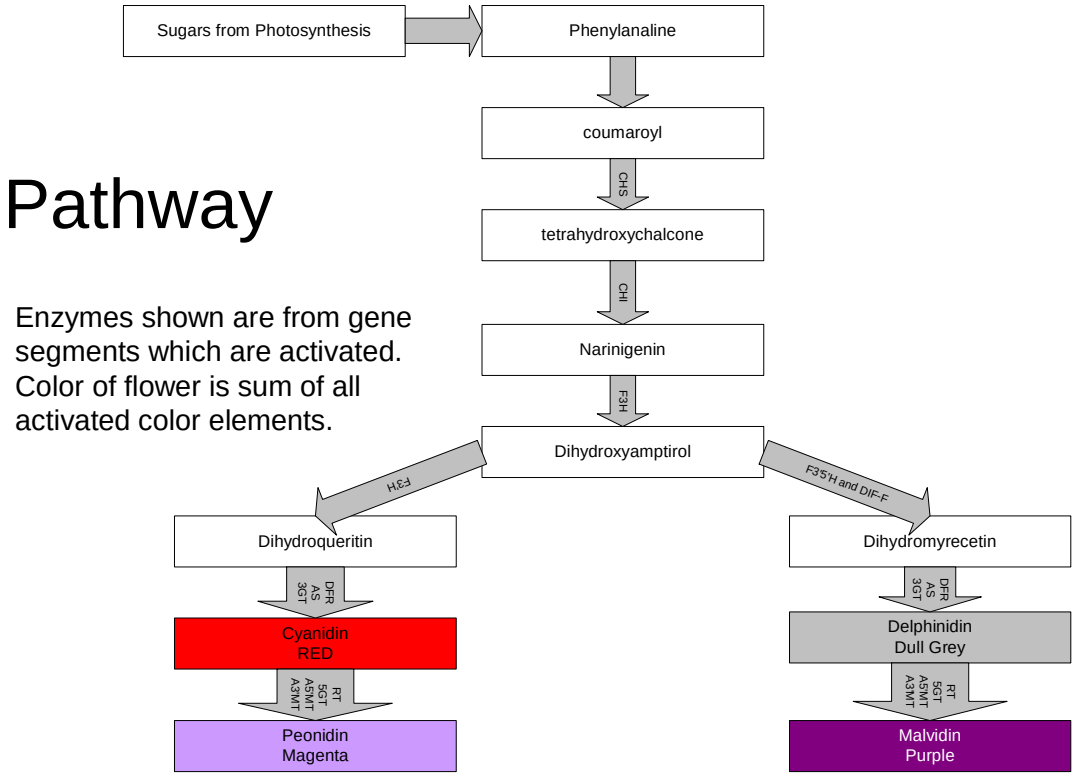
The following is the overall pathway for all elements.



The above shows how we start from CO2 and then go through a variety of other pathways. We will review those pathways in some detail since it is the enzyme control in them which is key.

4.6.1 Anthocyanin Pathway

The anthocyanin pathway with the controlling enzymes is shown below. The enzymes are presented in the arrows linking each step in this pathway. This pathways shows the start as a sugar element and then goes to phenyanaline and then down through the chain to one of the four indicated anthocyanins.

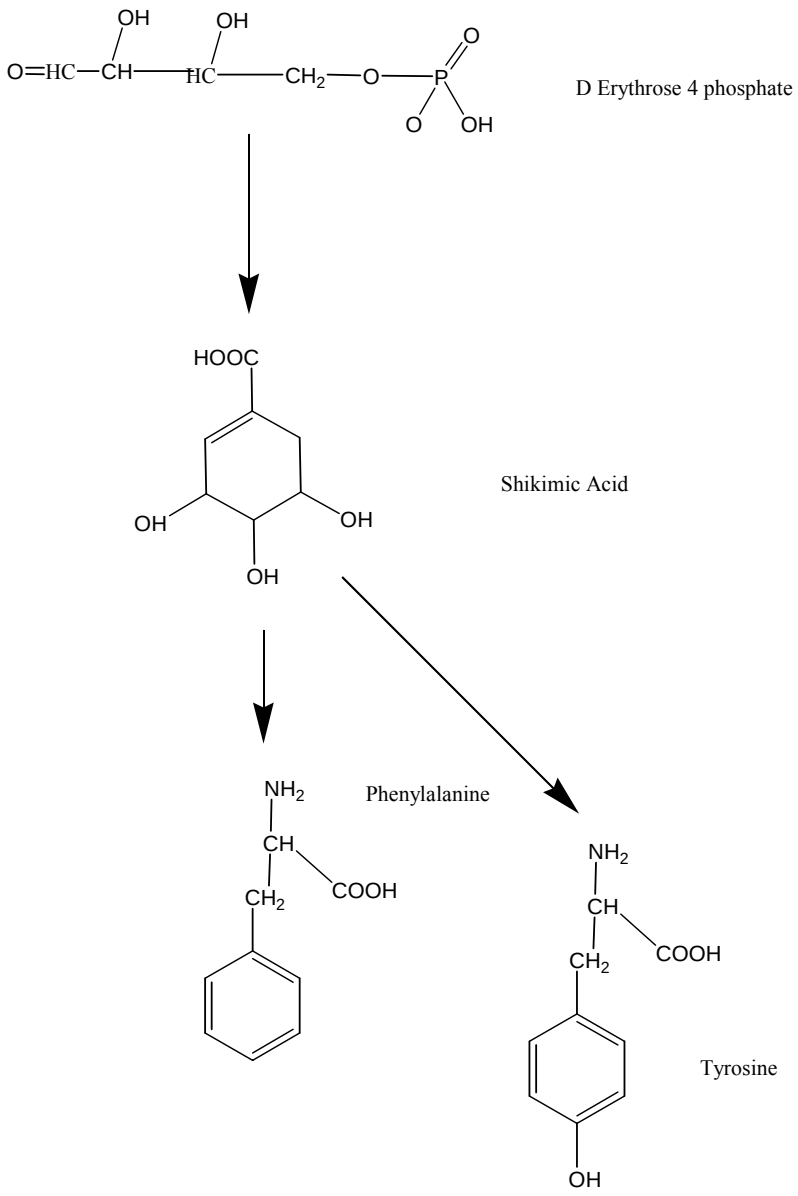


Note that at each step there is an enzyme element. The genetic loci for cloned flavonoid enzymes in Arabidopsis are shown in the following Table.[13]

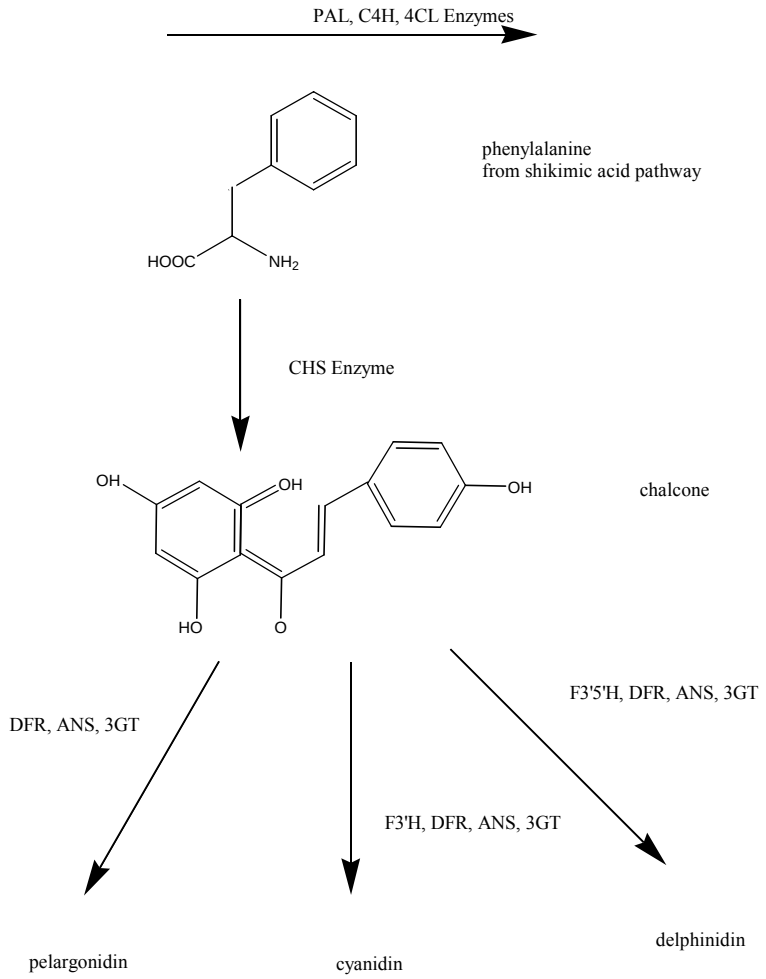
[13] See Similar information for maize, petunia, and snapdragon is described by Holton and Cornish (1995). b Based on the AGI map, 11/12/00; numbers in parentheses refer to P1 or bacterial artificial chromosome clones on which these sequences reside. c Transposon- tagged mutant for FLS1 (Wisman et al., 1998).

Enzyme	Locus	Chromosome	Map Position
CHS	tt4	5	7,050 kb (MAC12)
CHI	tt5	3	21,000 kb (T15C9)
F3H	tt6	3	19,600 kb (F24M12)
F39H	tt7	5	4,400 kb (F13G24)
FLS	fls1<Enc	5	FLS1: 4,700 kb (MAH20) FLS2-5;: 32,150 kb (MBK5) FLS6: 24,350 kb (MRH10)
DFR	tt3	5	23,800 kb (MJB21)
LDOX	tt19	4	16,900 kb (F7H19)
LCR	ban,ast d	1	26,800 kb (T13M11)

The pathway for the conversion of the sugar erythrose to penylanaline is shown in the reaction below. This accounts for the upper part of the pathway which we have shown. It uses the Shikimic pathways which we have shown in the initial discussion on the pathways.



The conversion details from pehnyanaline through chalcone to the anthocyanins is shown in the reaction below. We have reiterated by transition the enzymes which facilitate each step in this process.



What these process point out can be summarized as follows:

1. There are common pathways which are operational in all plants for the generation of the pigments.

2. Enzymes used as activators modulate the amount of production of the enzymes.

3. The products of these pathways, the anthocyanins, are driven by the concentration of the facilitating enzymes. Specifically we can write:

$$z_k = \min(B_j x_j; \forall j \in \Theta)$$

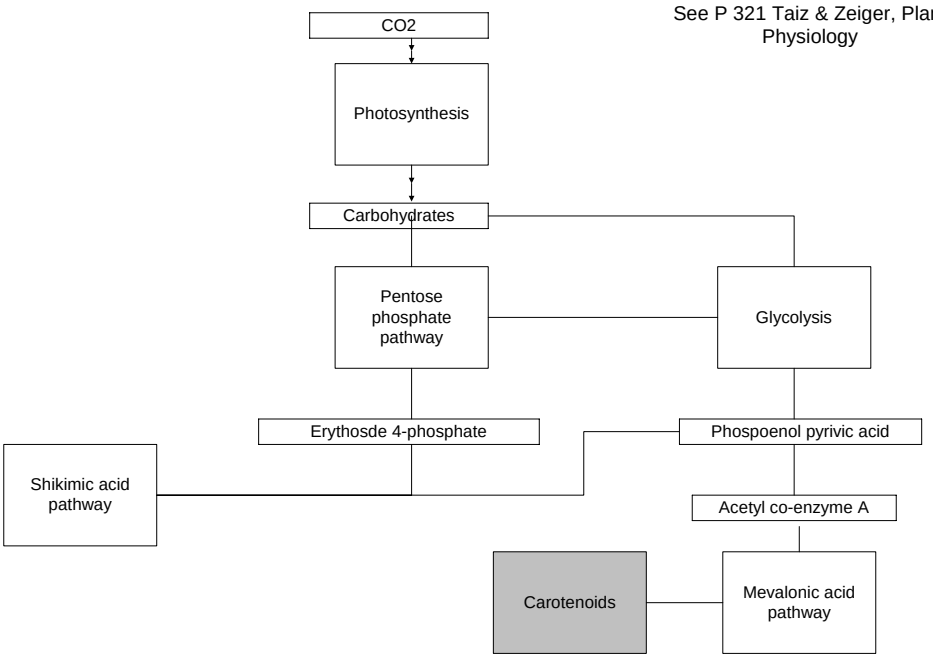
Namely the concentration of the secondary product, the z element, is proportional to the minimal concentration of the facilitating enzymes, namely the set θ.

Secondary products always have this type of production process. As we look at a cell, from a system point of view we see facilitating proteins and secondary products. The concentration of the secondaries are proportional, in some general way, to the concentration of the facilitating

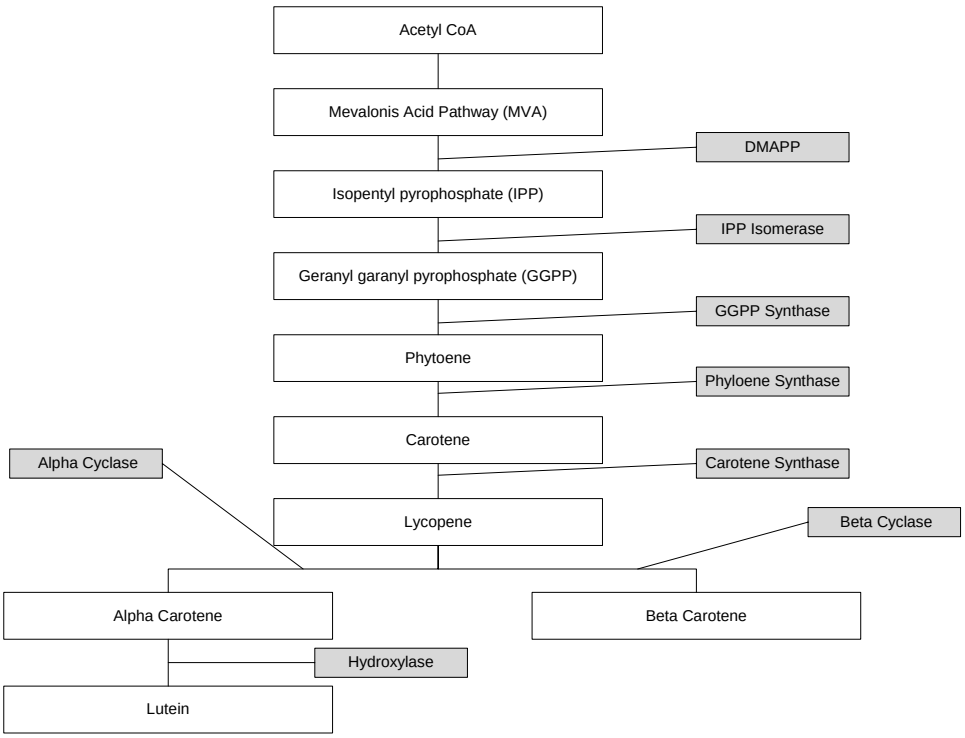
proteins. However we see there are many facilitating proteins which may make this a more complex analysis, however doable.

4.6.2 Carotenoid Pathway

We have shown the carotenoids as above. The carotenoid pathway is shown below. We have demonstrated this in general terms earlier but in this case below we see the specific details.



We show below the pathways and the facilitating enzymes. In many ways it appears identical to the anthocyanin pathway and the facilitating enzymes.



4.6.3 Flavonol Pathway

The flavonol pathway is identical to that of the anthocyanin. See Winkel-Shirley.

5 GENETIC INFLUENCE

In this section we briefly review the molecular genetics of a plant cell. We do not get into any significant details but merely review the elements which we can use letter in developing the mathematical models for plant regulation. As we have shown in the previous section, plant colors are the result of the expression of three types of secondary plant cell products; anthocyanins, flavones and carotenoids. We have focused mainly on the anthocyanins but have shown the details on all three. What we focused on is that the production of any one of these is a result of a specific pathway and that the production in that pathway is controlled by a set of enzymes. The enzymes are proteins produced within the cell. The proteins are the result of the expression of a set of genes.

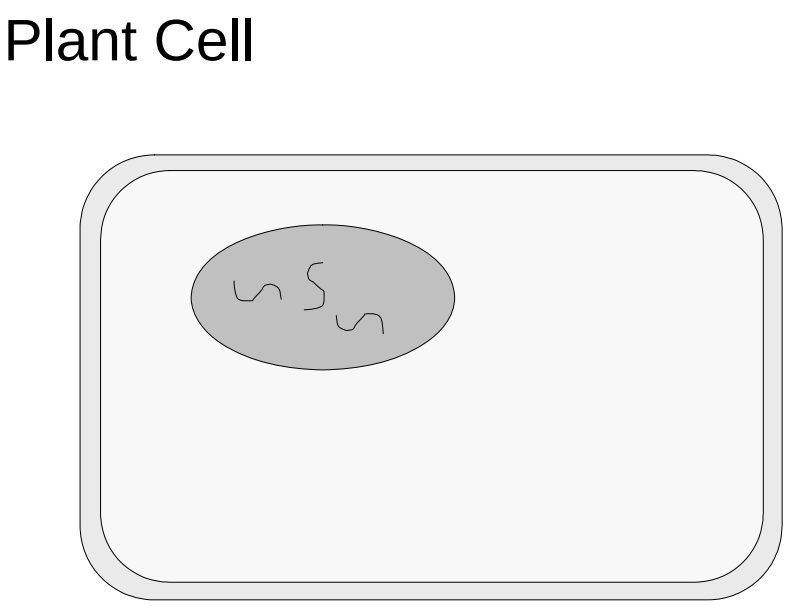
In this section we now by reviewing the current understanding of plant cell micro genetics show that the proteins are expressed by the normal process understood since Watson and Cricks seminal work and that there are factors which and activate their production, indeed enhance their production, or repress their production. These are the activators or repressor proteins. The activator and repressor proteins are in effect other genes expressing themselves. We will combine the last section with the results in this section to effect a dynamic system model for plant color generation in the next section.

What will be critical to understand here is that we just want to place the process of activators and repressors in context. We discuss in the next section what our overall design approach will be; that of an engineering model development and not a detailed understanding at the cell level. Frankly, we are not interested in the lower level detail, only gross modeling of cells, genes, and their proteins. They will become the inputs, outputs and control mechanisms of our design approach.

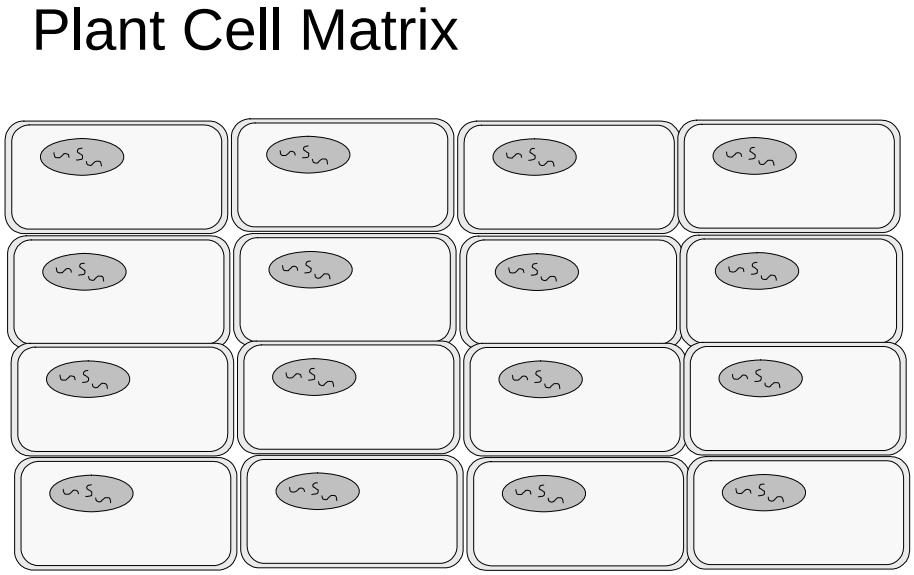
5.1 Plant Cells

Plant cells are a class of eukaryotic cells which are characterized primarily by have a rigid cell wall. In almost all other ways they are similar to animal cells. Plants generate all of the amino acids they need for protein generation unlike animal cells but other than that, for our purposes, they function very much the same. Thus as we develop a model for plants the model has no restrictions in its applications to animals as well.

The typical plant cell is shown below. The cell wall and the nucleus are depicted.



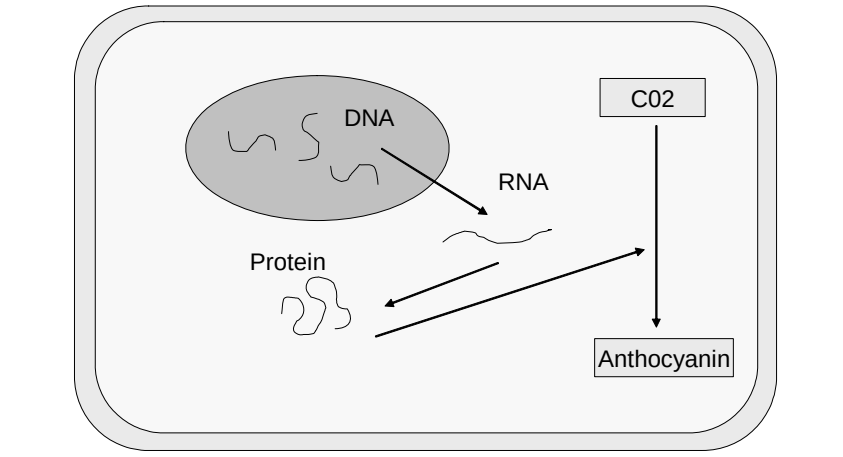
When we look at a collection of plant cells they appear as below. They are aligned and interconnect via various channels. Unlike animal cells plant cells have a much more rigid structure due to the cell wall however the general intercell signaling is identical.



Our interest will be to focus on both the intracell and intercell signaling and control of the pathways.

5.2 Plant DNA

Plant DNA processes are almost identical to those of animals. The graphic below summarizes the view we shall take. Each cell has DNA and the DNA uses a mRNA to create proteins. The proteins are then used in the management of the pathways to create the secondary products of the cell, in our case the anthocyanins.



For a single cell the model is quite straight forward. Gene expression causes RNA which causes Protein, which is enzyme in anthocyanin pathway generating the anthocyanin.

We do however want to stress certain issues. There are two extreme views of cells:

Micro/Time View: The micro view looks at a cell at each instant of time and considers what is happening. Is the cell generating a protein and a secondary and if so how and what is the sequence in which this process occurs. It is a focus on a single cell over some time period and we see many things happening.

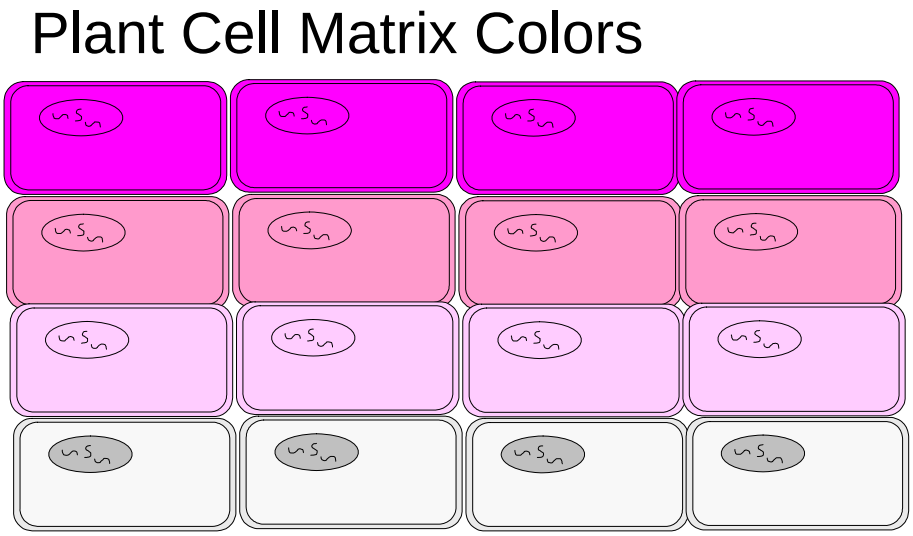
Ensemble View: In this case we look at the cell on average. Namely we say a cell can "on average" produce a protein and can then in turn produce a secondary.

These two views have analogs in mathematical analysis; they are the time averages versus the ensemble average. In mathematical statistics we have the concept of looking at a single cell and time averaging say the concentration of a certain secondary. We know how it is produced and thus over some time window we can look at the average of say pelargonidin and we than measure its average value. In contrast we can take a collection of similar cells and measure the pelargonidin in each cell and take that average. The latter is called the ensemble average. The equivalence of the two is called the Ergodic Theorem and was developed by Norbert Wiener[14]. The microbiologist typically focuses on the time view. We in this paper will focus on the ensemble view. The latter view will allow us to model, predict and control large collections of cells.

Now the figure below depicts a typical problem we want to understand. Consider an array of cells. Consider that they are arranged in ascending order up the petal of the flower, from base to

[14] See McGarty, Stochastic Systems and State Estimation.

outer edge. Consider now that at each vertical increase that the cells at the same level all have the same color yet at each level they have a differing shade of color. This implies that the anthocyanin concentrations are different at each level but identical at each cell within a level. We will assume we can understand a single cell from our discussions in the last section, if we understand the pathways and their enzyme controls. Now we ask how does one create a mathematical system model which can "explain" the color patterns we see below. This will be a critical question to answer.



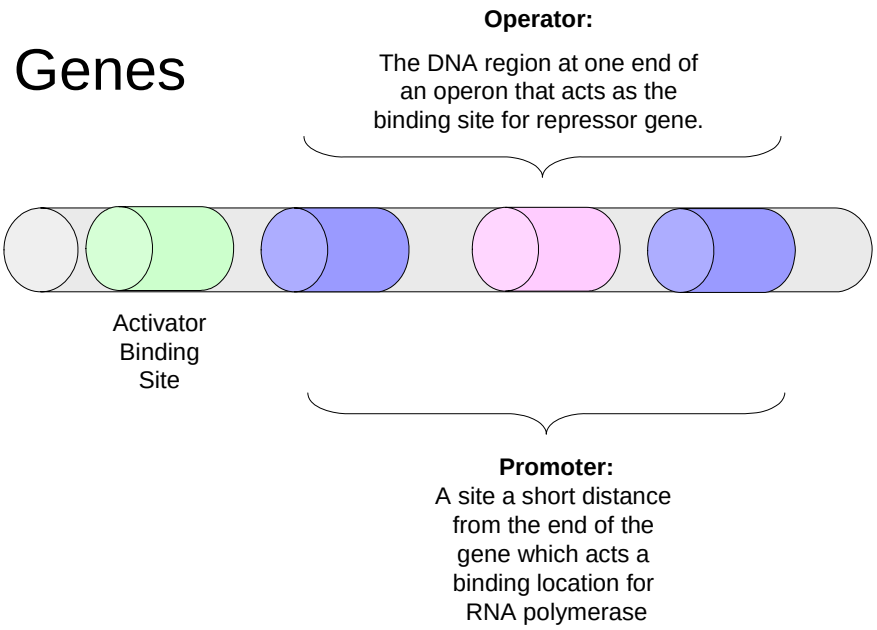
How do the cells communicate? Why does one cell generate more anthocyanin than other cells. Why is this not just random? What is the control mechanism?

Before we can answer this question we need to delve a bit deeper into the genetics of gene expression.

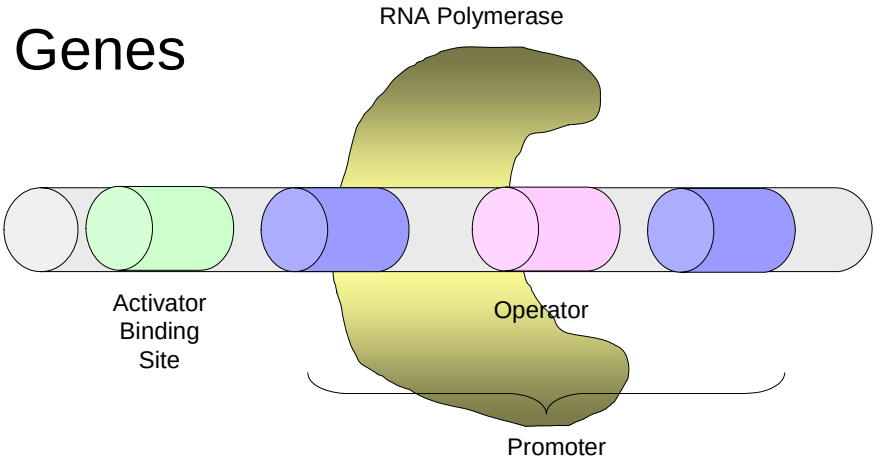
5.3 Plant Gene Processes

The processes in plant genes are generally identical to those in animal and thus human genes. The figure below shows a typical gene structure along with key sites. This structure shows the gene activator site which is where activator proteins can bind to start or enhance the expression of the gene. The operator sits and the overall promoter sequence are shown down from the activator site.[15]

[15] This is detailed in Watson et al. Also see Griffiths et al.



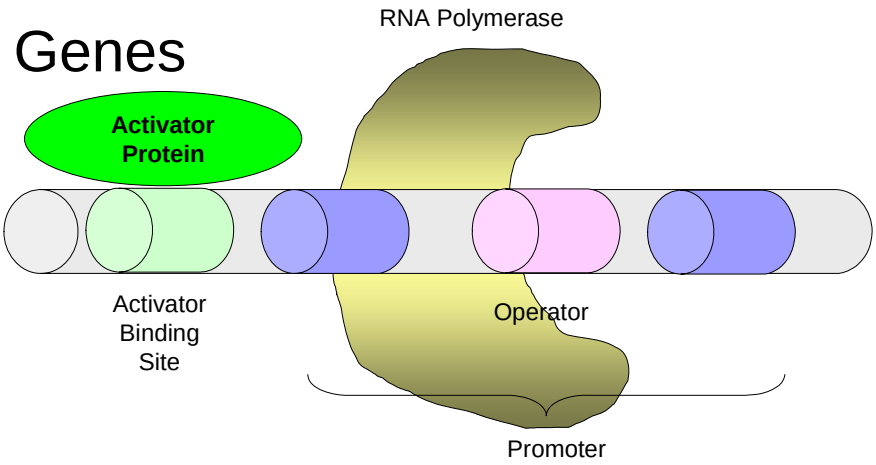
Genes express them selves with the assistance of RNA polymerase. The RNA polymerase is key in that it binds to the DNA and then opens it up to allow for the transcription creating the mRNA required for the translation process. In the figure below we show this process.



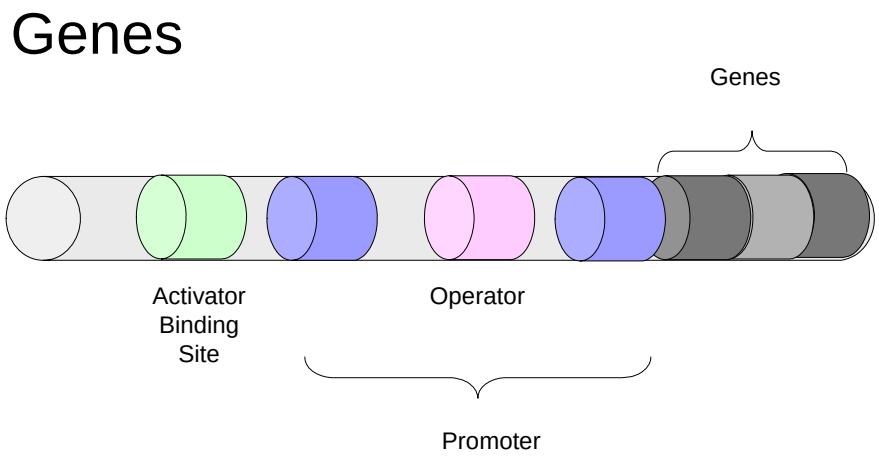
We will now focus on two actions which control the gene expression; activators and suppressors.

5.4 Activators

Activators are proteins which when attached to the gene assist in the expression of the gene. An activator is a protein resulting from another gene which can assist and facilitate the expression of a gene. Remember we want to look at the ensemble view, not the time view. Thus we assume that the RNA polymerase is continuously acting to produce proteins and that there is a continuous flow at some level of the activators. The cell process from the time view is shown below. An activator binds facilitates the RNA polymerase binding which in turn produces the mRNA and then in turn the proteins via the translation process.



If there is an activator then the gene can be readily expressed. The RNA polymerase then binds, creates the mRNA and this in turn produces the related protein. Activators stimulate this process. The Figure below depicts the location of the gene downstream from the activator and the promoter.



Now it is important to understand the activator from a time perspective and then from the ensemble perspective.

1. Activators are proteins generated by other genes in the cell.
2. Activators bind to the DNA and facilitate the production of the gene, which in turn produces another protein.
3. Activators can bind, release and then rebind. Each time the do that they produce another mRNA and that in turns produces another protein molecule.
4. From a time perspective, it is activator, produces gene reading, produces mRNA, produces protein.
5. From an ensemble perspective we have a concentration of activator proteins and then we get a concentration of result proteins.

This then leads to a simple model:

$$P_o = \text{Output Protein Concentration}$$
$$P_i = \text{Input Protein Concentration}$$
$$P_o = A_{o,i} P_i$$

But there is also a dynamic model which we can state; to some degree this model is a hybrid of the time and ensemble approach. The model states:

$$\frac{dP_o}{dt} = f\left(P_o(t), P_i(t), t\right)$$
$$P_o(0) = P_o^0$$
$$P_i(0) = P_i^0$$

Now we must remember that this simple two protein, two gene model is just a simplification. In reality we may have dozens of not hundreds of genes in this process. Now consider a simple linear model for this two gene system:

$$P_i(t) = P_i^0 \exp(-\lambda_i t)$$

$$\frac{dP_o(t)}{dt} = A_{o,i} P_i(t) + A_{o,o} P_o(t)$$

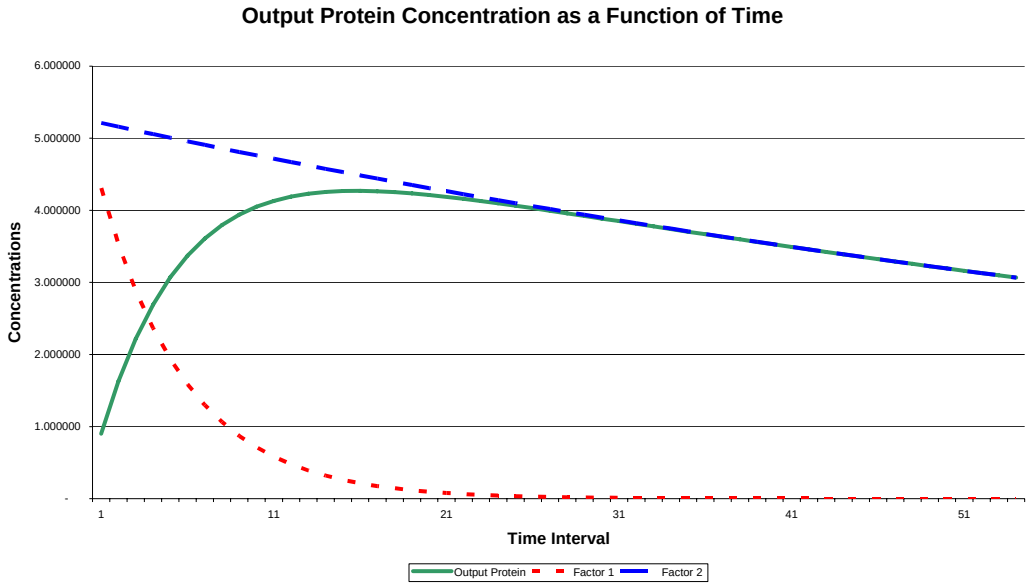
We can solve this differential equation. It is:

$$P_o(t) = k_{o,i} P_i(0) \left[\frac{\exp(-\lambda_i t) - \exp(-k_{o,o} t)}{\lambda_i - k_{o,o}} \right]$$

where;

$$A_{o,o} = -k_{o,o}$$
$$A_{o,i} = +k_{o,i}$$

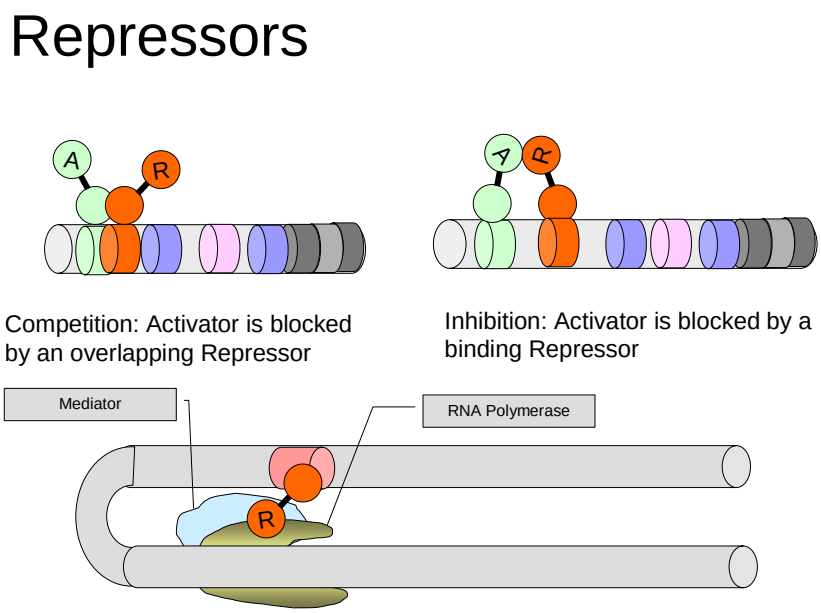
We have solved this for a simple example using constants of 0.01 and 0.2 respectively.



Note that the output protein concentration reaches a peak and then decays as per the driving protein. We will see this phenomenon again.

5.5 Repressors

In contrast to activators we also have genes which are suppressors. Three methods of suppressor action are shown below. A suppressor does the opposite of an activator. It suppresses the expression of a gene. The same logic will follow the repressor as was with activators. We again also want to view this from an ensemble perspective.



As we did with the activator, we see a repressor stops the generation of the protein. This it is nothing more than a negative driver to protein generation.

5.6 Summary of Actions

We can now summarize what we have presented here:

1. Color is the result of anthocyanin production.
2. Anthocyanin production is a product of a specific pathway.
3. Pathways are mediated by enzymes, which are proteins generated by genes in the cell.
4. Proteins are generated by genes.
5. Gene activation is modulated by activator proteins and suppressor proteins.
6. Activator and suppressor proteins are generated by other genes.
7. One can model this overall process by a linked set of equations, both of a time varying nature and an ensemble, average steady state, nature.
8. An overall state model can be developed for the genetic control of color in plants.

We can now take this set of conclusions and use it to construct the state model.

6 EXPRESSION ANALYSIS AND IMPLICATIONS

In this section we develop a systems approach to the problem of color analysis and synthesis. This work is based upon the recent work of Szallasi and others. However this also builds upon the work in McGarty (1971) which focused a systems approach to the overall identification problem.

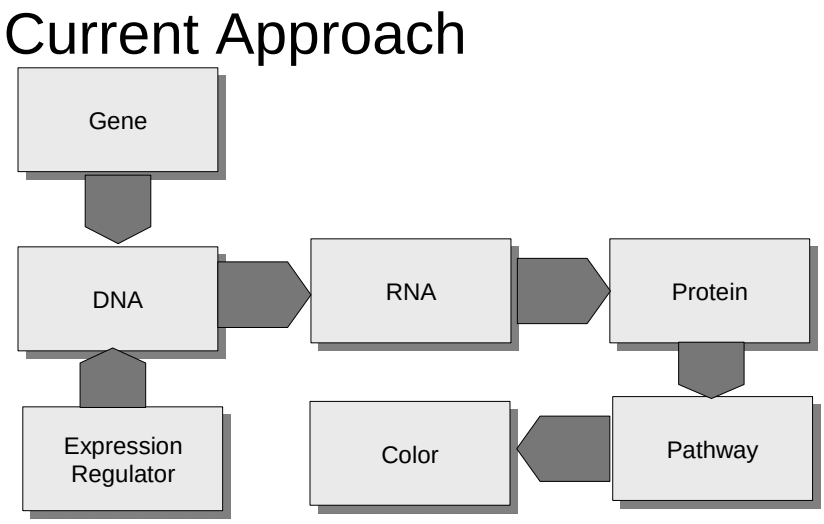
6.1 Approach: Engineering versus Science

The approach we take in this paper is an engineering approach rather than a biological approach.[16] Our interest is in developing a model or sets of models which allow us by a verifiable means to show how the genes react and interact to produce the plant colors. We can compare this to the engineering approach to circuit design of transistor circuits versus the science of understanding the semiconductor from the point of view of detailed quantum mechanical models. The biologist in our approach is akin to the physicists and engineers who approach the cell from the bottom up, trying to understand all of the intricate processes and steps that lead at the micro level to the developments we look at herein. In our approach it is akin to the engineer knowing that there is some function inside the semiconductors which may clearly be important but the engineer's interest is in designing and analyzing the transistor as a circuit element. Thus for an engineer, if we increase a current here we get a decrease or an increase at some other point. The engineer creates a world view of a macro set of processes and models the details of the biologists in our case with a few set of equations which show the results of increases and decreases. This model must then be valid table and verifiable. One must be able to make measurements to show that the processes predicted indeed occur, to a reasonable degree of accuracy. Then one can analyze a genetic circuit and then in addition one can design a genetic circuit. We then can understand where the colors come from and possibly engineer the genes to develop and deliver on colors we desire.

6.2 A Control Paradigm

The basic control paradigm is contained in the following Figure. The expression regulator may be an activator or suppressor. It may be a result of a gene expression in the cell itself or quite possibly as we shall discuss fed through from another cell. There are many of these regulatory cycles and they are all interconnected. This basic paradigm is one of hundreds or thousands of such interconnected flows.

[16] There has been a significant set of development recently in analyzing genetic data from a systems perspective. In this paper we have taken such an approach. The recent work by such authors as Perkins et al, Vohradsky, Hatzimanikatis et al, and the recent book by Szallasi are seminal. However, there is an issue here also or world view and what does one really want from the analysis. The bench scientists looks to understand all the details of the underlying processes. The engineer seeks to understand enough to model the process and to do so with a reasonable degree of accuracy but the ultimate goal for the engineer is control of the process and generation of new processes.

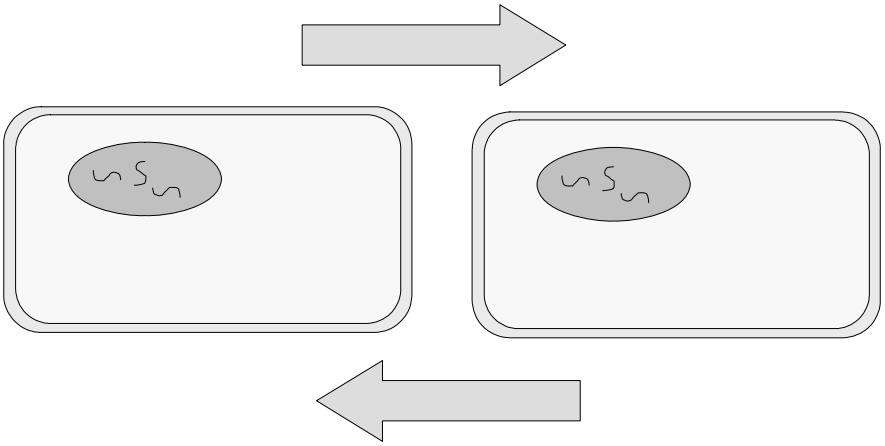


In developing our models we will use this construct. However, we can frequently focus on natural clusters of related genes. They may be a dozen or more such related genes in each cluster and possibly hundred of such clusters. Although cells and their proteins may affect all other cells, only a few of the genes regulated have a significant level of regulation. The low levels of "regulation" we shall consider just as noise.

6.3 Cell Signalling: Intra and Inter Cell

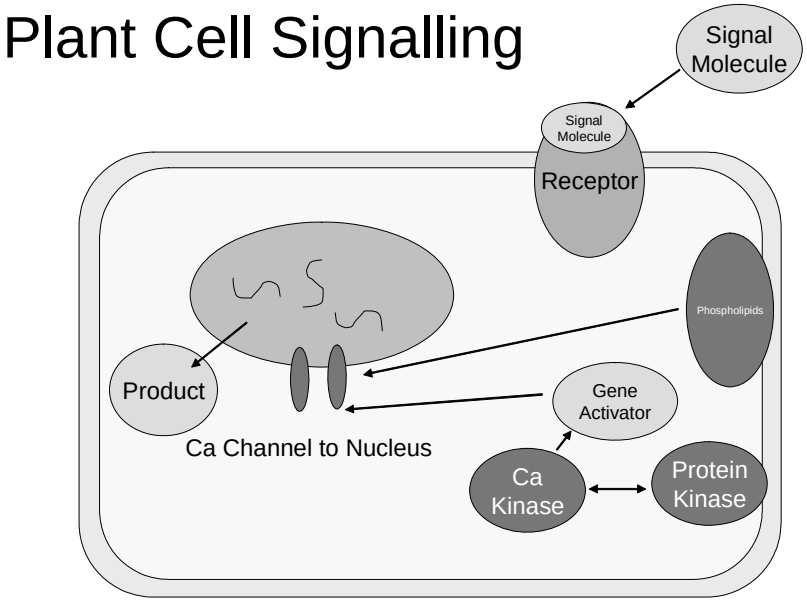
We must also better understand the inter cell signaling. Although we include it in this paper we have not as of yet produced a robust enough model for this set of processes. The Figure below presents the essence of the problem.

Plant Inter Cell Communications



What do the cells use to communicate and how. What are the elements? Proteins?

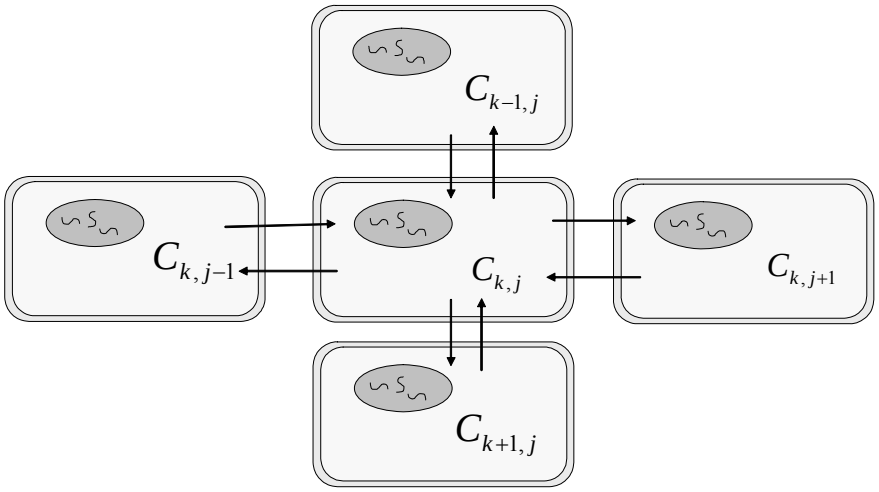
Key to intercell signaling will be the receptor elements which control the flow of the controlling elements. This means that we must be able to introduce certain additional elements in the model which at this time are not yet fully developed. The Figure below highlights the issues of concern in this area.



Dey, Plant Biochemistry, p. 373, 1997 Academic

Then we must be able to establish a full network view of the signally processes. There has been considerable work looking at this from a meta perspective as some neural network. However the approach does not yet provide an adequate refection of a gene by gene analysis.

Plant Inter Cell Communications



What do the cells use to communicate and how. What are the elements? Proteins?

6.4 A Model for Secondary Production

We can now commence the structuring of the model. Let us assume that x_k is a concentration of a plant cellular protein. Let us assume k goes from 1 to n. Then we have:

$$\frac{dx_i(t)}{dt} = f_i(x(t),u(t),t) + n(t)$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ . \\ . \\ . \\ x_n(t) \end{bmatrix}$$

and

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ . \\ . \\ . \\ u(_m(t) \end{bmatrix}$$

Now we also assume that there is a model for the transformation of a protein into a secondary such as an anthocyanin. The we have, where z_k is the secondary concentration of the k^{th} element;

$$z_k = g(x(t),t) + w(t)$$

Now we can obtain a color as we had describe before as a weighted combination of the secondary concentrations.

6.4.1 Linear Models

We can now provide a linear model:

$$\frac{dx(t)}{dt} = A(t)x(t) + B(t)u(t) + n(t)$$

where

$$A(t) = \begin{bmatrix} A_{11}(t)....A_{1,n}(t) \\ \\ A_{n1}(t)....A_{nn(t)} \end{bmatrix}$$

and B is an n by q matrix with u a q dimensional vector with entries akin to A.

We also have:

$$z(t) = C(t)x(t) + w(t)$$

where C is an m by n matrix also akin to A.

We will discuss latter how we estimate the entries in this set of equations.

6.4.2 Second Order Models

We now want to consider a second order model, one which contains product elements. This model has the ability to consider the characteristics of enzyme functionality found in the Michaelis-Menten theory and its enhancements.[17] It also allows for consideration of the complexities of the Voltera model establishing multiple stable points in a phase space.[18] Finally having this model we can demonstrates the existence of dramatic color change and patterns.[19]

Now consider the second order model. We base this upon the models in McGarty (p. 241):

[17] See Rubinow Chapter 2 for detailed discussions of the enzyme models.

[18] See the books by Cunningham and also by Andronov. The discussion of a phase plane analysis is key to understanding the development of abrupt changes in coloration.

[19] See Murray. In this book there is a set of detailed analyses for establishing the complexities we see in developmental biology. This is especially true regarding the books development of patterns in animals. The same follows with plants. The book however does not approach it from the genetic basis which we do herein.

$$\frac{dx(t)}{dt} = A(t)x(t) + \frac{1}{2}\sum_{i=1}^{n}\gamma_i x^T(t)\Lambda(t)x(t)$$

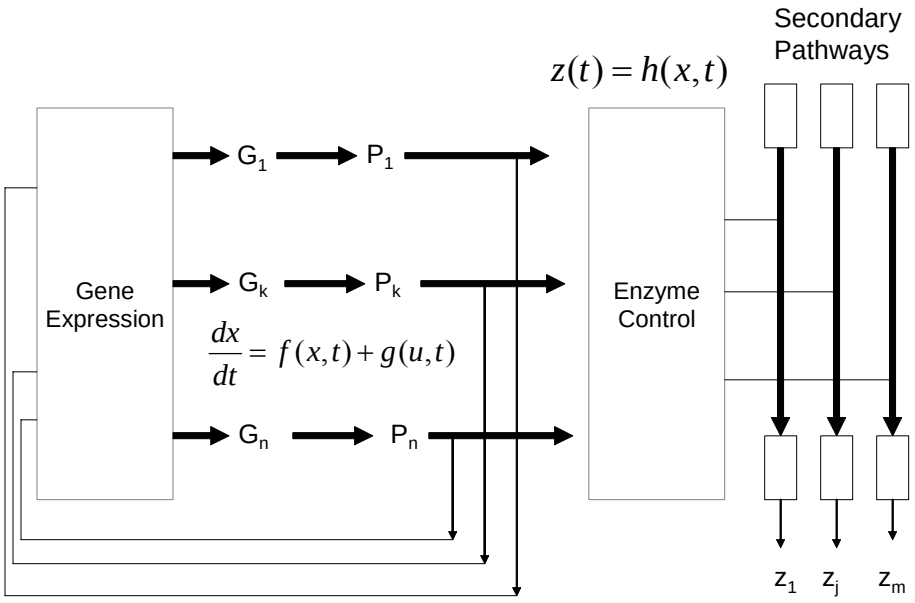
where

$$\gamma_i = \begin{bmatrix} 0 & 1 \\ 0 & .. \\ 1 & j \\ 0 & .. \\ 0 & .. \\ 0 & n \end{bmatrix}$$

and:

$$\Lambda(t) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} \cdots\cdots\cdots \frac{\partial f_1}{\partial x_n} \\ \\ \\ \frac{\partial f_n}{\partial x_1} \cdots\cdots\cdots \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

This model allows us to model the process as enzyme like in their functions. The totality is shown below:



We shall pursue the implications of this in the following. However it is important to make one final observation. In this model, which we shall expand in the next section, we see that we can change the secondary products in two ways; first we can find a u(t) process, a driving process, which in many ways is a drug driven process. That is we can use this approach to determine what drug, or sets of drugs can be used to obtain a desired result. Second, we can genetically engineer the process by changing the A matrix, by adding or deleting genes! Again if we have a desired result, we can attain that result by modifying the A matrix by the addition or deletion of elements, namely genes. This we believe is a key observation.

6.5 Pattern Initiation

The next question we pose is how can we obtain the patterns we see in the flower? We approach this by relying on work by Alan Turing in 1952.[20] The results herein rely heavily on what is also presented in Murray.

Let us pose the problem as follows. Assume that there are two secondary elements (anthocyanins for example), z_1 and z_2, and if one is greater than the other then the flower is Color 1 or Color 2 respectively. We further assume that we have a spatial-temporal domain and that these two secondary elements can compete in some manner with each other. We further assume that the process is a continuous process across some spatial domain, {x,y,z}, as well as in time. Namely we assume that the cells are so small that they are "points" in color space, and blend together.

[20] See A. Turing, The Chemical Basis of Morphogenesis, Phil Trans Royal Soc London B337 pp 37-72, 1959. This is also the same Turing who conceived of the Turing Machine, the first mathematically complete structure for the computer.

This assumption of "point" like behavior for a cell is akin to our assumption of ensemble averages. Now let us consider a cell as shown below. There are the two secondary elements in the cell. We now consider the following:

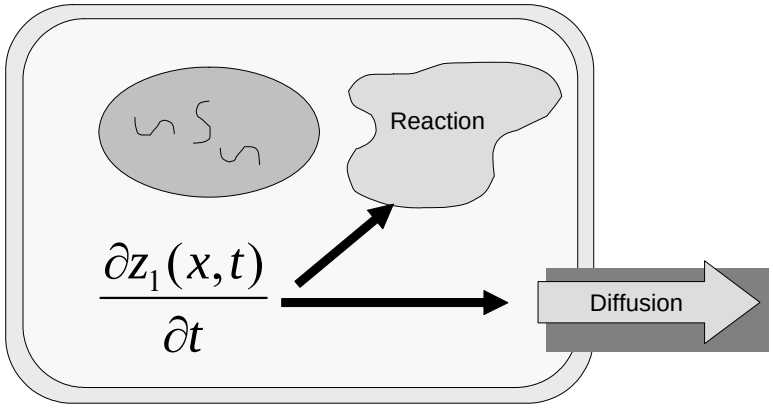
1. Consider a secondary and consider its concentration; that concentration is a function of time and space. In the previous part of this section we considered only time. Now we consider space as well.
2. Consider now that the secondary has a change over some time period. We then ask where does that change in concentration go. The answer is simple;
3. First, the concentration flows from the point, we also mean cell, to other cells or points in a diffusion like manner. We understand that from basic physics and thus we include a diffusion term.
4. Second, the concentration goes into combining with other reactions within the cell, or even at the boundary of the cell. This we shall call the reaction kinetics losses. The concentration of the secondary can increase or decrease depending on what the chemical reactions are within the cell.[21]
5. The result is that we have a model which states that the rate of change of concentration equals the changes due to reaction kinetics within the cell plus the flow to the outside of the cell.
6. We must continue to remember that a point and a cell is the nexus in this argument.

The following Figure depicts this model.

[21] See Atkins, Physical Chemistry pp 778-830 for a good overview of these issues. Also we have developed models along this line elsewhere in this paper. Note we have developed the models from the cell genetics upward. This model is from the chemical concentrations downward. The analysis herein complements the work of Murray by connecting the genetic elements and internal pathways with the work of Turing which is a macro level ensemble approach.

[Rate of change of concentration] = [Reaction Kinetics] + [Diffusion]

$z_1 = \text{concentration of protein or secondary } Z_1 \text{ in the cell}$



$$\frac{\partial z_1(x,t)}{\partial t} = F_1(z_1, z_2, x, t) + \lambda_1 \nabla^2 z_1(x,t)$$

We can now state the fundamental equation as follows:

[Rate of change of concentration] = [Reaction Kinetics] + [Diffusion]

or mathematically we have:

$$\frac{\partial z_1(x,t)}{\partial t} = F_1(z_1, z_2, x, t) + \lambda_1 \nabla^2 z_1(x,t)$$

$$\frac{\partial z_2(x,t)}{\partial t} = F_2(z_1, z_2, x, t) + \lambda_2 \nabla^2 z_2(x,t)$$

If we exclude the diffusion element we have a model which in many ways is similar to what we have developed herein. However, the reaction kinetics portion is not related to genes but only to their products. We have included the basic genetic structure as reflected by the expression of their individual proteins. We know that the proteins are key to the pathway reactions. This is a complex enzyme reaction analysis.

Murray presents several types of reaction kinetics models and we summarize them here. They are as follows:

Case 1: Schnakenberg Reaction

The Schnakenberg model is one of the simpler models. It reflects basic reaction kinetics models as are generally accepted by chemists and are found in enzyme kinetics.

$$F_1 = k_1 - k_2 z_1 + k_3 z_1^2 z_2$$

and

$$F_2 = k_4 - k_3 z_1^2 z_2$$

Case 2: Activator-Inhibitor

The activator-inhibitor model is in many ways the better model for what we see at the gene level. It is strongly akin to the activator-suppressor reactions we have discussed before.

$$F_1 = k_1 - k_2 z_1 + \frac{k_3 z_1^2}{z_2}$$

and

$$F_2 = k_4 z_1^2 - k_5 z_2$$

Case 3: Substrate-Inhibitor Reaction

The substrate-inhibitor model reflects a substrate implementation as in enzyme reactions and it includes a dampening inhibitor element. Recall that if the rate of change in time of the secondary concentration is negative we have a decaying amount of the secondary. In this model the reaction kinetics are strongly decreasing.

$$F_1 = k_1 - k_2 z_1 - H(z_1, z_2)$$

and

$$F_2 = k_3 - k_4 z_2 - H(z_1, z_2)$$

where

$$H(z_1, z_2) = \frac{k_5 z_1 z_2}{k_6 + k_7 z_1 + k_8 z_1^2}$$

Case 4: Activator-Inhibitor with Activator Inhibition

This is akin to the activator-inhibitor.

$$F_1 = k_1 - k_2 z_1 + k_3 \frac{z_1^2}{z_2}$$

and

$$F_2 = k_4 z_1^2 - k_5 z_2$$

Murray proceeds to solve the problem for several simple examples. We review and summarize one of them.

Consider the following model:

$$\frac{\partial z_1(x,t)}{\partial t} = \gamma(a - z_1(x,t) + z_1^2(x,t) z_2(x,t)) + \frac{\partial^2 z_1(x,t)}{\partial t^2}$$

and

$$\frac{\partial z_2(x,t)}{\partial t} = \gamma(b - z_1^2(x,t) z_2(x,t)) + d\frac{\partial^2 z_2(x,t)}{\partial t^2}$$

Now Murray determines a "solution" of the form:

$$z_1(x,t) \approx z_1(x_0,t_0) + \varepsilon \exp\left[\lambda\left(\frac{\pi^2}{p^2}\right)t\right]\cos\left(\frac{\pi x}{p}\right)$$

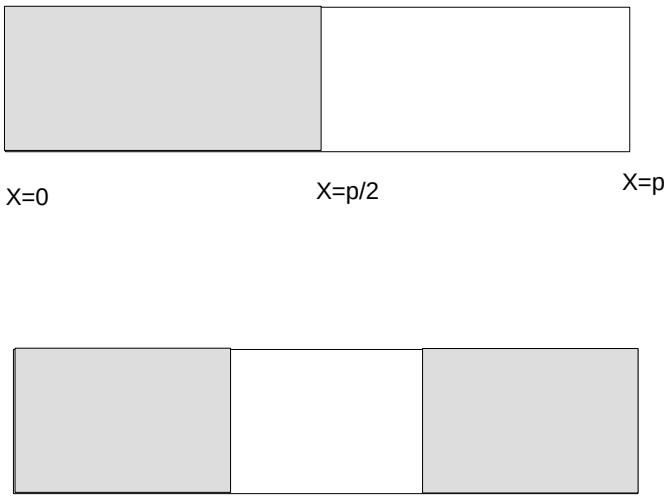
where p is the domain, namely we seek a solution over:

$$\text{x } \varepsilon \text{ the interval } (0, \text{p})$$

Now remember that:

$$\text{If } z_1 \geq z_2 \text{ then :}$$
$$\text{Color} = \text{Color}_1$$
$$\text{else}$$
$$\text{Color} = \text{Color}_2$$

This model of Murray is one where either one secondary dominates or the other, there is no blending. Thus Murray shows the result to be:



See Murray, pp: 390-392

Note that we have the ability to explain the banding and patterns. Murray continues this for many animal pattern developments.

The important factor to observe here is the use of what is akin to the Fokker-Planck equation. It is a diffusion equation with a system effected term. We have used this extensively in various stochastic processes and it is also used in the classic Schrödinger equation for quantum mechanics. McGarty had developed this in 1971 for the use in estimating nonlinear dynamic systems parameters. We have a great deal of experience in solving this equation.

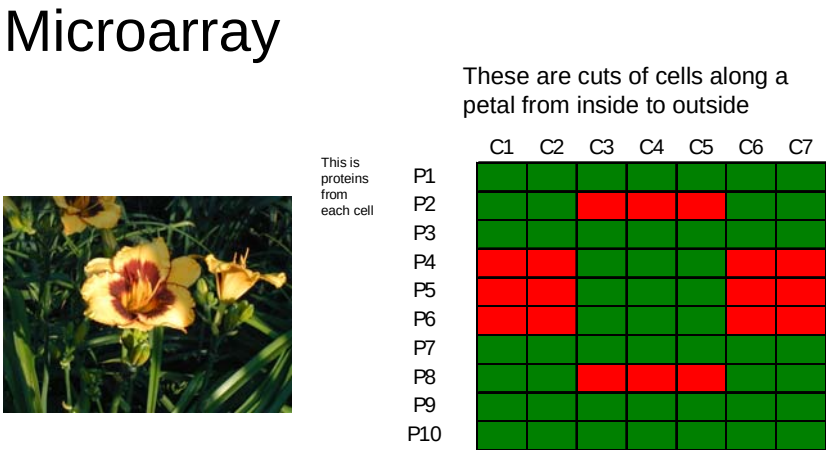
The major observation to be made here is in this subsection we have made the connection between the micro and the macro. We have developed a bottom up analysis from the gene level and then we have used a top down analysis for color and pattern development, and then we have the nexus at this point. We relate a point in the Turing worlds to a cell in our world. When we do that the two worlds combine.

6.6 Methods for Verification and Validation

The final issue is how do we determine the values for the model we have developed. We make a simple argument here for the use of microarray technology. Suppose we can take a single flower and make a lateral cut, if that is the direction we seek to analyze. The we segment that cut into many slices and then we place the slices across a microarry in the top row,

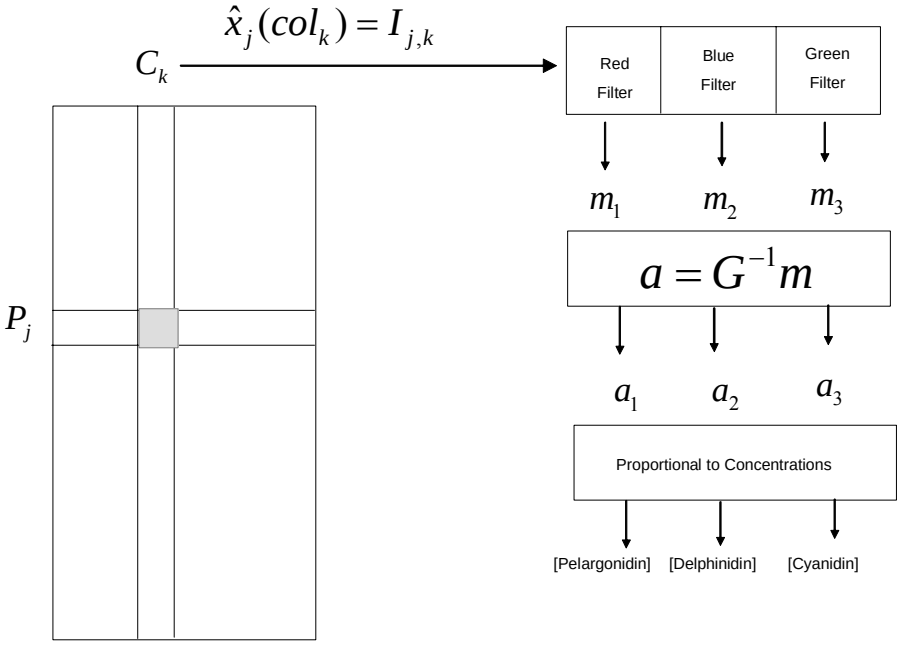
The we assume we know the cDNA which is important and then we go protein by protein or whatever tag we seek. From that we can determine the spatial expression matrix of the system. We can then use that spatial data base to solve the inversion problem to determine the system matrix A. We can also verify the coefficients in the Turing model for flower coloring.

We briefly show this below.



Question: Why does this eyezone appear here? Use microarray analysis and determined constants and dynamics. Do we have model of this in physical world?

We have separately developed the algorithms to perform these tasks and will publish them separately. We show this below:



Simply stated, we have a combination of the system identification problem and the observability problem.[22] We do the following:

1. First we measure the color,
2. Then we can infer the secondaries which have produced the color
3. We then use the micro array analysis to see what genes are expressing what at each slice where we know the output, namely the color or the implied secondaries.
4. Knowing the model we have for the overall process, we now know genes and outputs we can apply the inversion or identification analysis to determine the constants for the dynamic intra cellular model we have developed herein.

7 A SET OF EXAMPLES

Let us consider a simple example of a three color element and seven gene system. We assume that we have three genes which effect proteins which in turn effect secondary products, namely anthocyanidins. We also assume we have four other genes which are activators, or suppressors, of these genes. We then develop a simple models and from that model determine the color of the resulting combination using a Red, Green, Blue model for colorimetry. This will be example 1.

There are three problems:

1. **Analysis:** In the analysis problem we assume we know the expression dynamics and the secondary production model. We then ask given those two what color do we get. We develop this in detail.

2. **Identification:** The identification problem is one in which we know the secondary processes, we have many color samples and we know the protein concentrations which yield each color. Then we ask how do we determine the A matrix for the gene expression.

3. **Design:** This problem is of significant interest. We seek a desired output state, color in our example. We know the gene expression dynamics and the secondary model. We then ask how do we modify the gene expression model to obtain the desired output.

7.1 Example 1- Known Expression Model, Known Secondary Model, Unknown Color

This example assume we know the model for the gene expression and we know the model for the transformation and control of the secondary products. We are seeking to determine the output color.

Le x be a gene expression in terms of mRNA or protein concentrations. Then we have:

[22] See McGarty, Stochastic Systems and State Estimation.

(0.1) $$\frac{dx(t)}{dt} = Ax(t) + u(t)$$

where x is a n X 1 vector. We assume n = 7 in this case.

We choose A such that –A equals:

	1	2	3	4	5	6	7
1	5.00	(0.20)	(0.30)	0.00	(0.50)	(0.20)	(0.10)
2	(0.20)	3.00	0.00	(0.30)	0.00	(0.10)	(0.25)
3	(0.30)	0.00	6.00	0.10	(0.40)	(0.50)	(1.00)
4	0.00	(0.30)	(0.10)	1.00	0.00	(0.20)	(1.00)
5	(0.50)	0.00	(0.40)	0.00	2.00	0.00	(1.00)
6	(0.20)	(0.10)	(0.50)	(0.20)	0.00	4.00	(1.00)
7	(0.10)	(0.25)	(1.00)	(1.00)	(1.00)	(1.00)	4.00

In the above model we assume that the diagonals control the particular protein concentrations from the perspective of reaching their steady state concentrations. The negative off diagonals are indicative of activators to the related gene.

Furthermore we chosen u, the steady state constants, as follows:

5.00
9.00
6.00
2.00
4.00
8.00
16.00

Then the steady state solution to the protein concentrations are as follows:

(0.2) $$x = A^{-1}u$$

For this particular case we have for x:

3.48
7.21
5.08
23.05
12.24
8.32
16.71

Now we can relate the proteins to the pathway dynamics in a steady state format. Namely, if the following holds:

$$(0.3) \qquad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \text{Concentration of Enzyme effecting Pelargonidin} \\ \text{Concentration of Enzyme effecting Delphinidin} \\ \text{Concentration of Enzyme effecting Cyanidin} \end{bmatrix}$$

we can then state the concentrations of each of these anthocyanidins as:

$$(0.4) \qquad z = Cx$$

where:

$$(0.5) \qquad \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \text{Concentration of Pelargonidin} \\ \text{Concentration of Delphinidin} \\ \text{Concentration of Cyanidin} \end{bmatrix}$$

and we have x being only the first three protein concentrations. We have chosen C as:

3	0	0
0	4	0
0	0	2

This yields a concentration of the three anthocyanidins as:

10.45
28.83
10.17

Now we relate concentrations to color. We use the RGB format or red, green and blue. We use the 256 bit format as is common in video monitor displays as a common format. This follows the CIE standard[23]. From this standard we have for each of the anthocyanidins the following mix.

[23] See Levi, Applied Optics, Wiley, 1968, pp. 10-33.

	Red	*Green*	*Blue*
Pelargonidin	255	110	0
Delphinidin	150	125	245
Cyanidin	255	0	240

Now we can determine the RBG combination as a result of the three concentrations as follows:

(0.6)

$$R = \frac{\sum_{i=1}^{3} z_{iR} R_i}{\sum_{i=1}^{3} z_{iR}}$$

and

$$G = \frac{\sum_{i=1}^{3} z_{iG} G_i}{\sum_{i=1}^{3} z_{iG}}$$

and

$$B = \frac{\sum_{i=1}^{3} z_{iB} B_i}{\sum_{i=1}^{3} z_{iB}}$$

where the R. G .B elements in each are from the table above. The result is:

	Red	*Green*	*Blue*
Pelargonidin	71	27	-
Delphinidin	68	95	182
Cyanidin	69	-	62
Total	208	121	244

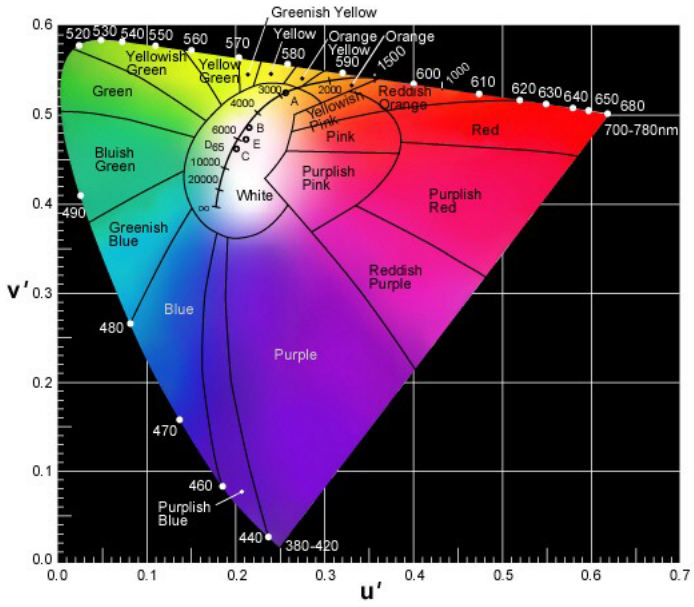
In this case we now have the output color. We also know from the CIE model that we can map the RGB coordinates into a perceived wavelength, namely color, namely:

(0.7)

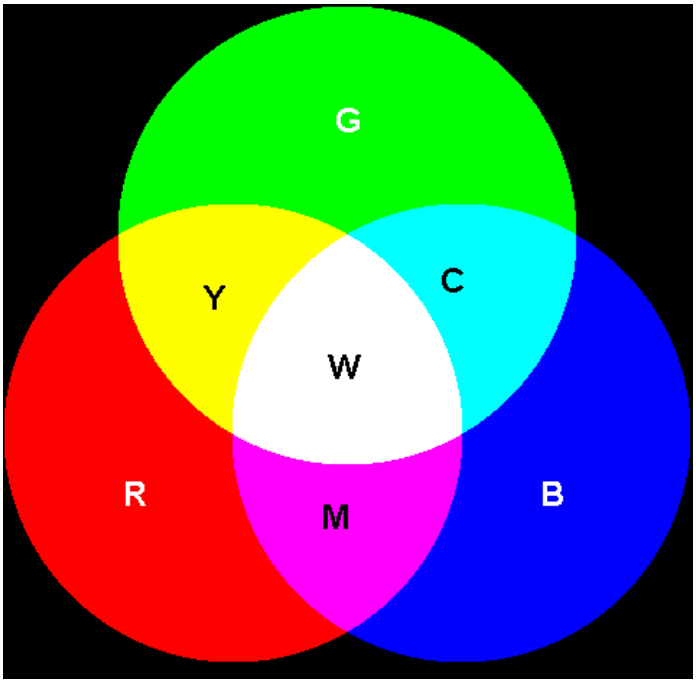
$$\lambda = h\left(R, G.B\right)(nm)$$

The CIE chart is shown below[24]:

[24] See Durrett, H. J., Color, Academic Press, 1987. also see http://home.wanadoo.nl/paulschils/06.00.html



The RGB chart is as follows:

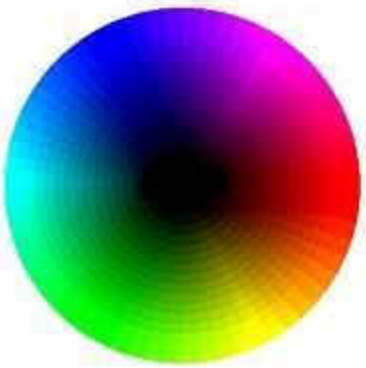


The Additive Primary Colors:

Additive colors are light based. Additive color mixtures are always lighter than any of the individual components. If the light of these primary colors is added the sensation of white light is produced as shown below.

The Subtractive Primary Colors.

Subtractive colors are pigments. Subtractive color mixtures are always darker than the components separately. If these Subtractive primary colors are mixed together in about equal amounts they will produce a neutral Hue. They can be mixed together in varying amounts to match almost any Hue. See below:



7.2 Example 2: Unknown Expression Model, Known Secondary Model, Known Colors from Many

The second problem is as follows. Let us assume we use the same model for the expression that we have shown above. Let us assume we have sample from N plants sections, each of a different color. Let us assume we perform a microarray analysis on each of these plant segments and we use cDNA from each of the seven genes. We then measure the expression intensity for each gene

and for each color. The question is, can we determine a good estimate for A given this data set. This is also called the Inversion Problem or the System Identification problem.

In the following Table we have varied the a11 and a22 entries respectively to yield various colors. The colors are in nm wavelength and we have used the formula to convert from RGB to nm which is:

(0.8) $$\lambda(nm) = 460(nm)R(\%) + 530(nm)G(\%) + 650B(\%)$$

This yields the following:

550	1	2	3	4	5	6	7	8	9
1	540	550	554	557	558	559	560	561	561
2	529	542	548	551	553	555	556	557	558
3	523	536	543	547	550	552	553	554	555
4	519	532	540	544	547	549	551	552	553
5	516	529	537	541	544	547	548	550	551
6	514	527	534	539	543	545	547	548	549
7	512	525	533	537	541	543	545	547	548
8	511	524	531	536	540	542	544	546	547
9	509	522	530	535	538	541	543	545	546

We will use the above formula in the following cases.

7.2.1 Case 1: Fixed Controlling Gene Variable Suppressor/Activator

Let us consider a case where we have a set of secondary controlling genes which do not effect other controlling genes and that there are secondary suppressor/activator genes which do not effect any secondary path but do suppress or activate the controlling genes. We show the result in the following sample from a quantized microarray analysis. The (row, column) values are effectively the concentrations of the first three controlling enzymes plus four suppressor/activator genes. Let us assume we have measured the following data, from 11 color samples, and we have done this in a microarry analysis.

				Color (nm)							
	Red					Green					Blue
Gene	400	425	450	475	500	525	550	575	600	625	650
1 (Pelargonidin)	1	1	1	2	2	2	3	3	4	4	5
2 (Delphinidin)	2	2	2	2	3	3	3	2	2	1	1
3 (Cyanidin)	1	2	3	4	5	6	7	8	9	10	11
4 (Activator 1)	4	4	4	3	3	3	2	2	2	1	1
5 (Activator 2)	5	5	5	5	5	5	9	9	9	9	9
6 (Suppressor 1)	2	2	2	2	1	1	1	-	-	-	-
7 (Suppressor 2)	1	1	1	1	1	1	2	2	2	2	2

The question then is; can we go from the microarray data to estimates of A? That is knowing the secondary pathways and their effect on color and then knowing the colors and their relative

expression of the relevant genes, can we estimate what A gives the best match? This is a bit different than the typical identification problem. Here we have 11 measurements and in each case the A matrix is different! But the difference is analytically constrained! The output constraint manifold is a surface mapped by the colors.

Consider the following simple problem. Two genes express color and the third gene determines suppressor or activator status on the other two. Namely Gene 1 and 2 are the ones in the pathway for color and gene 3 can suppress or activate either of the two operative genes. Here below we have chosen activation.

	1	2	3
1	5.00	0.00	(0.30)
2	0.00	3.00	(0.20)
3	0.00	0.00	6.00

Now we observe the following:

1. The diagonals are the same in every plant color.

2. The activator degree may be different from plant color to plant color. That is color is determined by activation/suppression alone and that there is some exogenous mechanism which drive this process. We will get to that latter.

3. All other entries are zero.

4. Three of the unknowns are constant in every sample. Two of the unknowns are different in every sample, but are constrained to yield a measured color.

The inverse matrix is:

	1	2	3
1	0.20	0.00	0.01
2	0.00	0.33	0.01
	0.00	0.00	0.17

Note that gene 1 will be affected by its constant and gene 2 by its constant and both will have a gene 3 impact as an activator. Namely we have:

$$x_1 = a^{11}c_1 + a^{13}c_3$$
$$\textit{and}$$
$$x_2 = a^{22}c_2 + a^{23}c_3 \tag{0.9}$$
$$\textit{and}$$
$$x_3 = a^{33}c_3$$

Recall the diagonals for the genes expression remain constant for all colors and the cross elements are the only ones changing, except perhaps for the suppressor/activator gene.

From the above analysis in Example 1 we can see that:

$$\lambda(nm) = \sigma_1 c_1 + \sigma_2 c_2 + \sigma_3 c_3$$

(0.10) *or*

$$\lambda(nm) = \kappa_1 a^{11} c_1 + \kappa_2 a^{22} c_2 + \left(\kappa_{13} a^{13} + \kappa_{23} a^{23}\right) c_3$$

where the constants are such that the first two are the same for all samples and the third changes sample by sample. We have normalized the data to eliminate the denominator normalization performed earlier. We also know all the constants except the a values. We know that a^{11} and a^{22} are the same and that a^{13}, a^{23} vary from color to color.

Let us now consider a simple example. As before we have the experiment with the resulting microarry results:

					Color (nm)						
	Red					***Green***					***Blue***
Gene	400	425	450	475	500	525	550	575	600	625	650
1	1	1	1	2	2	2	2	3	3	3	3
2	2	2	2	2	3	3	3	2	2	1	1
3	3	3	2	4	4	5	6	7	7	8	9

The entries represent gene expression values. Clearly since the diagonals are fixed we can take the concentrations of the two genes as constant. It is the third gene expression which is the causative factor. This we can take the average of the expression for the two as a baseline. Also gene 3 is increasing in its expression. Gene 3 also has no dependence on gene 1 and gene 2. This is by construct. Thus we can determine the dependence of gene 3 expression as some statistically reliable function of the color. The greater the wavelength the greater the gene expression.

We thus make the observations:

Gene 1 is fixed and variable. The fixed is the diagonal or self control element and the variable is the suppressor/activator element. Thus if we use a simple linear regression we find the intercept the diagonal and the slope the changing suppressor/activator element.

Gene 2 is identical to gene 1. Thus the same linear regression, or in this case we would use a quadratic regression. The type of curve fitting is one which gives the tightest fit, the average is the Gene itself and the variable is the suppressor/activator impact.

Gene 3 is in the third regression fit. We can see that this gene is driven by itself and thus the parameter for its diagonal is obtained directly from the data of the microarray results.

7.2.2 Case 2: Constant Controlling and Suppressor/Activator Genes

Another approach is to consider the following. We have six genes, two of which (Genes 1 and 2) we know effect expression and four are suppressor/activator genes (Genes 3 thru 6).

					Color (nm)						
	Red					***Green***					***Blue***
Gene	400	425	450	475	500	525	550	575	600	625	650
1	1	1	1	2	2	2	2	3	3	3	3
2	2	2	2	2	3	3	3	2	2	1	1
3	3	3	2	4	4	5	6	7	7	8	9
4	1	1	1	2	2	2	2	3	3	3	4
5	9	9	9	6	6	4	4	4	4	3	3
6	2	2	2	6	6	6	9	9	9	12	15

Now we further note that:

$$x_1 = a^{11}c_1 + a^{13}c_3$$
$$and$$
$$x_2 = a^{22}c_2 + a^{23}c_3 \quad (0.11)$$
$$and$$
$$x_n = a^{nn}c_n \text{ for } n = 3...6$$

The constraints are:

$$\lambda(nm) = \kappa_1 a^{11}c_1 + \kappa_2 a^{22}c_2 + \sum_{n=3}^{6}\left(\kappa_{1n}a^{1n} + \kappa_{2n}a^{2n}\right)c_n \quad (0.12)$$

The above is a constraint.

We now have a set of wavelengths and concentrations. We know all the values except the a values which is the identification issue. We have the measurement tuples:

$$\left\{\lambda_k, x_1, ..., x_n\right\} \quad (0.13)$$

We now seek the following:

$$\hat{a} = \begin{bmatrix} \hat{a}_1 \\ .. \\ \hat{a}_n \end{bmatrix}$$

(0.14) such that

$$\min\left[\sum_{i=1}^{M} c_i^2(\hat{\lambda}_i - \lambda_i)^2 + \sum_{j=1}^{N}\sigma_i^2\left(\hat{x}_j - x_j\right)^2\right]$$

Let us recall the simple optimization result:

$$h(a) = \left[\sum_{i=1}^{M} c_i^2 (\hat{\lambda}_i - \lambda_i)^2 + \sum_{j=1}^{N} \sigma_i^2 \left(\hat{x}_j - x_j \right)^2 \right]$$

and

$$\frac{\partial h(a)}{\partial a_n} = g_n(a) = 0 \tag{0.15}$$

is the optimal point, so we seek to solve the vector equation:

g(a)=0

We can now state the general solution in terms of Newton's Method[25]:

$$g(a) = 0$$

is the desired result. Define:

$$A(a) = -\left[\frac{\partial g(a)}{\partial a} \right]^{-1}$$

(0.16) where we define:

$$\left[\frac{\partial g(a)}{\partial a} \right] = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_n} \\ \frac{\partial g_n}{\partial a_1} & \cdots & \frac{\partial g_n}{\partial a_n} \end{bmatrix}$$

and the estimate at sample k+1 is:

$$\hat{a}(k+1) = \hat{a}(k) + A(\hat{a}(k)) g(\hat{a}(k))$$

Note that we use this iterative scheme as one of several means to achieve the result. For each tuple of data we do the following:

$\hat{a}(0) = a^0$, an n x 1 vector guess. Then we use the first data tuple:

$$\hat{a}(1) = \hat{a}(0) + A(\hat{a}(0)) g((\hat{a}(0))$$

(0.17) where we use the difference:

$$a_{k,measured}(0) - \hat{a}(0)$$

as the data entry element for each of the elements of a.

The Newton algorithm is but one of many possible algorithms.

To be a bit more specific, for the x values we have:

[25] See Athans et al, Systems, Networks and Computation, Multivariable Methods, McGraw Hill (New York) 1974, pp-115-122.

$$\Delta x_1(0) = [x_1^{\text{measured cell 1}} - \hat{a}^{11}(0)c_1 + \hat{a}^{13}(0)c_3]$$

and

(0.18) $$\Delta x_2(0) = [x_2^{\text{measured cell 1}} - \hat{a}^{22}(0)c_2 + \hat{a}^{23}(0)c_3]$$

and

$$\Delta x_n(0) = [x_n^{\text{measured cell 1}} - \hat{a}^{nn}(0)c_n] \text{ for } n = 3...6$$

We know the conditions for Newton convergence. We can also estimate the accuracy of this algorithm as well.

7.2.3 Case 3: Fixed Expression Gene and Present or Absent Suppressor/Activator Genes

In this case we assume that the expression genes are fixed across the color base but now the suppressor/activator genes are present or absent depending on the color. This is a more likely case than what we have seen before. Namely it occurs when we hybridize in species since we can generally be certain that they all have the same expression genes which drive the pathways but the suppressor/activator genes may vary from species to species and as we hybridize the plant we get a mix and it is that mix of the suppressor/activator genes which yield the color variation.

Let us now assume we have the following microarray data. We assume gene 1 and 2 are the secondary path expression genes and that genes 3 thru 6 are the suppressor/activator genes.

						Color (nm)					
	Red					***Green***					***Blue***
Gene	400	425	450	475	500	525	550	575	600	625	650
1	1	1	1	2	2	2	2	3	3	3	3
2	2	2	2	2	3	3	3	2	2	1	1
3	0	1	0	1	0	1	0	1	0	1	0
4	0	0	1	1	0	0	1	1	0	0	1
5	0	0	0	0	1	1	1	1	0	0	0
6	0	0	0	0	0	0	0	0	1	1	1

We see we have distinct suppressor/activator combinations in each column. A 0 indicate gene absence or senescence and a 1 indicates it is active. Namely an active gene had produced mRNA and thus cDNA is binding.

This model allows for a binary analysis of the suppressor/activator elements. The models for genes 1 and 2 remain as above but the gene for 2 thru 4 now are present or absent. We also not that with 4 suppressor/activator genes and being present or absent offers 2^4 or 16 possible outcomes. We can see that a model of this form creates a binary modeling of phenotypic expression levels. We have not in this model allowed for partial expression. Partial expression may be present in actuality and caused by other secondary suppressor/activators which are much lower level. We generally model these effects as noise.

Unlike Case 2 where we assumed that all constants were the same or in Case 1 where we assumed the suppressor/activator constants changed, in this case we assume that they are either

on or off. However we do assume when they are on that they have the same constant in their effects on the expression genes.

We now re-look at the measurement model:

$$\lambda_k(nm) = \kappa_1 a^{11} c_1 + \kappa_2 a^{22} c_2 + \sum_{n=3}^{6} \left(\kappa_{1n,k} a^{1n} + \kappa_{2n,k} a^{2n} \right) c_n \tag{0.19}$$

That is, the wavelength has the same model for each wavelength but we have introduced constants inside the summation which are dependent on whether the gene is active or not. Thus we have the same model as before yet we now have to make the κ dependent on the presence or absence, that is we admit 0 values and we obtain them from the microarray analysis.

7.2.4 Case 4: Noise

In this case we introduce noise in the system as well as the measurement. Noise can be from many sources; measurement error, unknown effects, interference, inaccuracy in modeling, and other as yet unknown factors.

The system model becomes:

$$\begin{aligned} &\frac{dx(t)}{dt} = Ax(t) + u(t) + n(t) \\ &where \\ &E\left[n(t)\right] = 0 \\ &and \\ &E\left[n(t) n^T(t+\tau) \right] = Q_0 \delta(\tau) \end{aligned} \tag{0.20}$$

Namely n(t) is Gaussian white noise. Q is an n x n matrix. We also have left open the use of the u(t) driver. This may be from exogenous causes.[26]

We may also have noise in the microarray measurements. In fact this is quite common.

Thus we would have to modify the Newton method to account for the noise. Specifically we would have:

[26] Mol et al (1998) How Genes Paint Flowers and Seed states various factors for color. They include: (i) the pigment itself which is the essence of the model we have developed, (ii) co-pigments and the presence of related metal ions which effect the secondary pathway in question, (iii) cell and vacuolar pH, and (iv) shape of the cell and/or its location. The last three of these can be modeled as drivers if known or one can attempt to add this to the noise load if a reliable statistical base is present.

$$h(a) = E\left[\sum_{i=1}^{M} c_i^2(\hat{\lambda}_i - \lambda_i)^2 + \sum_{j=1}^{N} \sigma_j^2\left(\hat{x}_j - x_j\right)^2\right]$$

and

$$\frac{\partial h(a)}{\partial a_n} = g_n(a) = 0 \tag{0.21}$$

is the optimal point, so we seek to solve the vector equation:

g(a)=0

where we have now added the expectation operator. The remaining analysis remains the same.

7.3 Example 3: To Be Determined Expression Model, Known Secondary Model, Desired Color

In this case we assume we have identified the A matrix and we know the secondary transformation and we are given a desired output color. Then we are required to determine the entries in the A matrix which will yield the desired color. This is the design problem.

Using the approach above, we can taken any gene expression system and with the known secondary we can determine the expression constants and then using them determine using standard control techniques what the parameters should be to obtain the desired $\lambda(nm)$.[27]

We provide two simple solutions to Case 2 and Case 3.

7.3.1 Case 2 Expression

In case 2 we have the following:

[27] See McGarty, Stochastic Systems and State Estimation, this is the argument for controllability for a linear system. Namely choose a determined end point and it will tell you how to get there.

A^* = desired transformation

$\hat{A}$ = estimated transformation from data analysis

λ^* = desired wavelength

$\hat{\lambda}$ = wavelength with estimated transformation

then we have:

(0.22) $\kappa^{\mathrm{T}}(A^* - \hat{A})c = \lambda^* - \hat{\lambda}$

choosing $\tilde{\kappa}$ and c̃ to create pseudo-inverse matrices:

$K = \tilde{\kappa}\kappa^T$

and

$C = \tilde{c}c^T$

we have:

$A^* = \hat{A} + K^{-1}(\lambda^* - \hat{\lambda})C^{-1}$

7.3.2 Case 3: Present/Absent Suppressor/Activator Genes

In this case we can determine what the answer would be for any combination of suppressor/activator genes. Then we can obtain the solution by exhaustive search in those cases of a reasonable number of suppressor/activator genes. Namely, we can use the following algorithm:

the actual wavelegnt produced is

$\lambda = KA^{-1}c$

the desired wavelength is

λ^*

(0.23)
$$A = \begin{bmatrix} a_{11}...a_{1m} & \delta_{1(m+1)}a_{1(m+1)}.....\delta_{1n}a_{1n} \\ ..... & \\ a_{m1}...a_{mm} & \delta_{m(m+1)}a_{m(m+1)}...\delta_{mn}a_{mn} \\ \delta_{(m+1)1}a_{(m+1)1}.....\delta_{mn}a_{mn} & \\ ..... & \\ \delta_{n1}a_{n1}.....\delta_{nn}a_{nn} & \end{bmatrix}$$

where

$\delta_{jk} = 1, 0$

now choose every possible δ_{jk} combination to find the desired λ

Various other algorithms can be chosen for this task. Also in an interesting manner the approach to case 3 is akin to a coding problem or signal design problem. We are looking for a signal constellation which optimizes the performance.

8 CONCLUSIONS

In this paper we have presented an interesting genus to study with respect to gene expression and ultimately the control of gene expression. The phenotypes are quite obvious in flower colors and in addition the hybridizing which has led to a wealth of examples has been done just in the past one hundred years. Also we have a reasonably clear understanding of the underlying species and we can readily assess the complexity of the species DNA structure.

We also have a well defined and understood set of pathways that give rise to the phenotype. We further know the effecting proteins, enzymes, and we also know the gene which effect the proteins in question. Finally we have well accepted models for the expression of the genes and we can use generally accepted models for the dynamics of gene expression.

This has led us in our final section to a modeling of gene expression as a set of definable dynamic systems. We have used a certain set of those systems to discuss examples. However certain key questions remain:

First, what are the dynamic models which can adequately and correctly describe the abrupt coloration of the flowers. We have a good understanding of many of the unstable dynamic systems models which can describe such phenomenon but what is the relationship between what occurs in the cell and what the models describe.

Second, we have used an ensemble approach versus the microbiologists time approach to modeling the system. We have posited an equivalency based upon the Ergodic Theorem. However there is no experimental proof of this fact. Such a study must be performed.

Third, in any systems approach, we always look at issues as observability and controllability. Observability simply is if we can see the outputs knowing the system model can we predict the initial condition. This must be validated experimentally. Controllability is simply can we drive the system to a desire point with a control function. The controllability question goes to the heart of flower color design. If we accept the validity of our models the answer appears to be determinable for any set of defines pathways.

Fourth, we have suggested a microarry approach to estimating the coefficients of the dynamic system. This is one of many possible techniques. The first part we should do is address this from an experimental perspective. Namely perform the microarray analysis. The second part is to investigate alternative methods of solving the system identification problem via alternative bench based validation tools.

Fifth, specific phenotypic design must be considered in more detail and experimentally validated.

Sixth, we use a stochastic model for the expression and pathway analysis. We used this as a way to account for dimensions we could not include because they were expressed at too low a level or because we had not knowledge of their existence. Thus we argued that like one would do in a system design analysis, noise may be true random processes or the aggregation of currently unknown tertiary processes. Experimental validation of this modeling element must be performed.

Seventh, can this approach be carried over to any other cell line? The answer we believe is yes it can and readily. What we have done herein is to focus on phenotypic characteristics and ones which are readily characterizable by well understood pathways. Such systems exist in many other systems including the human.

9 APPENDIX A: AN EXAMPLE

Let us consider a more complex example of a three color element and six gene system. We assume that we have two genes which effect proteins which in turn effect secondary products, namely anthocyanidins. We also assume we have four other genes which are activators, or suppressors, of these genes; two per gene. We then develop a simple models and from that model determine the color of the resulting combination using a Red, Green, Blue model for colorimetry.

There are the same three problems that we had discussed in the text, namely:

4. **Analysis:** In the analysis problem we assume we know the expression dynamics and the secondary production model. We then ask given those two what color do we get. We develop this in detail.

5. **Identification:** The identification problem is one in which we know the secondary processes, we have many color samples and we know the protein concentrations which yield each color. Then we ask how do we determine the A matrix for the gene expression.

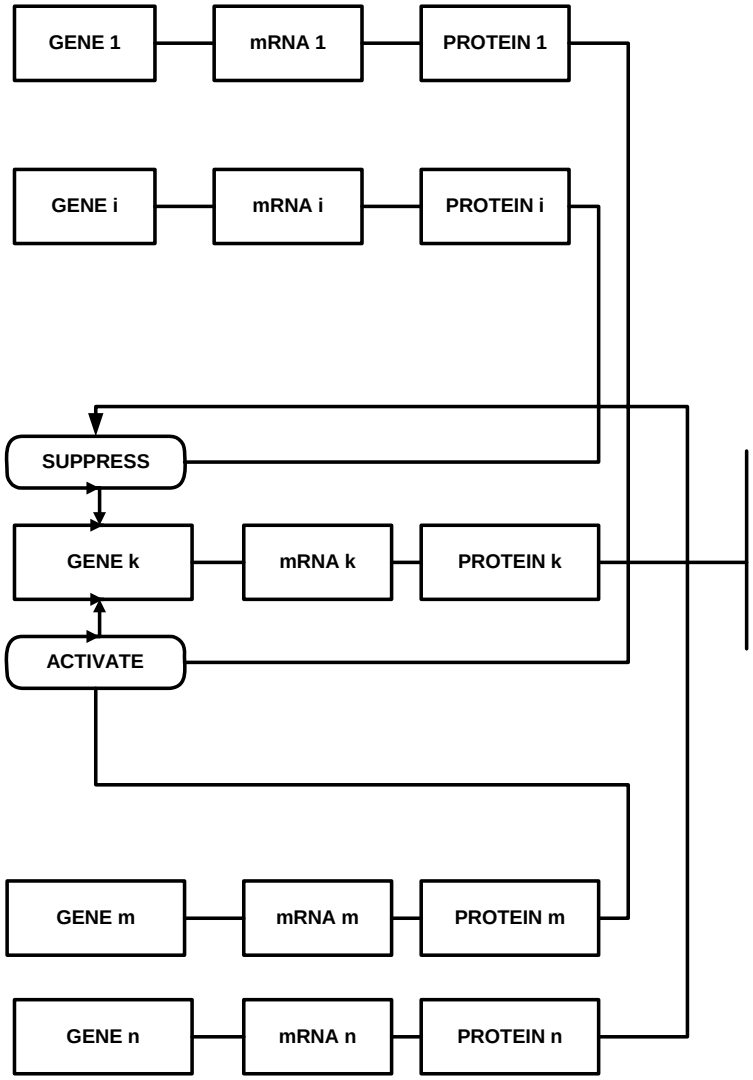
6. **Design:** This problem is of significant interest. We seek a desired output state, color in our example. We know the gene expression dynamics and the secondary model. We then ask how do we modify the gene expression model to obtain the desired output.

9.1 The Genetic Structure

The genetic model we use is quite simple. There is a set of genes which control secondary pathways. These genes are themselves controlled by activator or repressor genes, A/R genes, which indirectly control the pathway. In addition these A/R genes may be turned on or off depending on other variables. But, when on or off the A/R genes function in a determinable manner.

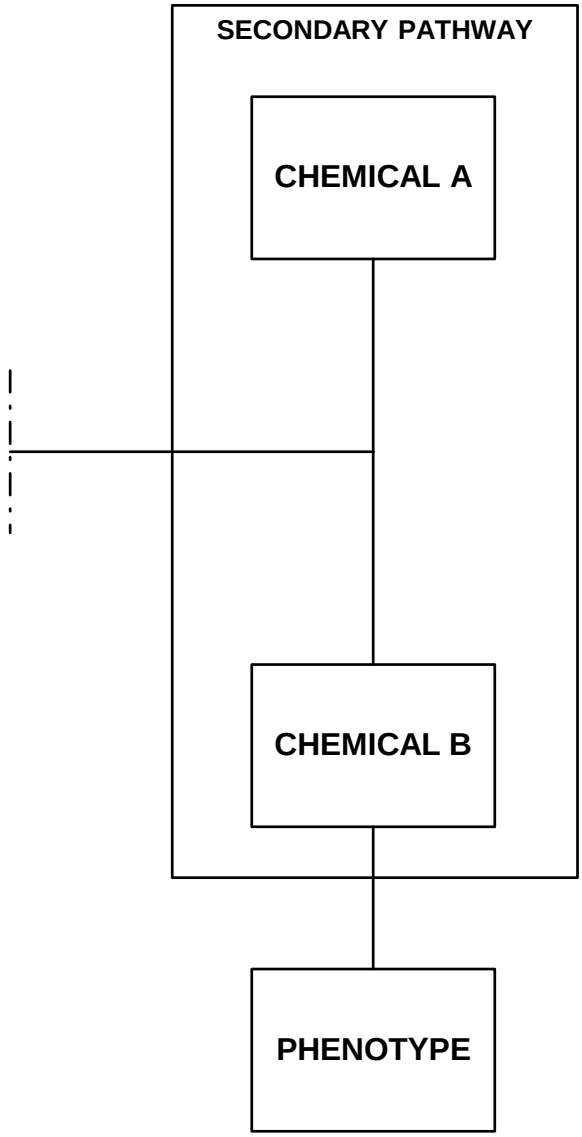
The following figure details what we are presenting.

Fig. 1a



The following Figure completes the detail showing the control of the secondary pathway. The results of the secondary pathway define the phenotype result.

Fig. 1b



9.2 Color Model

We now develop a simple color model. Let us use the standard approach to color by using the Red, Green, Blue model, using the variable R, G, and B. By weighting each of these elements we define the color. Thus we have for any color C the following:

$$C = m_1 R + m_2 G + m_3 B$$

Here the elements m are the amounts of R, G, and B respectively that generate the color C. We can define this as:

$$C = m^T M$$

where

$$m = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$$

and

$$M = \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Now we will assume three anthocyanins; Pelargonidin, Delphinidin and Cyanidin. Then we know that each of these anthocyanins has a color and they can be represented as:

$$C_P = \gamma_P^T M$$
$$C_D = \gamma_D^T M$$
$$C_C = \gamma_C^T M$$

where

$$\gamma_K = \begin{bmatrix} \gamma_{K1} \\ \gamma_{K2} \\ \gamma_{K3} \end{bmatrix}$$

This tells us that each of the three colors can be represented in weighted parts of the base color vector M.

Now if we have a solution of all three of these anthocyanins we assume that the resulting color is merely a weighted combination of each of the elements. Namely we assume we have relative concentrations for each as:

ρ_P = Relative Concentration of Pelargonidin
ρ_D = Relative Concentration of Delphinidin
ρ_C = Relative Concentration of Cyanidin
and

$$\sum_{i=1}^{3} \rho_i = 1$$

This yields a weighted color as:

$$Color = C =$$
$$\left[\rho_P\gamma_{P1} + \rho_D\gamma_{D1} + \rho_C\gamma_{C1}\right] +$$
$$\left[\rho_P\gamma_{P2} + \rho_D\gamma_{D2} + \rho_C\gamma_{C2}\right] +$$
$$\left[\rho_P\gamma_{P3} + \rho_D\gamma_{D3} + \rho_C\gamma_{C3}\right]$$

Now we can write this as follows:

$$C = \rho^T HM$$
where

$$H = \begin{bmatrix} \gamma_{P1}..\gamma_{D1}..\gamma_{C1} \\ \gamma_{P2}..\gamma_{D2}..\gamma_{C2} \\ \gamma_{P3}..\gamma_{D3}..\gamma_{C3} \end{bmatrix}$$

Thus we can relate M to the relative densities:

$$m^T = \rho^T H$$
or

$$\rho = \left[H^T\right]^{-1} m$$

This tells us that if we can measure the m elements in the color of the sample then we can determine the relative concentrations of the three anthocyanins. This can be generalized immediately to any set of such elements. The only issue will be how we interpret the inverse. The use of a generalized inverse may be required.

9.3 The Model

Let us consider a six gene model, two color modifying genes and four control genes, two each. The model is as follows:

$$\frac{dx(t)}{dt} = Ax(t) + u(t)$$

where

$$A = \begin{bmatrix} a_{11}..a_{12}..a_{13}..0..0..0 \\ 0...a_{22}..0...0...0...0 \\ 0...0...a_{33}..0...0...0 \\ 0...0...0...a_{44}..a_{45}..a_{46} \\ 0...0...0...0...a_{55}..0 \\ 0...0...0...0...0...a_{66} \end{bmatrix}$$

and

$$u(t) = \begin{bmatrix} u_1 \\ ... \\ u_6 \end{bmatrix}$$

Now we can define:

$$A = \begin{bmatrix} A_1...0 \\ 0...A_2 \end{bmatrix}$$

where we have partitioned the matrix into four submatrices. This shows that each gene and its controller are separate.

Now we can determine the concentrations of each protein in steady state as follows:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -A_1^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

and

$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix} = -A_2^{-1} \begin{bmatrix} u_4 \\ u_5 \\ u_6 \end{bmatrix}$$

Now clearly we can obtain the inverse to yield:

$$A_1^{-1} = \frac{\begin{bmatrix} a_{22}a_{33}.....0.....0 \\ a_{12}a_{33}..a_{11}a_{33}.....0 \\ a_{13}a_{22}.....0...a_{11}a_{22} \end{bmatrix}^T}{\det A_1}$$

yiedling

$$= \frac{\begin{bmatrix} a_{22}a_{33}..a_{12}a_{33}..a_{13}a_{22} \\ 0.....a_{11}a_{33}.....0 \\ 0.....0.....a_{11}a_{22} \end{bmatrix}}{a_{11}a_{22}a_{33}} = \begin{bmatrix} \frac{1}{a_{11}}...\frac{a_{12}}{a_{11}a_{22}}..\frac{a_{13}}{a_{11}a_{33}} \\ 0.....\frac{1}{a_{22}}.....0 \\ 0.....0.....\frac{1}{a_{33}} \end{bmatrix}$$

We can define this as:

$$A^{-1} = \begin{bmatrix} a_{11}^{-1}..a_{12}^{-1}..a_{13}^{-1} \\ 0....a_{22}^{-1}...0 \\ 0....0.....a_{33}^{-1} \end{bmatrix}$$

we will argue that finding either the matrix elements or their inverse relatives is identical. Thus we focus on the inverse elements.

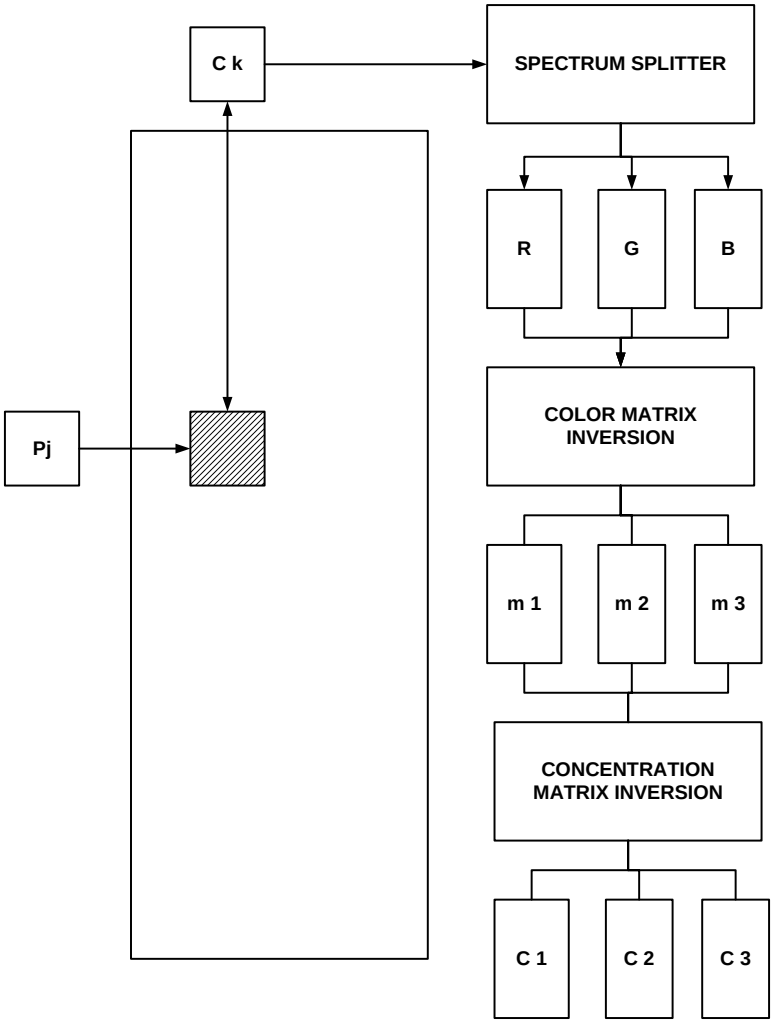
Now the concentrations of the anthocyanins is given by the 2 x 2 vector as follows:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} c_{11}...0...0...0...0...0 \\ 0...0...0...c_{24}...0...0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = Cx$$

9.4 Experiment and Data

The collection of data proceeds as shown below. We use a microarray with differing color sample in each column and different gene sequences, cDNA in each row. We measure the gene

concentration using standard means in each cell and we then measure for each color the RGB elements and then convert that to m values for the anthocyanins which are present and then convert that to the concentration conversion units, namely c.



Note from the earlier discussion that we can determine the concentrations of the anthocyanins from the conversion matrices we had developed beforehand using the methodology we have shown above.

9.5 Optimization

Now we can rephrase the optimization problem. Let us assume from the experiment we measure the gene concentrations and we measure the values of the color elements. Thus we have the x and m values. The we seek to select the "a" or the α values to minimize the following[28]:

[28] We select the s and σ values based upon the data set themselves. They are the calculated variances of the within group data sets. This is comparable to a typical analysis of variance model.

$$h(a) = \sum_{k=1}^{K} s_k^2 (m_k - \hat{m}_k)^2 + \sum_{j=1}^{J} \sigma_j^2 (x_j - \hat{x}_j)^2$$

Since the solution is to choose the a vector to minimize this for all selected data points, we need to determine an algorithm to obtain this. We first note that we are minimizing the weighted difference between the measurement and what the measurement should have been using the guess on the a's. The weights are not necessarily arbitrary. They can be shown to be the variance of the values in order to give a minimum mean square fit.

The goal is to solve the following equations:

$$g_i = \frac{\partial h(a)}{\partial a_i} = \sum_{k=1}^{K} s_k^2 (m_k - \hat{m}_k) \frac{\partial \hat{m}_k}{\partial a_i} + \sum_{j=1}^{J} \sigma_j^2 (x_j - \hat{x}_j) \frac{\partial \hat{x}_j}{\partial a_i} = 0; \nabla i = 1 \ldots n$$

This can be solved using the Newton method amongst others. We choose that method. We also note we can use a or α. We choose the latter for simplicity. We define α as:

$$\begin{bmatrix} \alpha_{11} \ldots\ldots \alpha_{13} \\ \ldots\ldots\ldots\ldots \\ \alpha_{31} \ldots\ldots \alpha_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{a_{11}} \ldots \frac{a_{12}}{a_{11}a_{22}} \ldots \frac{a_{13}}{a_{11}a_{33}} \\ 0 \ldots\ldots\ldots \frac{1}{a_{22}} \ldots\ldots\ldots 0 \\ 0 \ldots\ldots\ldots 0 \ldots\ldots\ldots \frac{1}{a_{33}} \end{bmatrix}$$

Thus we solve the problem:

$$g_i(\alpha) = \frac{\partial h(\alpha)}{\partial \alpha_i} = \sum_{k=1}^{K} s_k^2 (m_k - \hat{m}_k) \frac{\partial \hat{m}_k}{\partial \alpha_i} + \sum_{j=1}^{J} \sigma_j^2 (x_j - \hat{x}_j) \frac{\partial \hat{x}_j}{\partial \alpha_i} = 0; \nabla i = 1 \ldots n$$

Using the Newton method we obtain:

$$\hat{\alpha}(k+1) = \hat{\alpha}(k) + K^{-1}(k+1,k) g(\hat{\alpha}(k))$$

where we have for K

$$K = \begin{bmatrix} K_{11}.....K_{1p} \\ ..... \\ ..... \\ K_{p1}.....K_{pp} \end{bmatrix}$$

where

$$K_{ij} = \frac{\partial g_i}{\partial a_j}$$

In this case we have 5 possible values for a. Thus the a vector is a 5 x 1 vector. The K matrix is a 5 x 5 matrix.

We present in the following Figures the process in detail:

Fig. 3a

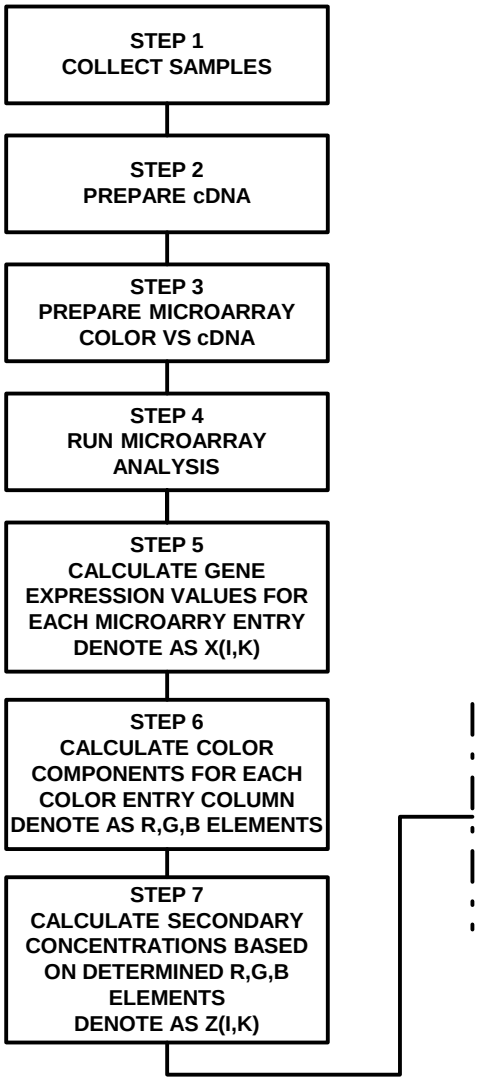
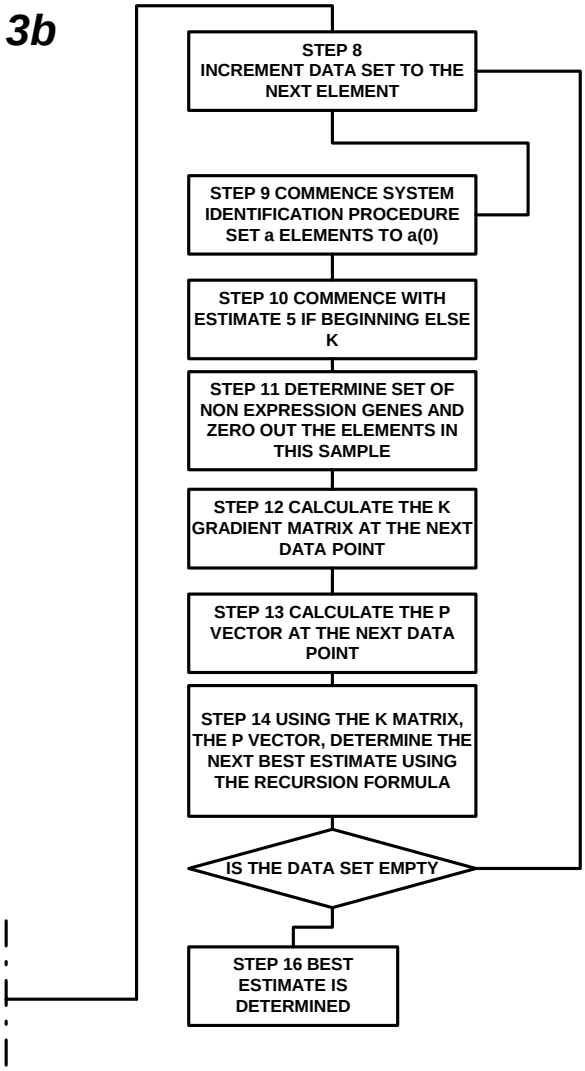

STEP 1
COLLECT SAMPLES
STEP 2
PREPARE cDNA
STEP 3
PREPARE MICROARRAY
COLOR VS cDNA
STEP 4
RUN MICROARRAY
ANALYSIS
STEP 5
CALCULATE GENE
EXPRESSION VALUES FOR
EACH MICROARRY ENTRY
DENOTE AS X(I,K)
STEP 6
CALCULATE COLOR
COMPONENTS FOR EACH
COLOR ENTRY COLUMN
DENOTE AS R,G,B ELEMENTS
STEP 7
CALCULATE SECONDARY
CONCENTRATIONS BASED
ON DETERMINED R,G,B
ELEMENTS
DENOTE AS Z(I,K)

Fig. 3b



9.5.1 Modification for on/off A/R Genes

There is a slight modification we must include. We must return to the beginning to best understand it. Namely if the a's in the system matrix are all constant then by definition the colors remain the same. However, if any one or more of the A/R genes are on or off then we can get variation. We first explore the implications of this and then we modify the estimator process accordingly. Let us review our model assumptions:

1. We assume that the genes directing the secondary pathway are always functioning.

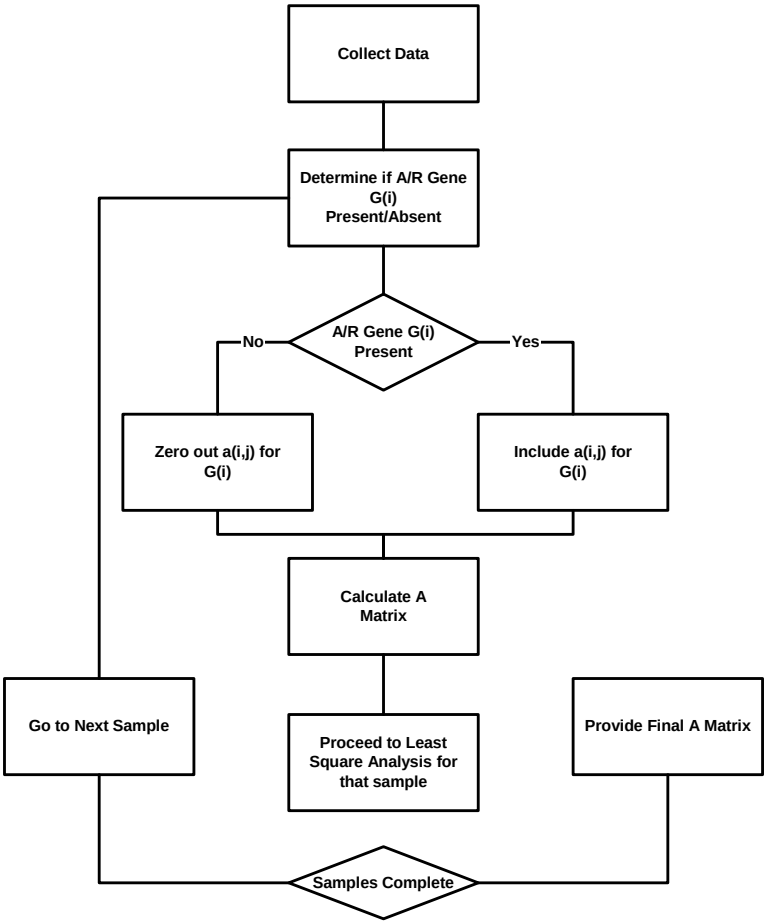
2. We assume that the constants in the gene expression model and secondary control are all constant and remain so.

3. We assume that the A/R gene may be on or off. They are controlled via some tertiary process yet to be determined.

Thus is we have a three gene system with two A/R gene per expression gene we have 4X4X4 possible states. This means we have 64 possible color states. If we have n A/R genes per expression gene and we have m expression genes we have 2^{nm} possible color states.

Now the above algorithm is a least squares estimate algorithm given an A/R gene state. We now propose a model where we first estimate the state of the A/R genes and then given that state we use the least squares approach to estimate the a values which remain. Thus if a specific A/R gene is in a 0 state we then zero out its effecting a value and estimate the remaining a values as we would have done before.

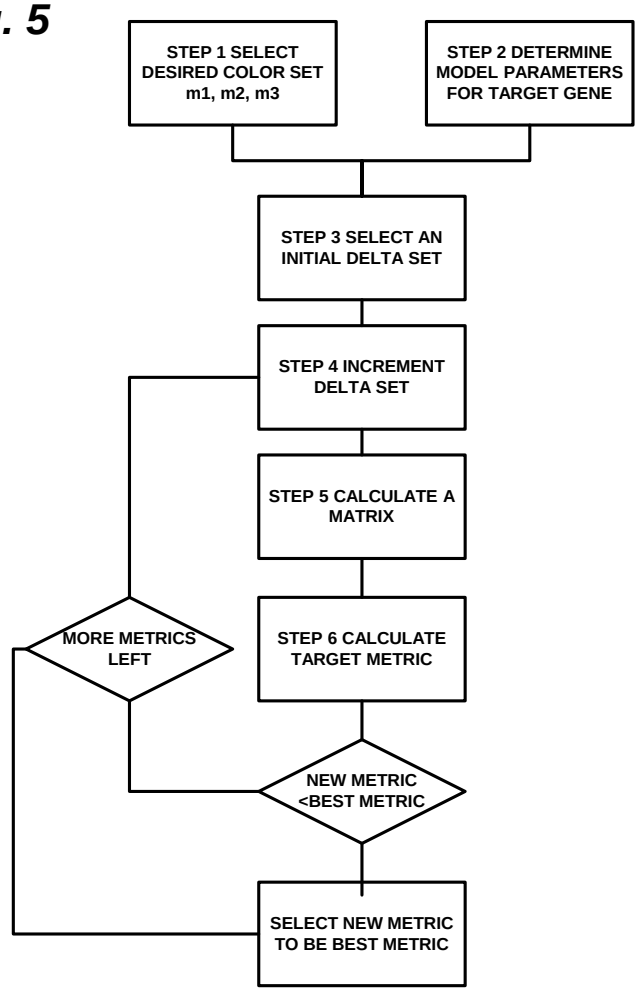
The algorithm for the calculation is shown below.



9.6 Design

We are now in the design phase of the process. In the above steps we have used data to determine the elements of the A matrix. This tells us what we need to go to the design stage.

Fig. 5



9.7 Some Issues on Estimates

There a few minor issues on probability and estimation which we review since we use this several times.

9.7.1 Averages of Functions

First, let us assume we want to determine the following:

$$E\left[f(x) \right]$$

Now use a Taylor series to obtain:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \ldots$$

Choose the expansion point as:

$$x_0 = E[x]$$

The we have:

$$E[f(x)] = f(E[x]) + \frac{1}{2}\sigma^2 f''(E[x]) + o(E[x])$$

We generally assume a small variance and small second derivative so that the average of the function is the function of the average.

9.7.2 Conditional Estimates

In the above analysis we see the case where we have the following:

Problem 1: Determine the best estimate given that all genes are present.

Problem 2: Given the best estimate given some genes are inactive. What is the probability they are inactive?

Problem 1 is determined by the least square approach where we assume the a values remain constant for all genes throughout.

Problem 2 divides into two estimation problems. First we estimate the on or off state of a gene for any A/R gene and then assuming that state we use the least square approach for the remaining estimates.

These problems are handled as we do herein in the literature at large.

10 APPENDIX B: ESTIMATION

The method of estimating the structural elements of the gene expression can be structured using a standard set of methodologies. In particular we use the two approaches used in McGarty 1971[29] and McGarty 1974. The McGarty 1971 approach was applied to estimating the constituent chemical concentrations of the upper atmosphere, namely the inversion problem, using transmitted light as the probe mechanism. In this case we seek to estimate the gene expression matrix using the concentrations of secondary chemicals as expressed in color concentrations. This is in many ways a similar problem.

10.1 The Model

Let us consider a six gene model, two color modifying genes and four control genes, two each. The model is as follows:

$$\frac{dx(t)}{dt} = Ax(t) + u(t) + n(t)$$

where

$$A = \begin{bmatrix} a_{11}..a_{12}..a_{13}..0..0..0 \\ 0...a_{22}..0...0...0...0 \\ 0...0...a_{33}..0...0...0 \\ 0...0...0...a_{44}..a_{45}..a_{46} \\ 0...0...0...0...a_{55}..0 \\ 0...0...0...0...0...a_{66} \end{bmatrix}$$

and

$$u(t) = \begin{bmatrix} u_1 \\ ... \\ u_6 \end{bmatrix}$$

and

$$E\left[n(t)\right] = 0$$

and

$$E\left[n(t)n(s)\right] = N_0 I\delta(t-s)$$

[29] See, McGarty, The Estimation of the Constituent Densities of the Upper Atmosphere by Means of **a** Recursive Filtering Algorithm, IEEE AC 1971.

Now we can define:

$$A = \begin{bmatrix} A_1 \ldots 0 \\ 0 \ldots A_2 \end{bmatrix}$$

where we have partitioned the matrix into four submatrices. This shows that each gene and its controller are separate.

Now we can determine the concentrations of each protein in steady state as follows, neglecting the Gaussian noise element for the time being:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -A_1^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

and

$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix} = -A_2^{-1} \begin{bmatrix} u_4 \\ u_5 \\ u_6 \end{bmatrix}$$

Now clearly we can obtain the inverse to yield:

$$A_1^{-1} = \frac{\begin{bmatrix} a_{22}a_{33} \ldots\ldots 0 \ldots\ldots 0 \\ a_{12}a_{33} \ldots a_{11}a_{33} \ldots\ldots 0 \\ a_{13}a_{22} \ldots\ldots 0 \ldots a_{11}a_{22} \end{bmatrix}^T}{\det A_1}$$

yiedling

$$= \frac{\begin{bmatrix} a_{22}a_{33} \ldots a_{12}a_{33} \ldots a_{13}a_{22} \\ 0 \ldots\ldots\ldots a_{11}a_{33} \ldots\ldots 0 \\ 0 \ldots\ldots\ldots 0 \ldots\ldots\ldots a_{11}a_{22} \end{bmatrix}}{a_{11}a_{22}a_{33}} = \begin{bmatrix} \frac{1}{a_{11}} \ldots \frac{a_{12}}{a_{11}a_{22}} \ldots \frac{a_{13}}{a_{11}a_{33}} \\ 0 \ldots\ldots\ldots \frac{1}{a_{22}} \ldots\ldots\ldots 0 \\ 0 \ldots\ldots\ldots 0 \ldots\ldots\ldots \frac{1}{a_{33}} \end{bmatrix}$$

We can define this as:

$$A^{-1} = \begin{bmatrix} a_{11}^{-1}..a_{12}^{-1}..a_{13}^{-1} \\ 0....a_{22}^{-1}...0 \\ 0....0.....a_{33}^{-1} \end{bmatrix}$$

we will argue that finding either the matrix elements or their inverse relatives is identical. Thus we focus on the inverse elements.

Now the concentrations of the anthocyanins is given by the 2 x 2 vector as follows:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} c_{11}...0...0...0...0...0 \\ 0...0...0...c_{24}...0...0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = Cx$$

10.2 Color Model

We now develop a simple color model. Let us use the standard approach to color by using the Red, Green, Blue model, using the variable R, G, and B. By weighting each of these elements we define the color. Thus we have for any color C the following:

$$C = m_1 R + m_2 G + m_3 B$$

Here the elements m are the amounts of R, G, and B respectively that generate the color C. We can define this as:

$$C = m^T M$$

where

$$m = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$$

and

$$M = \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Now we will assume three anthocyanins; Pelargonidin, Delphinidin and Cyanidin. Then we know that each of these anthocyanins has a color and they can be represented as:

$$C_P = \gamma_P^T M$$
$$C_D = \gamma_D^T M$$
$$C_C = \gamma_C^T M$$

where

$$\gamma_K = \begin{bmatrix} \gamma_{K1} \\ \gamma_{K2} \\ \gamma_{K3} \end{bmatrix}$$

This tells us that each of the three colors can be represented in weighted parts of the base color vector M.

Now if we have a solution of all three of these anthocyanins we assume that the resulting color is merely a weighted combination of each of the elements. Namely we assume we have relative concentrations for each as:

ρ_P = Relative Concentration of Pelargonidin
ρ_D = Relative Concentration of Delphinidin
ρ_C = Relative Concentration of Cyanidin

and

$$\sum_{i=1}^{3} \rho_i = 1$$

This yields a weighted color as:

$$Color = C =$$
$$\left[\rho_P\gamma_{P1} + \rho_D\gamma_{D1} + \rho_C\gamma_{C1}\right] +$$
$$\left[\rho_P\gamma_{P2} + \rho_D\gamma_{D2} + \rho_C\gamma_{C2}\right] +$$
$$\left[\rho_P\gamma_{P3} + \rho_D\gamma_{D3} + \rho_C\gamma_{C3}\right]$$

Now we can write this as follows:

$$C = \rho^T HM$$
where
$$H = \begin{bmatrix} \gamma_{P1}..\gamma_{D1}..\gamma_{C1} \\ \gamma_{P2}..\gamma_{D2}..\gamma_{C2} \\ \gamma_{P3}..\gamma_{D3}..\gamma_{C3} \end{bmatrix}$$

Thus we can relate M to the relative densities:

$$m^T = \rho^T H$$
or
$$\rho = \left[H^T\right]^{-1} m$$

This tells us that if we can measure the m elements in the color of the sample then we can determine the relative concentrations of the three anthocyanins. This can be generalized immediately to any set of such elements. The only issue will be how we interpret the inverse. The use of a generalized inverse may be required.

10.3 The Estimator Model

The system model is as follows. Let us begin with a model for the to be determined vector a:

$$\frac{da(t)}{dt} = 0$$

where

$$a(t) = \begin{bmatrix} a_1 \\ \dots \\ a_5 \end{bmatrix}$$

In this case we have assumed a is a 5 x 1 vector but it can be any vector.

The measurement system equation is given by:

$$z(t) = g(a,t) + w(t)$$

where z is an m x 1 vector. In this case however we have for the measurement the following:

$$z(t) = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ x_1 \\ \dots \\ x_6 \end{bmatrix} = g(a,t) + w(t)$$

we now expand in a Taylor series the above g function:

$$g(a,t) = g(a_0,t) + C(a_0,t)\left[a(t) - a_0(t)\right] + \frac{1}{2}\sum_{i=1}^{N} \gamma_i \left[a - a_0\right]^T F_i \left[a - a_0\right] + \dots$$

where we have:

$$C = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \dots & \frac{\partial g_1}{\partial a_n} \\ .. & .. & .. \\ \frac{\partial g_m}{\partial a_1} & \dots & \frac{\partial g_m}{\partial a_n} \end{bmatrix}$$

Thus we have for the measurement:

$$z(t) = C(t)a(t) + \left[g(a_0) - C(a_0)a_0(t)\right]$$

We now use standard Kalman theory to determine the mean square estimate;

$$\frac{d\hat{a}(t)}{dt} = P(t)C^T(t)K^{-1}(z - C(t)\hat{a}(t))$$

where

$$\frac{dP(t)}{dt} = -P(t)C^T(t)K^{-1}C(t)P(t) +$$
$$\sum_{i=1}^{N} PF_i P\gamma_i{}^T K^{-1}(z - g(a_0)$$

where

$$K\gamma(t-s) = E\left[w(t)w^T(s)\right]$$

In discrete time we have the equation:

$$\hat{a}(k+1) = \hat{a}(k) + PCK^{-1}\left[z(k) - \hat{z}(k)\right]$$

which is identical to the equation we derived from the Newton method.

10.4 Optimization

Now we can rephrase the optimization problem. Let us assume from the experiment we measure the gene concentrations and we measure the values of the color elements. Thus we have the x and m values.

The "General Model" is to select a values so that we minimize a weighted mean square error as may be provided below:

$$\min\left[a(t)-\hat{a}(t)\right]^2$$

or

$$\min \quad \mathrm{E}[(m-\hat{m})^T S(m-\hat{m})]+E[(x-\hat{x})^T \Sigma(x-\hat{x})]$$

The first minimization is what we would attain in a MMSE approach using a Kalman like filter. The second approach is one wherein we attempt to get a best fir to the data obtained. We argue that there is an equivalence to both approaches.

Then, if we simplify, we may seek to select the "a" or the α values to minimize the following[30]:

$$h(a)=\sum_{k=1}^{K} s_k^2(m_k-\hat{m}_k)^2+\sum_{j=1}^{J}\sigma_j^2(x_j-\hat{x}_j)^2$$

Since the solution is to choose the a vector to minimize this for all selected data points, we need to determine an algorithm to obtain this. We first note that we are minimizing the weighted difference between the measurement and what the measurement should have been using the guess on the a's. The weights are not necessarily arbitrary. They can be shown to be the variance of the values in order to give a minimum mean square fit.

The goal is to solve the following equations:

$$g_i=\frac{\partial h(a)}{\partial a_i}=\sum_{k=1}^{K} s_k^2(m_k-\hat{m}_k)\frac{\partial \hat{m}_k}{\partial a_i}+\sum_{j=1}^{J}\sigma_j^2(x_j-\hat{x}_j)\frac{\partial \hat{x}_j}{\partial a_i}=0;\nabla i=1...n$$

This can be solved using the Newton method amongst others. We choose that method. We also note we can use a or α. We choose the latter for simplicity. We define α as:

$$\begin{bmatrix}\alpha_{11}.....\alpha_{13}\\ .....\\ \alpha_{31}.....\alpha_{33}\end{bmatrix}=\begin{bmatrix}\frac{1}{a_{11}}...\frac{a_{12}}{a_{11}a_{22}}..\frac{a_{13}}{a_{11}a_{33}}\\ 0.....\frac{1}{a_{22}}.....0\\ 0.....0.....\frac{1}{a_{33}}\end{bmatrix}$$

[30] We select the s and σ values based upon the date set themselves. They are the calculated variances of the within group data sets. This is comparable to a typical analysis of variance model.

Thus we solve the problem:

$$g_i(\alpha) = \frac{\partial h(\alpha)}{\partial \alpha_i} = \sum_{k=1}^{K} s_k^2 (m_k - \hat{m}_k) \frac{\partial \hat{m}_k}{\partial \alpha_i} + \sum_{j=1}^{J} \sigma_j^2 (x_j - \hat{x}_j) \frac{\partial \hat{x}_j}{\partial \alpha_i} = 0; \forall i = 1...n$$

Using the Newton method we obtain:

$$\hat{\alpha}(k+1) = \hat{\alpha}(k) + K^{-1}(k+1,k) g(\hat{\alpha}(k))$$

where we have for K

$$K = \begin{bmatrix} K_{11}.....K_{1p} \\ ..... \\ ..... \\ K_{p1}.....K_{pp} \end{bmatrix}$$

where

$$K_{ij} = \frac{\partial g_i}{\partial a_j}$$

In this case we have 5 possible values for a. Thus the a vector is a 5 x 1 vector. The K matrix is a 5 x 5 matrix.

10.4.1 Modification for on/off A/R Genes

There is a slight modification we must include. We must return to the beginning to best understand it. Namely if the a's in the system matrix are all constant then by definition the colors remain the same. However, if any one or more of the A/R genes are on or off then we can get variation. We first explore the implications of this and then we modify the estimator process accordingly. Let us review our model assumptions:

4. We assume that the genes directing the secondary pathway are always functioning.

5. We assume that the constants in the gene expression model and secondary control are all constant and remain so.

6. We assume that the A/R gene may be on or off. They are controlled via some tertiary process yet to be determined.

Thus is we have a three gene system with two A/R gene per expression gene we have 4X4X4 possible states. This means we have 64 possible color states. If we have n A/R genes per expression gene and we have m expression genes we have 2^{nm} possible color states.

Now the above algorithm is a least squares estimate algorithm given an A/R gene state. We now propose a model where we first estimate the state of the A/R genes and then given that state we use the least squares approach to estimate the a values which remain. Thus if a specific A/R gene is in a 0 state we then zero out its effecting a value and estimate the remaining a values as we would have done before.

The mathematical analysis to justify the algorithm uses the MAP or maximum a posteriori estimate approach.[31] Specifically, we try to maximize the following:

$$\frac{\partial \ln\ p_{a/z}(A/Z)}{\partial A} = 0$$

is the MAP approach. Van Trees shows the equivalence to the MMSE approach or the Bayesian analysis.[32] Clearly the Z variable are binary, namely the on/off, estimates and continuous, namely if on what value. Thus we write the MAP as:

$$\frac{\partial \ln\ p_{a/z_C,z_D}(A/Z_{Continuous},Z_{Discrete})}{\partial A} = 0$$

is;

$$\frac{\partial \ln\ p_{a/z_c} p_{z_c/z_D}}{\partial A}$$

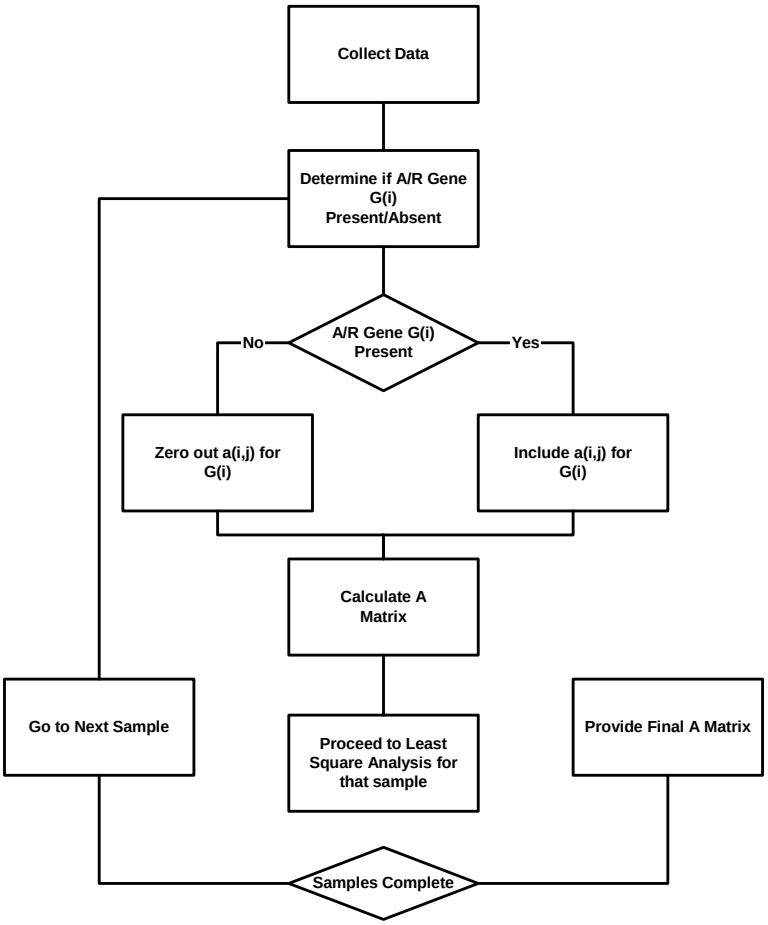
clearly from the above we can separate the two optimizations, first estimate the binary of digital, D, elements and then use the standard approach to obtain the continuous, C, elements. The best estimator for the digital part is a standard MAP estimator using a threshold.

The algorithm for the calculation is shown below. We perform the functions as we stated above.[33]

[31] See Van Trees, Detection, Estimation and Modulation Theory, Wiley (New York) 1968.

[32] Van Trees pp. 56-58.

[33] It should also be noted that we can eliminate outlier data as described in McGarty, Bayesian Outlier Rejection and State Estimation, IEEE AC 1975, pp. 682-687. In this work the author allows for the process of estimation plus the optimal rejection of outliers from the estimates. Performance bounds are also provided.



10.5 Model Variants

We now want to consider several variants on how color can be generated. In our model we assumed that Expression constants and drivers remained the same throughout but that they were turned on and off thus generating differing colors. However there are many possibilities,

1. Expression Constants Vary: In this case the a values vary from color to color. From a gene expression perspective we cannot find this an attractive alternative. However it may be conceivable that there are secondary controller A/R genes which may be playing a role which may be unidentified at the time of the experimental analysis.

2. Expression Drivers Vary: The u values we have used to be the steady state drivers may be affected by various factors, including local factors such as plant acidity and location. We have assumed we know these variables. This assumption is based upon some past experimental analyses. However these may vary and must be taken into account. To do so we can expand the model to estimate them as well.

3. Expression Constants are on/off: This is the model we have developed herein. This model assumes that all variable are constant and that we have just an on and off process of A/R genes.

4. Expression Drivers are on/off: This is an intriguing alternative with no known physical embodiment but it may be the case from time to time. However the effect is the same for A/R expression as if we assumed the constants went on or off.

Any combination of the above may also occur.

11 REFERENCES

1. Ando, S., H. Iba, *Quantitative Modeling of Gene Regulatory Network*, Univ Tokyo.
2. Andronov, A., et al, **Theory of Oscillators**, Dover (New York) 1987.
3. Atkins, P. **Physical Chemistry**, Freeman (New York) 1990.
4. Bailey, L.H., et al, **Hortus Third**, Macmillan (New York) 1976.
5. Baker, C.A.H., et al, *GeneVis: Visualization Tools for Genetic Regulatory Network Dynamics*, Univ Calgary.
6. Bartei, B., S. Matsuda, *Seeing Red*, Science, Vol 299, 17 Jan 2003, pp 352-353.
7. Brennan, J. R., *The Chromosomes of Hemerocalis*, Daylily Journal, Vol 47 No 1 1992, pp. 73-77.
8. Brodacki, A., at al, *Genetic Distances in Hens Estimated with Protein Genes*, Jrl of Polish Agr Univ, Vol 6 No 2 2003
9. Campbell, A., L. Heyer, **Genomics, Proteomics, and Bioinformatics**, Benjamin Cummings (New York) 2003.
10. Causton, H. et al, **Microarray Gene Expression and Analysis**, Blackwell (Malden, MA) 2003.
11. Chapman, S., et al *Evaluating Plant Breeding Strategies by Simulating Gene Action*, Agronomy 2003 pp 99-113.
12. Chen, T., et al, *Modeling Gene Expression with Differential Equations*, Pacific Symposium on Biocomputing, 1999 pp. 29-40.
13. Chung, M., J. Noguchi, *Geographic spatial correlation of morphological characters of Hemerocalis middendorfii complex*, Ann Bot Fennici Vol 35, 1998, pp. 183-189.
14. Chung, M., *Spatial Structure of three Populations of Hemerocalis hakuuensis*, Bot. Bull. Acad. Sci., 2000, Vol 41, pp. 231-236.
15. Crampin, E., et al, *Extracting Biochemical Reaction Kinetics from Time Series Data,* LNAI 3214 pp 329-336 2004 Springer Verlag.
16. Cunningham, W., **Nonlinear Analysis**, McGraw Hill (New York) 1958.
17. Dahlgren, R.M.T., T**he Families of Monocotyledons**, Springer (New York) 1985.
18. Dey, P. M., J. B. Harborne, **Plant Biochemistry**, Academic Press (New York) 1997.
19. Doyle, F., D. Lauffenberger, *Network Interference*.
20. Dressler, D., H. Potter, **Discovering Enzymes**, Scientific American Press (New York) 1991.
21. Dunn, G., B. Everett, **Mathematical Taxonomy**, Dover (New York) 2004.
22. Durrett, H., **Color**, Academic Press (New York) 1987.

23. Erhardt, W., **Hemerocalis**, Timber Press (Portland, OR) 1992.

24. Federoff, N., *Transposons and Genome Evolution in Plants*, PNAS, Vol 97, June 2000, pp 7002-7007.

25. Gardenes, J., et al, On the robustness of complex heterogeneous gene expression networks, Biophys Chem, 118 2005 pp. 225-228.

26. Gatlin, F., *Daylilies, A Fifty Year Affair*, AHS Press (Edgerton, MO), 1995.

27. Goodwin, T.W., **Chemistry and Biochemistry of Plant Pigments**, Vols 1 and 2, Academic Press (New York) 1976.

28. Grenfell, D., **Daylilies**, Timber Press (Portland, OR) 1998.

29. Griffiths, A., et al, **Genetic Analysis 5th Ed**, Freeman (New York) 1993.

30. Hatzimanikatis, V., *Dynamical Analysis of Gene Networks Requires Both mRNA and Protein Expression Information*, Metabolic Engr, Vol 1, 1999, pp. 275-281.

31. Holton, T., E. Cornish, *Genetics and Biochemistry of Anthocyanin Biosynthesis,* The Plant Cell, Vol 7, 1995, pp 1071-1083.

32. Jaakola, L. et al, *Expression of Genes Involved in Anthocyanin Biosynthesis*, Plant Physiology, Vol 130 Oct 2002, pp 729-739.

33. Judd, D. B. et al, **Color** 2ND Edition, Wiley (New York) 1963.

34. Kato, M., et al, *Accumulation of Carotenoids and Expression of Carotenoid Biosynthetic Genes during maturation of Citrus Fruits*, Plant Physiology, Feb 2004 Vol 132 pp. 824-837.

35. Kohane, I., et al, **Microarrays for an Integrative Genomics**, MIT Press (Cambridge) 2003.

36. Krane, D., M. Raymer, **Bioinformatics**, Benjamin Cummings (New York) 2003.

37. Krubasik, P. et al, *Expression and Functional Analysis of Gene Cluster Involved in the Synthesis of Decaprenoxxanthin reveals The Mechanism for c50 Carotenoid Formation*, Eur J Biochem Vol 268 pp 3702-3708, Feb 2001.

38. Laurentin, H., P. Karlovsky, *Genetic Relationship and Diversity in a Sesamum indicum using AFLP*, BMC Genetics, Vol 7 No 10 Feb 2006.

39. Lensnaw, J., S. Ghabrial, *Tulip Breaking*, Plant Disease, Vol 84, 2000, pp 1052-1060.

40. Lesk, A., **Bioinformatics**, Oxford (New York) 2002.

41. Levi, L., **Applied Optics**, Wiley (New York) 1968.

42. Liaaen-Jensen, S., *Carotenoids of Lower Plants*, Pure and Applied Chemistry, Vol 57 No 6, pp 649-658 1985.

43. Ljung, L. System Identification, Prentice Hall (Englewood Cliffs) 1987.

44. McGarty, T., **Stochastic Systems and State Estimation**, Wiley (New York) 1974.

45. McMurry, J., Begley, T., **The Organic Chemistry of Biological Pathways**, Roberts & Company Publishers, 2005.

46. Misawa, N. et al, *Elucidation of Erwinia uredovora Carotenoid Biosynthetic Pathway*, Jrl Bacteriology, Dec 1990 pp 6704-6712, Vol 172 No 12.

47. Mjolsness, E., et al *Gene Regulation Networks with Signaling and Cell Division in Development Satiations*, IEEE Biomed Engr Houston TX Oct 2002.

48. Mohr, H., P. Schopfer, **Plant Physiology**, Springer (New York) 1995.

49. Mol, J, et al, *How Genes Paint Flowers and Seeds, Trends in Plant Science*, Vol 3 June 1998, pp 212-217.

50. Mol, J., et al, *Novel Colored Plants*, Current Opinion in Biotechnology, Vol 10, 1999, pp 198-201.

51. Mueller, U., L. Wolfenbarger, *AFLP Genotyping and Fingerprinting*, Tree, Vol 14 No 10, Oct 1999 pp. 389-394.

52. Munson, R., **Hemerocalis, The Daylily**, Timber Press (Portland, OR) 1989.

53. Murray, J., **Mathematical Biology**, Springer (New York) 1989.

54. Murrell, J., *Understanding Rate of Chemical Reactions*, University of Sussex.

55. Naik, P. S., et al, *Genetic manipulation of carotenoid pathway in higher plants*, Current Science, Vol 85, No 10, Nov 2003, pp 1423-1430.

56. Noguchi, J., H. De-yuan, *Multiple origins of the Japanese nocturnal Hemerocalis citrina*, Int Jrl Plant Science, 2004, Vol 16, pp. 219-230.

57. Peat, J.P., T.L. Petit, **The Daylily**, Timber Press (Portland, OR) 2004.

58. Percus, J., **Mathematics of Genome Analysis**, Cambridge (New York) 2004.

59. Perkins, T., et al, *Inferring Models of Gene Expression Dynamics*, Journal of Theoretical Biology, Vol 230, 2004, pp. 289-299.

60. Petit, T., J. Peat, **The Color Encyclopedia of Daylilies**, Timber Press (Portland, OR) 2000.

61. Rubinow, S., **Mathematical Biology**, Dover (New York) 2002.

62. Sage A. P., J. Melsa, **System Identification**, Academic Press (New York) 1971.

63. Schabell, J., *The Ancestor*, Daylily Journal Vol 45 No 2 1990, pp. 159-160.

64. Schabell, J., *The Daylily, 5000 Years of Glory*, Daylily Journal Vol 45 No 4 1990, pp. 348-353.

65. Schabell, J., *The Historical Species, Conclusion*, Daylily Journal Vol 50 No 1 1995, pp. 48-53.

66. Schabell, J., *The Historical Species, H aurantiaca*, Daylily Journal Vol 49 No 1 1994, pp. 96-101.

67. Schabell, J., *The Historical Species, H aurantiaca*, Daylily Journal Vol 49 No 2 1994, pp. 224-228.

68. Schabell, J., *The Historical Species, H citrina*, Daylily Journal Vol 47 No 2 1992, pp. 168-170.

69. Schabell, J., *The Historical Species, H dumortierii*, Daylily Journal Vol 47 No 3 1992, pp. 283-286.

70. Schabell, J., *The Historical Species, H forestii* et al, Daylily Journal Vol 49 No 4 1994, pp. 387-391.

71. Schabell, J., *The Historical Species, H fulva rosea*, Daylily Journal Vol 46 No 2 1991, pp. 163-165.

72. Schabell, J., *The Historical Species, H fulva*, Daylily Journal Vol 46 No 1 1991, pp. 42-44.

73. Schabell, J., *The Historical Species, H fulva*, Daylily Journal Vol 46 No 4 1991, pp. 388-391.

74. Schabell, J., *The Historical Species, H fulva, Doubles*, Daylily Journal Vol 46 No 3 1991, pp. 239-243.

75. Schabell, J., *The Historical Species, H lilioasphodelus*, Daylily Journal Vol 48 No 2 1993, pp. 174-177.

76. Schabell, J., *The Historical Species, H middendorfii*, Daylily Journal Vol 48 No 1 1993, pp. 59-61.

77. Schabell, J., *The Historical Species, H minor, H multiflora, H altissima*, Daylily Journal Vol 49 No 3 1994, pp. 310-316.

78. Schabell, J., *The Historical Species, H thunbergii*, Daylily Journal Vol 48 No 3 1993, pp. 300-303.

79. Schabell, J., *The Over Time and Space*, Daylily Journal Vol 45 No 3 1990, pp. 282-284.

80. Schnell. S, T. Turner, *Reaction Kinetics in Intracellular Environments with Macromolecular Crowding*, Biophys and Molec Bio vol 85 2004 pp. 235-260.

81. Simon, W., **Mathematical Techniques for Biology and Medicine**, Dover (New York) 1986.

82. Springer, P., *Gene Traps*, The Plant Cell, Vol 12 July 2001, pp. 1007-1020.

83. Stout, A.B., **Daylilies**, Saga Press (Millwood, NY) 1986.

84. Szallasi, Z. **System Modeling in Cellular Biology: From Concepts to Nuts and Bolts**. MIT Press (Cambridge) 2006.

85. Taiz, L., E. Zeiger, **Plant Physiology**, Benjamin Cummings (Redwood City, CA) 1991.

86. Tinoco, I. et al, **Physical Chemistry**, Prentice Hall (Englewood Cliffs, NJ) 1995.

87. Tobias, A., *Directed Evolution of Biosynthetic Pathways to Carotenoids with Unnatural Carbon Bonds*, PhD Thesis, Cal Tech, 2006.

88. Tomkins, J. R., *DNA Fingerprinting in Daylilies*, Part I, Daylily Journal, Vol 56 No 2 2001, pp. 195-200.

89. Tomkins, J. R., *DNA Fingerprinting in Daylilies*, Part II, Daylily Journal, Vol 56 No 3 2001, pp. 343-347.

90. Tomkins, J. R., *How much DNA is in a Daylily*, Daylily Journal, Vol 58 No 2 2003, pp. 205-209.

91. Tomkins, J., et al, *Evaluation of genetic variation in the daylily (Hemerocalis) using AFLP markers*, Theor Appl Genet Vol 102, 2001, pp. 489-496.

92. Turing, A., *The Chemical Basis of Morphogenesis*, Phil Trans Royal Soc London B337 pp 37-72, 1959.

93. Vohradsky, J., *Neural Network Model of Gene Expression*, FASEB Journal, Vol 15, March 2001, pp. 846-854.

94. Watson, J., et al, **Molecular Biology of the Gene**, Benjamin Cummings (San Francisco) 2004.

95. Wessler, S., *Plant Transposable Elements*, Plant Physiology, January 2001 Vol 125, pp. 149-151.

96. Winkel-Shirley, B., *Flavonoid Biosynthesis*, Plant Physiology, Vol 126 June 2001 pp 485-493.

97. Wu., F.X. et al, *Modeling Gene Expression From Microarray Expression Data with State Space Equations*.

98. Zak, D., *Continuous Time Identification of Gene Expression Models,* Journal of Integrative Biology, Vol 7 No 4 2003 pp 373-386.

12 WEB SITES

http://www.ibiblio.org/pfaf/cgi-bin/arr_html?Hemerocallis++species&CAN=LATIND

http://www.ibiblio.org/pfaf/cgi-bin/arr_html?Hemerocallis+hakunensis&CAN=LATIND

http://www.ibiblio.org/pfaf/cgi-bin/arr_html?Hemerocallis+citrina&CAN=LATIND

This is the Poldeck site. It is excellent for structure and photos:

http://www.hemerocallis-species.com/HS/HS_homee.htm

http://www.hort.purdue.edu/rhodcv/hort640c/secprod/se00013.htm

http://www.daylilies.org/daylilies.html

http://pubs.caes.uga.edu/caespubs/pubcd/C545.htm

http://historicdaylily0.tripod.com/

This is the National Center for Biological Information, under NLM (National Library of Medicine), which has significant sequencing information for Hemerocallis:

http://www.ncbi.nlm.nih.gov/BLAST/ and specifically
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=42415328

http://www.nybg.org/