

## **Distribution Agreement**

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

Signature:

---

Yi-Han Lin

---

Date

THE BROAD-HOST-RANGE PATHOGENESIS OF *AGROBACTERIUM*  
*TUMEFACIENS*: SUCCESSFUL COUPLING THE MOTIONS AND DOMAIN  
INTERACTIONS WITHIN VIRA TO INTEGRATE SIGNAL SENSING

By

Yi-Han Lin  
Doctor of Philosophy

Chemistry

---

Dr. David G. Lynn  
Advisor

---

Dr. Justin P. Gallivan  
Committee Member

---

Dr. Stefan Lutz  
Committee Member

Accepted:

---

Lisa A. Tedesco, Ph.D.  
Dean of the James T. Laney School of Graduate Studies

\_\_\_\_\_ Date

THE BROAD-HOST-RANGE PATHOGENESIS OF *AGROBACTERIUM*  
*TUMEFACIENS*: SUCCESSFUL COUPLING THE MOTIONS AND DOMAIN  
INTERACTIONS WITHIN VIRA TO INTEGRATE SIGNAL SENSING

By

Yi-Han Lin  
B. S., National Taiwan University, 2003

Advisor: David G. Lynn, PhD.

An abstract of  
A dissertation submitted to the Faculty of the  
James T. Laney School of Graduate Studies of Emory University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in Chemistry  
2011

## ABSTRACT

### THE BROAD-HOST-RANGE PATHOGENESIS OF *AGROBACTERIUM TUMEFACIENS*: SUCCESSFUL COUPLING THE MOTIONS AND DOMAIN INTERACTIONS WITHIN VIRA TO INTEGRATE SIGNAL SENSING

By Yi-Han Lin

*Agrobacterium tumefaciens* has provided a nice example of a wide-host-range pathogen mediating the only known inter-kingdom gene transfer strategy for transferring higher plants. The system employs a simple two-component system (TCS) scheme to perceive and integrate multiple plant-derived signals and regulates virulence. The VirA/VirG two-component protein adopts a conserved phosphorylation catalytic core, as other TCS proteins, but incorporates signal regulation by coordinating with three regulatory domains. In this dissertation, I have attempted to dissect the motions and molecular mechanisms of these regulatory domains. I started by analyzing the interface of the critical phenol sensing linker domain and the kinase core, and the data supports a rotational motion propagating through what appears to be a rigid helical bundle, spanning from the linker domain to the kinase. The phenol binding within the linker domain and signal specificity were probed, both computationally and genetically, and the data indicates a possible long-range motion connecting the helical interface to the proposed phenol binding site. I further analyzed the mechanisms that coordinate the other two regulatory domains: the periplasmic domain for sugar sensing and the C-terminal receiver domain for regulation of the phenol response. A critical convergence point regulating both phenol perception and sugar perception was identified at a single tyrosine residue, aa293. The role of this point in mediating the wide-host-range properties of the pathogen is discussed.

The findings in this dissertation shed light on protein evolution, grounded in the mechanism of inner-domain motion regulation and the long-range inter-domain coordination, to allow for this truly remarkable pathogen to emerge. New experimental approaches have emerged from this dissertation that may well define the plant signaling landscape and the intricate system-wide chemistry that occurs at the host-pathogen interface. This unique insight into plant-bacterial interaction continues to define the intricate cellular commitments important for pathogen resistance and eukaryotic cell development.

THE BROAD-HOST-RANGE PATHOGENESIS OF *AGROBACTERIUM*  
*TUMEFACIENS*: SUCCESSFUL COUPLING THE MOTIONS AND DOMAIN  
INTERACTIONS WITHIN VIRA TO INTEGRATE SIGNAL SENSING

By

Yi-Han Lin  
B. S., National Taiwan University, 2003

Advisor: David G. Lynn, PhD.

A dissertation submitted to the Faculty of the  
James T. Laney School of Graduate Studies of Emory University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in Chemistry  
2011

## ACKNOWLEDGEMENTS

Before I came to the United States to pursue my Ph.D degree, many people told me that Ph.D is a long journey that needs extremely strong dedication and persistence. Indeed, seven years were not a short time, so I would like to express my deepest gratitude to my advisor Prof. David Lynn. Without his mentorship, it would really have been a long journey as it seems. I would like to thank Dave to have inspired my science spirit, to guide me when I was confused and lost, and to always support me to mature to be a confident scientist. I started my graduate courses in Prof. Stefan Lutz's biochemistry class, and continued on Prof. Justin Gallivan's class of enzyme mechanisms. In both classes, I have learned greatly not only from the course material, but also on how to develop a good scientific proposal. I would also like to thank Prof. Stefan Lutz and Prof. Justin Gallivan for their guidance and supervision during my graduate study.

I joined the lab and worked with Dr. Rong Gao and Dr. Fang Fang on the "Agro" project since 2005. I was able to pick up this project quickly and started my own research because Rong and Fang have set up a solid foundation and brought up many interesting questions for me to pursue. I enjoyed working with them, and all the discussions we have had on this project. I would also like to thank the graduate students, undergraduates, and visiting students who have contributed to this project. I would also like to thank our collaborators Prof. Andrew Binns and Dr. Arlene Wise at University of Pennsylvania for their stimulating thoughts on this research project.

I enjoyed working in the Lynn Lab because people here are always helpful and friendly. The lab life was full of joy because we not only talk about science, but also have

fun together. I enjoyed every birthday celebrations and group trips we had. I have to thank many lab members, Anil, Yue, James, Savannah, Erin, Chenrui, Tolu, Jay, Junjun, Dibyendu, Lisa, Li, and Kaya. I also want to give special thanks to Anil, Savannah, James, Yue, and Kaya, for critically reviewing my dissertation. The graduate school life was also colorful because of many good friends. Seth, Kitty, Betül, James, Yue, Weilin and I joined the department together, and I would like to thank them for their friendship, support, and to grow up with me in scientific fields and in life. I would also like to thank my Taiwanese friends: Yi-Hung, I-Lin, Yu Tsao, Chia-Hung, and Howard, for their company and support during the past seven years.

I want to give big thanks to my friends Weilin and I-Lin. Weilin and I shared our experiences in cumulative exams, easy and uneasy experiments, annual reports, and proposal together. I would like to thank her for her friendship and support at all time. I also feel very lucky to have met my good friend, I-Lin, again, on the other side of the world, and to share with her the sweet and bitterness during graduate study and in my life. She is more than my friend, but more like a sister to me. I also want to thank my husband Fu for his love and support, and my son Ashton who made my graduate life even more memorable.

Finally, I would like to give my greatest appreciation to my parents. There were great moments and bad times during my graduate study, but no matter in what situation, Mom and Dad were always my biggest supporters. I want to thank them to have always been present in the big moments in my life. I also want to thank my Dad for his wisdom advises when I was in difficult times, and my Mom to have given me her endless and selfless love. I dedicate this work to them.

## TABLE OF CONTENTS

CHAPTERS	PAGE
1. THE PRINCIPLE OF <i>AGROBACTERIUM TUMEFACIENS</i> PATHOGENESIS AND THE VIRA/VIRG TWO-COMPONENT SYSTEM	1
<i>Agrobacterium</i> pathogenesis and the plant-derived signals	
The VirA/VirG TCS	
Phenol perception by the linker domain and the linker structure model	
2. MEDIATING THE ROTATIONAL MOTION BETWEEN THE PHENOL SENSING LINKER DOMAIN AND THE KINASE IN VIRA	25
2. Introduction	25
The HAMP domain	
Materials and Methods	32
Results	45
GCN4 fusions	
Library screen of constitutive on mutants in aa426-437 (linker $\alpha$ 4)	
Kinase coiled-coil insertion	
Discussion	73
3. GENETIC, COMPUTATIONAL, AND BIOPHYSICAL CHARACTERIZATION OF PHENOL PERCEPTION OF THE LINKER DOMAIN	80
Introduction	80
The GAF domain	
Phenol specificity of different <i>Agrobacterium</i> strains and Y293T	
Materials and Methods	94
Results	103
Computation modeling phenol binding in GAF	
Genetic analysis of phenol perception	



Regulation of phenol specificity at the helix region <i>In vitro</i> characterization of the GAF-like linker domain	
Discussion	129
4. INTEGRATING PHENOL AND SUGER PERCEPTION FOR THE BROAD-HOST-RANGE IN <i>AGROBACTERIUM TUMEFACIENS</i>	134
Introduction	134
Host-range determination of <i>Agrobacterium</i> The two library screening results	
Materials and Methods	139
Results	144
VirA(Y293F) VirA(Y293T)	
Discussion	157
5. SUMMARY	168
REFERENCES	177

## LIST OF ILLUSTRATIONS

FIGURE		PAGE
CHAPTER 1		
Figure 1.1	Induction of <i>virB::lacZ</i> in <i>Agrobacterium</i> by different phenols	6
Figure 1.2	Induction of <i>virB::lacZ</i> by different inducing sugars	7
Figure 1.3	The two-component signaling scheme	14
Figure 1.4	Available structures of the two-component proteins	16
Figure 1.5	Schematic diagram of VirA/VirG two-component signal transduction and domain organization	19
Figure 1.6	The linker domain is homologous to the GAF domain	23
CHAPTER 2		
Figure 2.1	Coiled-coil prediction of VirA-LK(285-711)	28
Figure 2.2	Sequence alignment of the VirA kinase domain aa(426-711) and TM0853	29
Figure 2.3	HAMP domain structure	31
Figure 2.4	aa426-437 represses VirA kinase activity	46
Figure 2.5	The yeast transcription factor GCN4	50
Figure 2.6	GCN4-fused 450 <sup>K</sup> and 450 <sup>KR</sup>	52
Figure 2.7	GCN4-fused 438 <sup>K</sup>	55
Figure 2.8	GCN4-fused 426 <sup>K(G665D)</sup>	58
Figure 2.9	Library screening for constitutive on mutation in linker $\alpha$ 4 (aa426-437)	63
Figure 2.10	$\beta$ -galactosidase activity of the constitutive on mutants in 426 <sup>K</sup>	65

Figure 2.11	The constitutive on mutant C435K is resulted from $\alpha 4$ - $\alpha 4'$ salt-bridge formation	66
Figure 2.12	The constitutive on mutation E430K is resulted from $\alpha 1$ - $\alpha 4$ salt-bridge formation	67
Figure 2.13	Direct amino acid insertion in kinase coiled-coil	70
Figure 2.14	The linker $\alpha 4$ constitutive ON mutations with the OFF kinase coiled-coil insertions	72
Figure 2.15	The two ON conformations in LZ-450 <sup>K</sup> and LZ-426 <sup>K</sup>	75
Figure 2.16	Proposed phosphorylation model of VirA and TM0853	78
CHAPTER 3		
Figure 3.1	Representative GAF domain architectures in different organisms	82
Figure 3.2	Crystal structure of GAF domains	83
Figure 3.3	Crystal structure of GAF domains binding with their ligands	87
Figure 3.4	Proposed phenol binding site in VirA and the “proton transfer” model	91
Figure 3.5	Phenol specificity and the library design	92
Figure 3.6	Parameters used in Autodock	95
Figure 3.7	Computational docking of phenolic compounds into the modeled linker structure	104
Figure 3.8	The docking of phenols to the linker structure generated by SAM-T06 method	108
Figure 3.9	Alanine scanning of the aromatic residues at the predicted phenol binding site	112
Figure 3.10	Engineering phenol specificity at the helices	117
Figure 3.11	<i>In vitro</i> characterization of the linker domain	122
Figure 3.12	Linker(W341A)	124

Figure 3.13	Intrinsic tryptophan fluorescence of the wild-type and mutant linkers	126
Figure 3.14	Chemical cross-link of the VirA linker and receiver domains	127
CHAPTER 4		
Figure 4.1	HDMBOA decomposition	136
Figure 4.2	The library screening probing linker-receiver interaction	140
Figure 4.3	AS induction in LKR(Y293F)	147
Figure 4.4	Phenol specificity of LKR(Y293F)	148
Figure 4.5	AS induction in PLKR(Y293F)	149
Figure 4.6	AS dose response of PLKR(Y293F)	150
Figure 4.7	Phenol specificity of PLKR(Y293F)	151
Figure 4.8	Glucose dose response of PLKR(Y293F)	152
Figure 4.9	DMP dose response of LKR(Y293T).	154
Figure 4.10	PLKR(Y293T)	155
Figure 4.11	The “piston-type” motion	159
Figure 4.12	Proposed model of sugar-mediated kinase activation transmitted through Y293	161
Figure 4.13	<i>vir</i> -box in <i>virA</i>	166
Figure 4.14	Fluorescence image of the plants co-cultivating with <i>Agrobacterium</i>	167
CHAPTER 5		
Figure 5.1	Phylogenetic tree of ten <i>Agrobacterium</i> strains	174
Figure 5.2	Proposed model of VirA/VirG domain coordination	175

TABLE		PAGE
Table 2.1	Bacterial strains and plasmids used in Chapter 2	33
Table 2.2	Primers used in Chapter 2	40
Table 3.1	Bacterial strains and plasmids used in Chapter 3	96
Table 3.2	Primers used in Chapter 3	98
Table 3.3	Chemical structures of phenolic inducers used in Chapter 3	99
Table 4.1	Bacterial strains and plasmids used in Chapter 4	142
Table 4.2	Primers used in Chapter 4	143

## ABBREVIATIONS

aa	amino acid
Ap	ampicillin
AS	acetosyringone
ASBr	$\alpha$ -bromo acetosyringone
ATP	adenosine-5'-triphosphate
AV	acetovanillone
bp	base pair
BCIP	5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt
BSA	bovine serum albumin
CA	catalytic ATP-binding domain
CD	circular dichroism
dd	double distilled
DHp	dimerization histidine phosphotransfer domain
DMP	2,6-dimethoxyphenol
DMS	dimethyl sulfoxide
x g	rcf or centripetal force of rotation
GA	guaiacol
GAF	cGMP-regulated cyclic nucleotide phosphodiesterases, <u>A</u> denylyl cyclases, and bacterial transcription factor <u>F</u> hlA
GFP	green fluorescence protein
hr	hour

HAMP	<u>H</u> istidine kinases, <u>A</u> denyl cyclases, <u>M</u> ethyl-accepting proteins, and <u>P</u> hosphatases
HAP	4-hydroxyacetophenone
HK	histidine kinase
HPt	histidine phosphotransfer domain
I. M.	induction medium
IPTG	isopropyl $\beta$ -D-thiogalactopyranoside
Km	kanamycin
kb	kilobases
kDa	kilo daltons
<i>lac</i>	lactose operon
LB	Luria-Bertani broth
LZ	leucine zipper
$\mu$ g	micro gram
$\mu$ L	micro liter
$\mu$ M	micro molar
mg	milli gram
mL	milli liter
mM	millimolar
MBP	maltose binding protein
MCP	methyl-accepting chemotaxis protein
MES	2-(4-Morpholino)ethanesulfonic acid
min	minute
MS	mass spectrometry

NBT	nitro-blue tetrazolium chloride
Ni-NTA	nickel-nitriloacetic acid
NMR	nuclear magnetic resonance
OD	optical density
OH-AS	$\alpha$ -hydroxyacetosyringone
PAGE	polyarylamide gel electrophoresis
PAS	<u>P</u> er- <u>A</u> rrnt- <u>S</u> im
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PDE	phosphodiesterases
r	resistance
REC	receiver domain
RFP	red fluorescence protein
RR	response regulator
RT	room temperature
rpm	rotations per minute
SAE	syringaldehyde
SDS	sodium dodecylsulfate
SPA	sinapinic acid
Spec	spectinomycin
TBE	Tris-borate-EDTA
TBS	Tris buffered saline



TCS	two-component signaling system
T-DNA	transferred DNA
Tc	tetracyclin
Ti	tumor inducing
TM	transmembrane
Tris	tris-(hydroxy methyl) aminomethane
<i>vir</i>	virulence regulon
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

**CHAPTER 1**  
**THE PRINCIPLE OF *AGROBACTERIUM TUMEFACIENS* PATHOGENESIS**  
**AND THE VIRA/VIRG TWO-COMPONENT SYSTEM**

“Life will always find a way,” said Dr. Ian Malcolm in *Jurassic Park* (1993). Evolution holds true as individuals, facing the pressure of natural selection and evolution, seek to survive. For a pathogen to survive, it has to specifically recognize, attach to, and colonize a host, always overcoming the host’s innate defense responses. The success of the pathogen and the host-pathogen interaction has been studied for decades, leading to huge changes in our understanding of the design of pathogenesis strategies and the development of better drugs. Clearly, host determination, which triggers the commitment of pathogenesis, is tightly controlled and very energy intensive. The host range of a pathogen can be very broad or as narrow as a specific organ, tissue, or even a cell type, and if we could control this most fragile element of host determination, pathogenesis could be regulated.

The Gram-negative *Agrobacterium tumefaciens* is a soil-borne, broad-host-range plant pathogen that mediates the only known inter-kingdom gene transfer mechanism in dicotyledonous plants (Gelvin 2000). The gene transfer mediates crown gall tumor formation and the synthesis of opines, molecules that can serve as the sole carbon and nitrogen source for the bacterium. Because of the natural mechanism of mediating lateral gene transfer, *Agrobacterium* has been widely engineered and applied in food and agricultural science. Even though its native host range appears to be limited to the

dicotyledonous plants, the bacterium has also been successfully engineered to transform cereal crops, such as rice, wheat, and maize (Shrawat and Lörz 2006; Goetz, Christine et al. 2009). Further genetic transformation of fungus and animal cells has been demonstrated under laboratory conditions (Kunik, Tzfira et al. ; Bundock, den Dulk-Ras et al. 1995; de Groot, Bundock et al. 1998). These results establish that the machinery for broad host recognition exists in *Agrobacterium tumefaciens* and that host range is likely defined by early signal perception.

### ***Agrobacterium* pathogenesis and the plant-derived signals**

*Agrobacterium* pathogenesis is initiated via the perception of plant-derived molecules at the wound site. Although it is still unclear whether the wounding of the plant is absolutely required for gene expression (Brencic, Angert et al. 2005), some of the molecules released from the host plant have been characterized. Monosaccharides and phenols, two precursors of plant cell wall formation, and a general acidic environment (~ pH 5.5), a characteristic of plant wounding, in combination elicit *Agrobacterium* pathogenesis. Those signaling molecules are either directly or indirectly sensed by VirA, a histidine autokinase of *Agrobacterium*. The so-called two-component signaling system (TCS) involves the interaction of VirA with its cognate response regulator, VirG. The *virA* and *virG* genes, together with more than twenty other virulence genes (e.g. *virB1-11*, *virC1*, *virD1-2*), are encoded in the *vir* operon of the ~200,000 bp tumor-inducing (Ti) plasmid in *Agrobacterium*. Those virulence genes, including *virA* and *virG* themselves, are up-regulated by the signals to mediate gene transfer. The transferred DNA (T-DNA), encoding the oncogenic genes, is excised from the Ti-plasmid, processed and packaged

by multiple Vir proteins, and transported into the plant host via the type IV secretion system (Christie 2004; Li, Wolf et al. 2005). Finally, the T-DNA is targeted to the plant nucleus, and integrated within the genome (for reviews see (Gelvin 2000; Zhu, Oger et al. 2000; Tzfira and Citovsky 2002; Gelvin 2003).

## 1. Phenol

Plant phenol perception likely determines *Agrobacterium*'s host range, but this structural class, derived largely from the plant cell wall, contains the largest structural diversity of plant secondary metabolites. The phenol structures were first identified from co-cultivation of *A. tumefaciens* with tobacco root cells, and two potent phenolic inducers, acetosyringone (AS) and  $\alpha$ -hydroxyacetosyringone (OH-AS), were identified in root exudates (Stachel, Messens et al. 1985). These same compounds were also extracted from the plant leaf. More than fifty phenols with different functional substitutions have now been identified as active in *vir* gene induction (Fig. 1.1)(Stachel, Messens et al. 1985; Melchers, Regensburg-Tuïnk et al. 1989; Duban, Lee et al. 1993), and the results indicate that each phenol has a wide range of effective concentrations. The active structures generally contain the following characteristics: a phenol hydroxyl group, one methoxy, but two methoxy substituents positioned *ortho* to the phenol hydroxyl significantly enhance *vir* induction, and a *para* substituent that can be accommodated by many different alkyl substituents (Melchers, Regensburg-Tuïnk et al. 1989; Duban, Lee et al. 1993). Some of these phenols are known to be products or intermediates of the lignin phenylpropanoid pathway in plants (Dixon, Achnine et al. 2002). Based on the structural characteristic of the active phenols, a "proton transfer" model has been proposed for VirA

activation (Hess, Dudley et al. 1991; Duban, Lee et al. 1993; Lee, Tzeng et al. 1993). Different natural and synthetic phenols have been tested (Lee 1997), but a better understanding of the phenol receptor is now required to further examine the proton transfer model and test other possible mechanisms involved in phenol induction.

## 2. pH

A study by Stachel et al. indicated that *vir* induction has an optimal pH between 4.8 and 6.0 (Stachel, Nester et al. 1986), and this requirement has been further examined in other labs (Gao and Lynn 2005; Yuan, Liu et al. 2007). Since plant vacuoles are normally acidic, a low pH could be diagnostic of the sap at the wound site, and indeed the response to pH is complex. Early studies indicated that the second promoter region (P2) of *virG* is up-regulated by an acidic pH (Winans 1990; Mantis and Winans 1992). Another study found that ChvG/ChvI, a chromosomally encoded two-component system of *Agrobacterium*, activates sets of genes, including *virB* and *virE*, in response to low pH (Li, Jia et al. 2002). And finally, the histidine kinase VirA can also be induced by acidic pH through the sugar binding protein ChvE (Chang and Winans 1992; Gao and Lynn 2005). Therefore, low pH seems to be a co-activator that works in concert with the phenols and the sugars to regulate host response.

## 3. Sugar

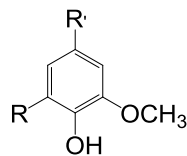
Monosaccharides were later found to serve as effective co-inducers of *Agrobacterium* virulence, and the range of active structures proved to be quite large (Ankenbauer and Nester 1990; Shimoda, Toyoda-Yamamoto et al. 1990). Among the

sugars characterized, the acidic monosaccharide derivatives, such as D-glucuronic acid and D-galacturonic acid, were the most potent inducers of *vir* gene expression (Fig. 1.2), and both of these sugars are components of plant cell walls (Ankenbauer and Nester 1990). The sugar receptor in *Agrobacterium*, ChvE, is now well characterized with low pH being responsible for sugar binding (Cangelosi, Ankenbauer et al. 1990). The *chvE* gene is encoded on the chromosome, and sequence comparisons suggested that ChvE belongs to a family of periplasmic sugar-binding proteins. Genetic studies confirmed the interaction between ChvE and the VirA periplasmic domain (Shimoda, Toyoda-Yamamoto et al. 1993; Doty, Yu et al. 1996), and mutations that decouple sugar and pH sensing were also identified (Gao and Lynn 2005). The structure of ChvE has been recently solved, helping to explain the previously identified residues that mediate sugar binding, map the interaction interface between ChvE and VirA, and better define sugar-mediated VirA activation (He, Nair et al. 2009).

Among the three *vir*-inducing signals, the phenols are indispensable. Neither the sugar nor acidic pH induce *vir* expression independently, but their presence can enhance phenol-induced *vir* expression, even though such effects are attenuated when phenols are added at high concentrations (Ankenbauer and Nester 1990). Another synergistic effect of sugar and acidic pH is to broaden the phenol recognition profile (Peng, Lee et al. 1998). Unlike sugar and acidic pH, which are sensed by a separate protein ChvE, phenol sensing is believed to be perceived directly by the histidine autokinase VirA. Even though the affinity labeling using isotope-incorporated phenolic analogs and affinity

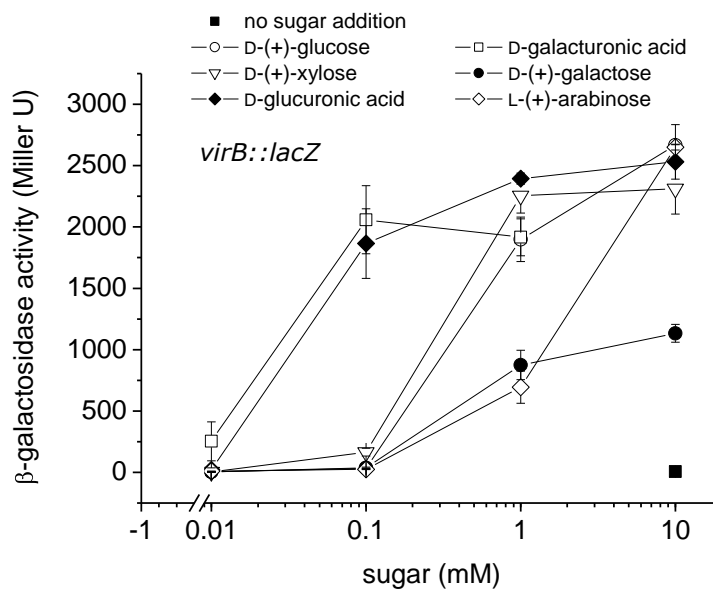
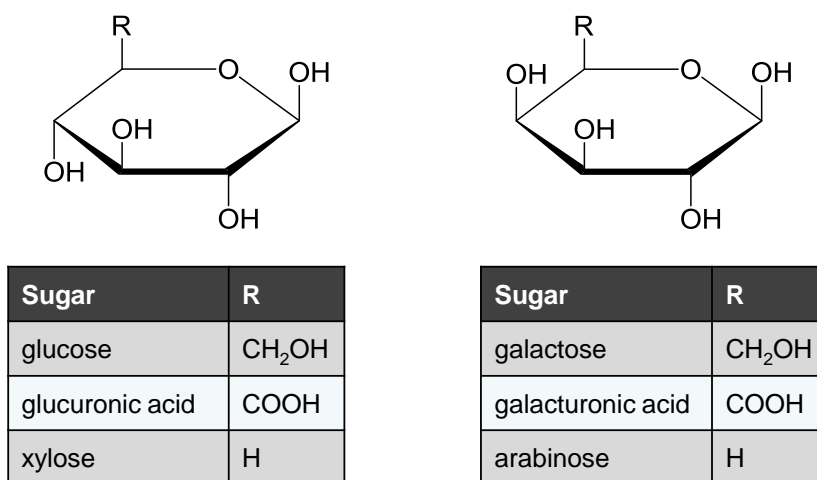
**Fig. 1.1 Induction of *virB::lacZ* in *Agrobacterium tumefaciens* by different phenols.**

The chemical structures and relative activities of phenol inducers, compiled from (Duban, Lee et al. 1993) are shown. The reported % activity was based on the  $\beta$ -galactosidase assay using 200  $\mu$ M phenolic inducers.



Phenol	R	R'	Reported % activity
acetosyringone (AS)	OCH <sub>3</sub>	COCH <sub>3</sub>	100
acetovanillone	H	COCH <sub>3</sub>	23
syringic acid	OCH <sub>3</sub>	COOH	23
coniferyl alcohol	H	CH=CHCH <sub>2</sub> OH	91
synapinic acid	OCH <sub>3</sub>	CH=CHCOOH	53
ferulic acid	H	CH=CHCOOH	16
2,6-dimethoxyphenol	OCH <sub>3</sub>	H	73

**Fig. 1.2 Induction of *virB::lacZ* by different inducing sugars.** As reported by (Ankenbauer and Nester 1990), the chemical structures of the sugars and their respective  $\beta$ -galactosidase activity in 2.5  $\mu$ M acetosyringone supplemented with 0.5% glycerol are shown.





chromatography did not successfully provide direct physical evidence (Lee, Dudley et al. 1992; Dye and Delmotte 1997), Lee et al. found that the phenol recognition profile is defined by the Ti-plasmid, even when transformed into different chromosomal backgrounds (Lee, Jin et al. 1995; Lee, Jin et al. 1996). Further, phenol inducibility can be successfully reconstituted in heterologous bacteria, *E. coli*, by expressing *virA*, *virG*, and *rpoA*, the gene encoding the  $\alpha$  subunit of *Agrobacterium* RNA polymerase (Lohrke, Nechaev et al. 1999; Lohrke, Yang et al. 2001). The sugar effect can also be reconstituted by moving in *chvE* (Lohrke, Yang et al. 2001). While the combined evidence supports VirA as the direct phenol sensor, the mechanism of how multiple inputs are mechanistically integrated through a single VirA receptor to signal broad-host-range pathogenesis has not been defined.

### **The VirA/VirG TCS**

The histidine kinase/response regulator two-component system (TCS) is one of the best examples of how simple elements can be used for many functions in an ever-changing environment. Since its discovery almost 20 years ago (Parkinson and Kofoid 1992), TCS has been extensively studied and reviewed (for reviews see (Robinson, Buckler et al. 2000; Stock, Robinson et al. 2000; West and Stock 2001; Gao and Stock 2009). A total of ~50,000 two-component proteins have been identified as being used across all three kingdoms of life. In bacteria, the number of TCS proteins identified ranges from 62 in *E. coli* and 27 in *Streptococcus pneumoniae*, to none in *Mycoplasma genitalium* and *Methanococcus jannaschii* (Robinson, Buckler et al. 2000). Likewise, while a few examples of TCS proteins exist in yeast and in plants, none have been found

in animals. Consequently, the prevalence of TCS in bacteria suggests a potential target for antibiotic development (Lin, Gao et al. 2008).

The functions controlled by different TCS varies from cell mobility to sporulation and biofilm formation, and each of these various pathways are regulated by different signals, ranging from small molecules, to temperature, and light. The signal transduction pathway involves a histidine sensor kinase (HK), which auto-phosphorylates at a conserved histidine residue. The phosphate can be transferred to a conserved aspartate of its cognate response regulator (RR). This RR contains an effector domain, which, upon activation by the HK, promotes the corresponding response (Fig. 1.3). Multiple phosphoryl-transfer reactions are also possible (Fig. 1.3B)(Hoch 2000). One of the best characterized phospho-relay system is the KinA→Spo0F→Spo0B→Spo0A pathway, controlling the sporulation in *Bacillus subtilis* (Stephenson and Lewis 2005). Such a pathway offers multiple points of control for tighter regulation. Progress in TCS research has been significantly enhanced with the structures solved of the HK and RR. More than 200 TCS protein structures are now available, and their function, the HK-RR interaction, and the signal regulation for activation are beginning to emerge.

#### 1. Histidine kinase (HK)

The histidine autokinase, or histidine sensor kinase, usually couples a conserved kinase core, the dimerization histidine phosphotransfer domain (DHp), with diverse signal sensing modules. Such design allows a simple phosphorylation reaction to be regulated by various inputs. These autokinases function as dimers, which form through a four-helix bundle architecture (Fig. 1.4A) that contains the conserved phosphorylated

histidine in  $\alpha 1$ . The direct helical linkage of the DHp domain to the upstream signal sensing domain implies a direct mechanical coupling for signal transmission (Szurmant, White et al. 2007; Tetsch and Jung 2009), which will be discussed in Chapter 2. Compared to the all- $\alpha$  helical structure in the DHp domain, the catalytic ATP-binding domain (CA) is composed of an  $\alpha/\beta$  structure and exists as a distinct domain (Fig. 1.4B). The CA domain contains four conserved motifs, N, G1, F, G2 boxes, for ATP binding. The ATP-bound CA domain phosphorylates the conserved histidine in the DHp domain of the alternate subunit of the dimer (Ninfa, Atkinson et al. 1993), and is referred to as a *trans*-phosphorylation mechanism, most readily seen in the HKs such as EnvZ and NtrB (Cai and Inouye 2003). However, the crystal structure of the first entire cytoplasmic histidine kinase, HK0853 of *Thermotoga maritima* (Fig. 1.4C), was best explained by a different *cis*-phosphorylation mechanism (Marina, Waldburger et al. 2005; Casino, Rubio et al. 2009). More details about the phosphorylation mechanism of HK and VirA will be discussed in Chapter 2.

The conserved kinase domain appears to be regulated by various environmental stimuli for signal sensing through input domains that can be either covalently attached to the kinase core or function as separate modules, through which signals originating outside the cell or within the cytoplasm can all be coupled. For example, in bacterial chemotaxis (Falke and Hazelbauer 2001; Wadhams and Armitage 2004), the CheA histidine kinase is coupled with different receptors, each responding to different nutrients (e.g. Tar and Trg sense aspartate and sugar, respectively). While those signal sensing domains have little similarity in primary sequence, some of these signal receptors have been grouped into protein families according to their tertiary structures. One of the largest

families identified is the PAS (Per-Arnt-Sim) domain. The PAS domains adopt similar folds with very high structural plasticity, and are widely distributed as modules in various signaling pathways (Taylor and Zhulin 1999; Hefti, Francoijs et al. 2004). From the SMART database, nearly 33% of HK are identified with at least one PAS domain (Gao and Stock 2009). An example is oxygen sensing in nitrogen fixing bacteria, such as the FixL/FixJ pathway in *Bradyrhizobium japonicum* (Perutz, Paoli et al. 1999; Gilles-Gonzalez and Gonzalez 2004). This PAS domain, coordinated with a heme group (Fig. 1.4D), binds to oxygen and regulates FixL phosphorylation activity. Another regulation motif commonly observed in HK is the HAMP domain (identified in Histidine kinases, Adenyl cyclases, Methyl-accepting proteins, and Phosphatases), which has 31% occurrence (Gao and Stock 2009). The HAMP domain has been widely implicated in mediating signal transmission by its coiled-coil architecture. More details of the HAMP domain and the possible signal transmission mechanisms will be discussed in Chapter 2.

## 2. Response regulator (RR)

The response regulator is also composed of two parts, a receiver domain containing the conserved aspartate residue as the phosphate acceptor and an effector domain for eliciting the output response. The receiver domain, similar to the DHp and CA domains in the HKs, is conserved in primary sequence. The structures of the receiver domains from different TCS also reveal a similar  $(\alpha/\beta)_5$  fold structure (Fig. 1.4E)(Madhusudan, Zapf et al. 1997; Robinson, Wu et al. 2003; Toro-Roman, Mack et al. 2005). Upon phosphorylation of the receiver domain, the conformation changes and activates the effector domain (Cho, Lee et al. 2000; Lee, Cho et al. 2001). The effector

domain has a broad functional scope, similar to that seen for the signal input domains throughout HKs. These functions vary from DNA and RNA binding to enzymatic activity, but most of the effector domains bind DNA and act as transcription regulators (Gao and Stock 2009).

### 3. VirA/VirG two-component system

The *virA* and *virG* genes were first identified using *vir::lacZ* induction from transposon insertional *vir* regulon mutants (Stachel, An et al. 1985; Stachel and Zambryski 1986), and were later categorized as the two-component regulatory proteins (Charles, Jin et al. 1992). The VirA autokinase contains 829 residues with the conserved histidine located at aa474. The response regulator VirG has 241 residues with the conserved aspartate at aa52. As shown in Fig. 1.5A, the auto-phosphorylation of VirA, and the following phosphoryl-transfer to VirG, activates virulence gene expression (Jin, Roitsch et al. 1990; Jin, Prusti et al. 1990).

VirA is remarkable in its ability to integrate multiple environmental stimuli, and the modularity of the regulatory domains has allowed their functions to be assigned (Chang and Winans 1992). VirA is an integral membrane homodimer with two transmembrane regions (TM1: aa18-39; TM2: aa260-278). The periplasmic domain (P) responds to sugar and acidic pH by coordinating with ChvE. The region between the membrane and the conserved kinase core has been labeled the linker domain (L), which, as seen in Fig. 1.5B, has been assigned as responsible for phenol sensing (Chang and Winans 1992; Gao and Lynn 2005). The conserved kinase domain (K) starts at around aa450 according to the BLAST search, and the CA domain ends at around aa691. An

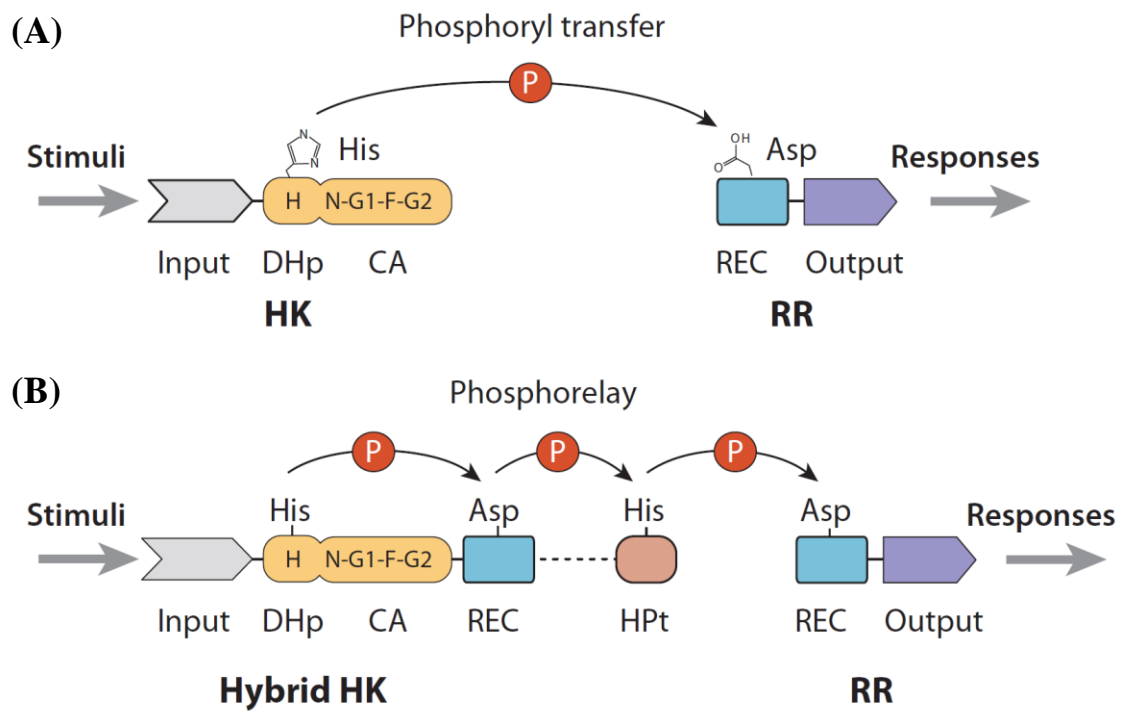
extra receiver domain (R) (aa712-829) is located at the C-terminus of VirA. This receiver domain does not seem to participate in phospho-relay (Fig. 1.5B), and when both VirA and VirG are expressed constitutively, the receiver domain appears to inhibit kinase activity (Chang, Zhu et al. 1996; Brencic, Xia et al. 2004). Other regulatory roles played by the VirA receiver domain will be introduced in Chapter 4.

While many TCS proteins have been identified via sequence alignments, the lack of knowledge of the exact signal has severely hampered studies of signal perception and transmission. The VirA/VirG TCS of *Agrobacterium tumefaciens* has served as a model for the study of TCS regulation, not only because the signal inputs are characterized, but also because regulation within the multi-domain VirA requires global domain coordination for signal integration.

### **Phenol perception by the linker domain and the linker structure model**

Since the most important regulation of VirA's activity is from the perception of phenols, phenol sensing in the linker domain becomes critical. Several questions related to phenol perception include: what kind of physical interaction is involved between the phenol and the linker domain? What kind of biophysical property does the linker domain have to constitute such broad phenol specificity? Is the proton-transfer model reasonable and accurate? Does phenol perception by the linker domain define host range? How are the multiple inputs integrated? All those questions require basic structural information of the linker domain, but as with most signaling domains, the linker domain was not found to be homologous to any identified protein. However, the secondary structure of the linker domain, predicted by the SAM-T02 method (Karplus, Karchin et al. 2003),

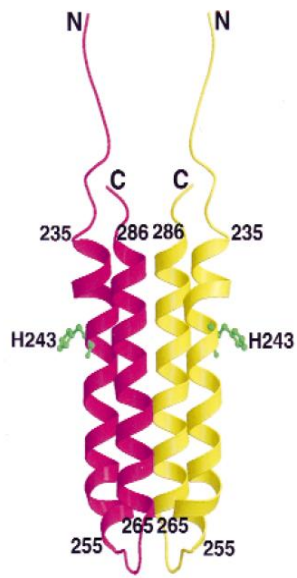
**Fig. 1.3 The two-component signaling scheme.** **A)** The HK/RR two-component signaling with one phosphoryl-transfer reaction (Gao and Stock 2009). The conserved histidine in the H-box is phosphorylated upon signal sensing at the input domain. After transferring the phosphate to the conserved aspartate of the receiver (REC) domain of RR, the output response turns on. The conserved motifs in the CA domain (N-G1-F-G2) of HK are as indicated. **B)** The two-component signaling pathway involving multiple phosphoryl-transfer reactions. A hybrid HK with a covalently linked REC domain and an intermediate HPt (histidine phosphotransfer) domain mediate the multiple phosphoryl-transfer reactions for the final signaling output.



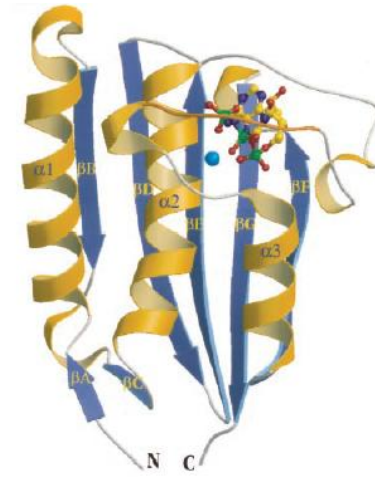


**Fig. 1.4 Available structures of the two-component proteins.** **A)** The DHp domain of *E. coli* EnvZ histidine kinase (Tomomori, Tanaka et al. 1999). The DHp domain forms a dimer with the four-helix bundle structure with one subunit shown in magenta and the other in yellow. The conserved His243 is shown in ball-and-stick. **B)** The typical  $\alpha/\beta$  structure of the CA domain of *E. coli* PhoQ histidine kinase (Marina, Mott et al. 2001). The ATP analog, adenosine 5'-( $\beta,\gamma$ -imino)triphosphate, is shown in ball-and-stick, and the coordinated  $Mg^{2+}$  is in blue sphere. **C)** The entire HK domain of *Thermotoga maritima* TM0853 (Marina, Waldburger et al. 2005). The central four-helix bundle DHp domain is flanking by two CA domains. One subunit is shown in green and the other one in yellow. The conserved His260 and ADP $\beta$ N molecule are shown in ball-and-stick. **D)** The heme-bound PAS domain of *Bradyrhizobium japonicum* FixL (Gong, Hao et al. 1998). The structure displays a palm-like folding with the ligand-free heme group shown in ball-and-stick. **E)** The receiver domain of *E. coli* CheY (Stock, Martinez-Hackert et al. 1993). The highly conserved residues are shown in ball-and-stick. The conserved Asp57 receives the transferred phosphate, assisted by a coordinated  $Mg^{2+}$  (yellow sphere).

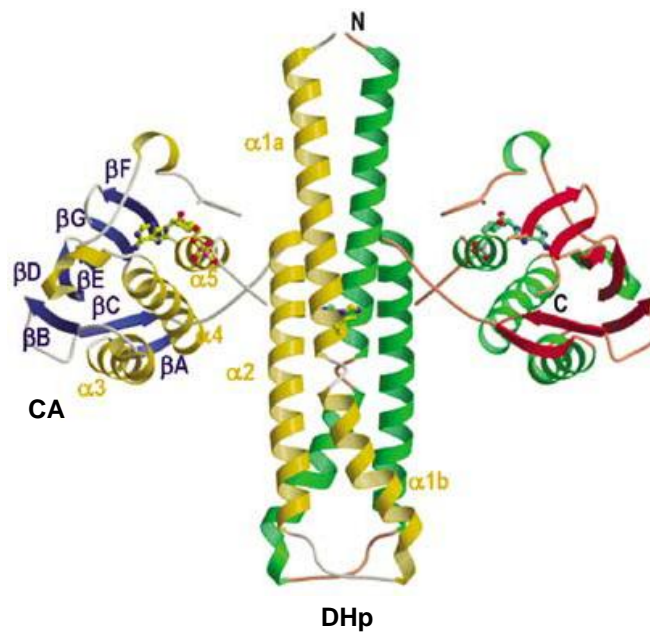
(A)

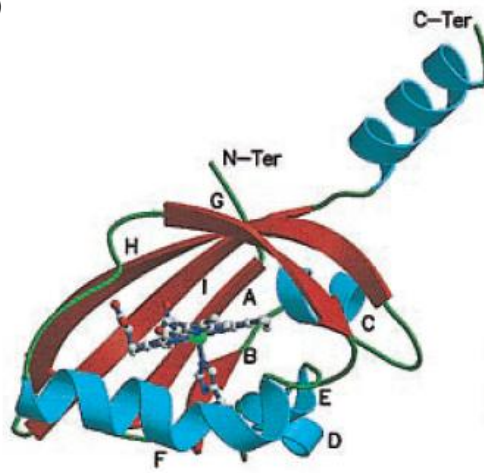
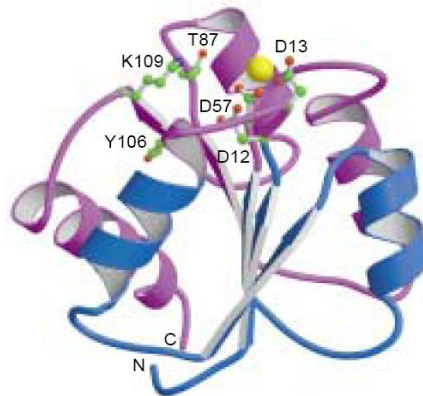


(B)

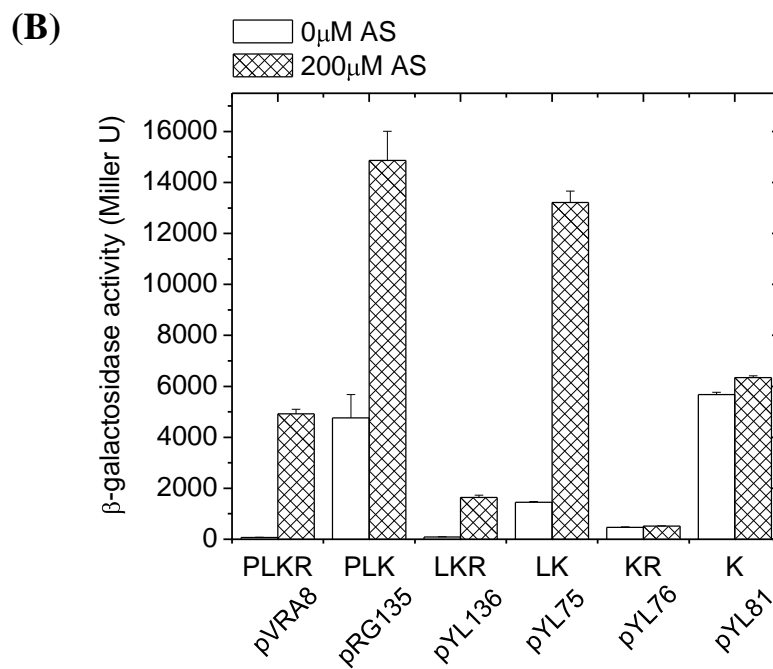
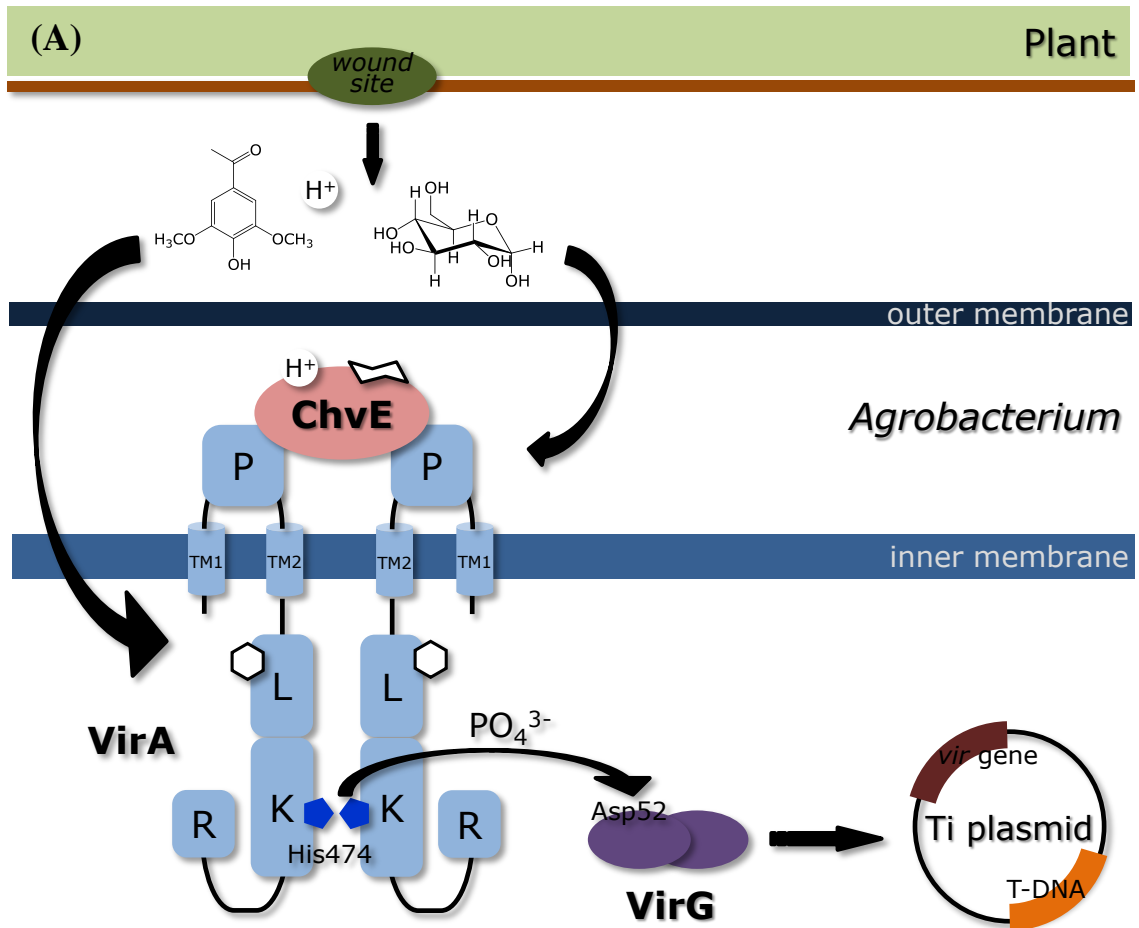


(C)



**(D)****(E)**

**Fig. 1.5 Schematic diagram of VirA/VirG two-component signal transduction and domain organization.** **A)** The signaling is initiated from the molecules released at the plant wound site: phenols, H<sup>+</sup>, and sugar, are detected by *Agrobacterium* VirA. H<sup>+</sup> and sugar are perceived at the periplasmic (P) domain of VirA, mediated by the sugar-binding protein ChvE. The phenols, penetrating through the inner membrane, are perceived at the linker (L) domain. Upon signal sensing, the kinase (K) auto-phosphorylates the conserved His474 (pentagon), and subsequently phosphoryl-transfer to the conserved Asp52 of VirG and activates *vir* gene expression. An additional receiver domain (R) locates at the C-terminus of VirA. **B)** *vir* gene expression of different truncated VirAs. *Agrobacterium* strain A136 carrying pRG109, containing constitutively expressed VirG, and different truncated VirAs were analyzed for *vir* gene expression, represented by  $\beta$ -galactosidase activity, with or without 200  $\mu$ M AS and supplemented with 14 mM glucose.



identified homology to another small-molecule-binding family known as the GAF domains (identified in cGMP-regulated cyclic nucleotide phosphodiesterases, Adenylyl cyclases, and bacterial transcription factor FhlA) (Fig 1.6A)(Gao and Lynn 2007). The GAF domains are observed in all three kingdoms of life and generally appear to regulate various cellular functions via small-molecule binding (Ho, Burden et al. 2000; Martinez, Beavo et al. 2002). This domain also has a ~9% occurrence in HKs (Gao and Stock 2009). More detailed description of the GAF domains, their ligand binding characteristics, and their regulatory roles will be introduced in Chapter 3.

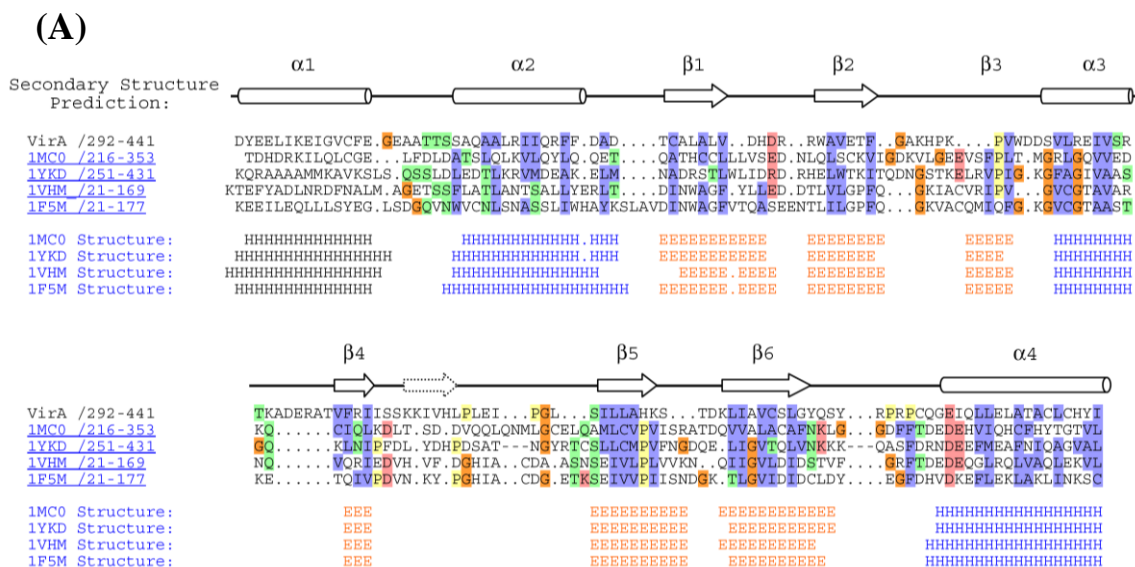
The similarity of the linker domain with the GAF domains allowed the development of a structure model for the linker. YKG9 (Ho, Burden et al. 2000) was selected to serve as the template for threading the linker sequence, VirA(292-441), using Swiss-Model Workspace (Arnold, Bordoli et al. 2006; Kiefer, Arnold et al. 2009), and the structure model is shown in Fig. 1.6B. The central core of the structure contains a  $\beta$ -sheet, arranged in a 2-1-5-4-3 order. On one side of the sheet are the  $\alpha$ 2- $\alpha$ 1- $\alpha$ 4 helices, all of which are close to the dimerization interface. On the other face of the sheet is a structurally less defined region that normally constitutes the ligand binding site in GAF domains (Ho, Burden et al. 2000; Levnikov, Blagova et al. 2006). Structurally validating the model of the GAF-like linker domain, and probes of phenol binding by computational and experimental methods will be described in Chapter 3.

The following chapters aim to elucidate the general signal perception and transmission mechanisms of VirA, explore how phenol perception is coupled and integrated with the periplasmic signals, and better define how this complicated receptor

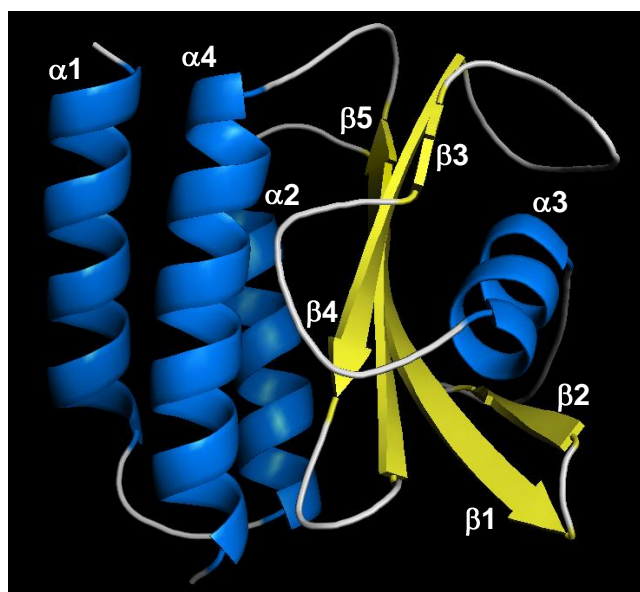
might dictate host range. I will focus primarily on the biophysical characterization of the linker domain and the possible phenol binding site, and use that information to investigate signal transmission across VirA. Through these approaches, I hope to illustrate how nature finds its way, taking *Agrobacterium* pathogenesis as the example, by utilizing the simple but robust TCS machine.

**Fig. 1.6 The linker domain is homologous to the GAF domain.** **A)** Sequence alignment of VirA linker domain with known GAF structures (Gao and Lynn 2007). The alignment was modified according to the secondary structures. The secondary structure of VirA linker was predicted by SAM-T02 method, and shown in cylinder as  $\alpha$ -helices, and arrow as  $\beta$ -sheets. The observed structures of GAF domains are represented by H as  $\alpha$ -helices and E as  $\beta$ -sheets. The selected GAF structures are: 1MC0, 1YKD, 1VHM, and 1F5M. Residues with distant homology are color coded as blue for hydrophobic residues, red for charged residues, orange for Gly, yellow for Pro, and green for Ser, Thr, Asn, and Gln. **B)** Homology structure of the VirA linker domain was developed by threading the sequence of VirA(292-441) into the GAF structure YKG9 (PDB ID: 1F5M) using Swiss-Model workspace. The structure model is colored in blue for  $\alpha$ -helices and yellow for  $\beta$ -sheets.





**(B)**



## CHAPTER 2

### MEDIATING THE ROTATIONAL MOTION BETWEEN THE PHENOL SENSING LINKER DOMAIN AND THE KINASE IN VIRA

#### Introduction

As discussed in Chapter 1, the autokinase domain of VirA, a typical type I HK, contains an N-terminal dimerization domain and a C-terminal catalytic ATP-binding domain. The dimerization domain contains the conserved His474 required for the kinase activity (Pan, Charles et al. 1993). In HKs, a coiled-coil motif is commonly observed in front of the conserved H-box, and its role in facilitating signal transmission has been discussed (Singh, Berger et al. 1998; Tao, Malone et al. 2002). Using the coiled-coil prediction program COILS (Lupas, Dyke et al. 1991), VirA is predicted to form a coiled-coil structure between aa440-462 (Fig. 2.1), a region previously included in the linker domain (Gao and Lynn 2007).

The ATP-binding domain of VirA (aa531-711) phosphorylates the conserved His474. Current evidence suggests that the VirA autokinase adopts a *trans*-phosphorylation mechanism and phosphorylates His474 via the other subunit within the dimer, similar to the phosphorylation mechanism of *E. coli* EnvZ (Cai and Inouye 2003; Brencic, Xia et al. 2004). From the BLAST search, the VirA autokinase is most homologous to HK TM0853 of *Thermotoga maritima* (Fig. 2.2), which adopts a different *cis*-phosphorylation mechanism (Casino, Rubio et al. 2009). TM0853 provided the first entire HK structure containing both the DHP and the CA domain (Fig. 1.4C)(Tanaka,

Saha et al. 1998; Tomomori, Tanaka et al. 1999; Marina, Waldburger et al. 2005). Several possible mechanisms of signal transmission for phosphorylation were proposed, including a possible helical rotation in the DHp domain for a proximal CA domain location for the phosphorylation (Utsumi, Yamada et al. 2008). The homology of VirA and TM0853 now allows for experimental design of the two phosphorylation mechanisms.

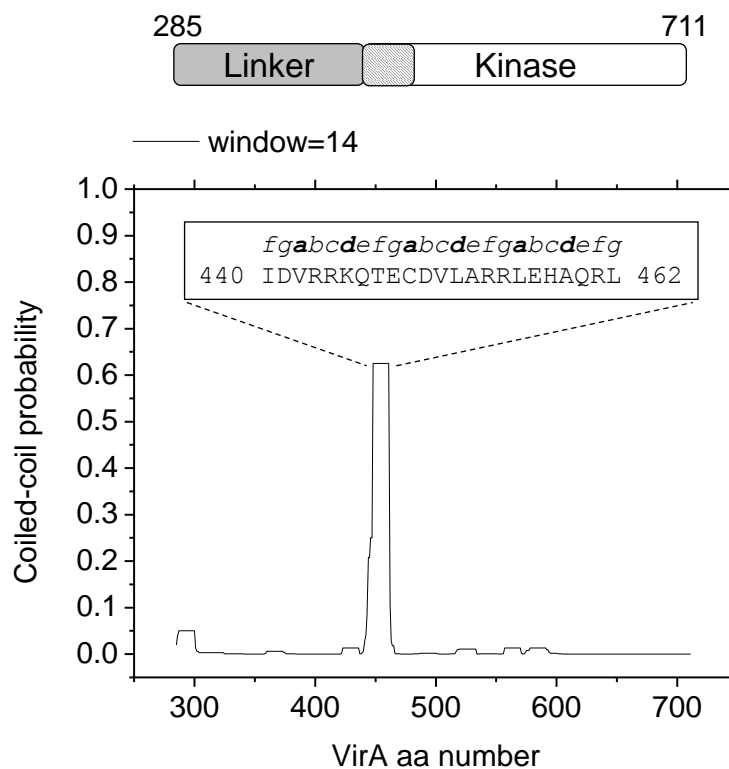
### **The HAMP domain**

The HAMP domain, derived from *H*istidine kinases, *A*denyl cyclases, *M*ethyl-accepting proteins, and *P*hosphatases (Aravind and Ponting 1999), are generally 50-residue motifs containing two amphipathic heptad repeated helices, linked by an unstructured loop. The canonical parallel, coiled-coil architecture (Fig. 2.3)(Hulko, Berndt et al. 2006) has been shown to be critical in mediating transmembrane signaling (Falke and Hazelbauer 2001; Zhu and Inouye 2003; Airola, Watts et al. 2010). Several successful swaps of HAMP domains from different HKs have been described, including the generation of an aspartate-sensing EnvZ by fusing Tar, the chemotaxis sensor (Utsumi, Brissette et al. 1989; Zhu and Inouye 2003); another example is the chimeras generated by fusing the HAMP domain of CpxA sensor kinase to the HK of NarX (Appleman, Chen et al. 2003). These data suggest the HAMP domains share common signal transmission mechanisms.

The VirA histidine autokinase, a membrane-embedded sensor kinase, also has a predicted HAMP-like region in the linker domain (Wang, Gao et al. 2002; Gao and Lynn 2007). Unlike a typical HAMP domain, which only has an unstructured loop connecting

the two helices, the two amphipathic helices (aa292-323, aa425-) of VirA are separated by a 100-residue insert (aa324-424) in a position that is likely to serve as the phenol binding region. This kind of separated HAMP architecture was also observed in the yeast osmosensor Sln1 (Aravind and Ponting 1999), and several molecular motions of HAMP helices have been proposed for mediating kinase activity. One possible motion is the sliding of the  $\alpha$ -helical structure perpendicular to the membrane surface. This “piston-like” motion is usually linked to the periplasmic signal sensing domains, as seen in the chemosensor Tar and Trg (Falke and Hazelbauer 2001; Peach, Hazelbauer et al. 2002), the nitrate binding protein NarX (Cheung and Hendrickson 2009), and the trimethylamine-N-oxide (TMAO) sensor TorS (Moore and Hendrickson 2009). Another possible mechanism would be rotation of the two helices along the helical axis (Hulko, Berndt et al. 2006). This rotational motion has been examined in VirA HAMP by fusing each amphipathic helix to the thermodynamically stable, coiled-coil motif GCN4, and both fusions displayed ratcheted activity by different adapters (Wang, Gao et al. 2002; Gao and Lynn 2007). Additional motions that can accompany the piston-like or rotational motion to propagate signal transmission include the scissor-like motion or changing the helical register, as observed in the concatenated HAMP domain of *Pseudomonas aeruginosa* (Airola, Watts et al. 2010).

**Fig. 2.1 Coiled-coil prediction of VirA-LK(aa285-711).** The coiled-coil formation tendency of VirA aa(285-711) was predicted by COILS (Lupas, Dyke et al. 1991). Amino acid sequence of aa440-462 with high coiled-coil forming probability is shown with the according heptad repeats inside the figure.

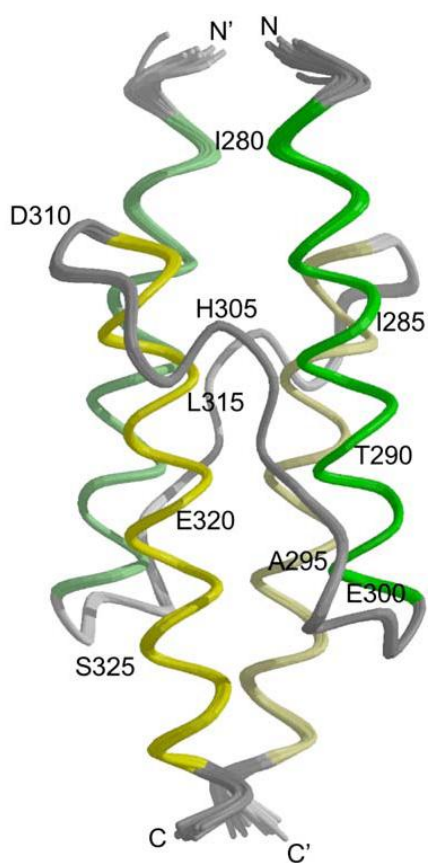


**Fig. 2.2 Sequence alignment of the VirA kinase domain (aa426-711) and TM0853.**

Sequence alignment was extracted from an entire alignment of *Agrobacterium tumefaciens* VirA(426-711), *Thermotoga maritima* TM0853 (PDB:2c2a), *Thermotoga maritima* ThkA(517-755), *E. coli* NtrB (PDB:1R62), and *E. coli* EnvZ (PDB:1BXD). Secondary structure of the VirA kinase domain was predicted by PROF (Ouali and King 2000). Predicted  $\alpha$ -helices are shown in cylinder and  $\beta$ -sheets in arrow. The structure of TM0853 was according to its crystal structure 2c2a, and  $\alpha$ -helices are denoted by H and  $\beta$ -sheets by E. The conserved motifs in the catalytic ATP-binding domain are shown in boxes.



**Fig. 2.3 HAMP domain structure.** The coiled-coil structure of *Archaeoglobus fulgidus* HAMP domain Af1503 is shown with the homodimer structure of the two amphipathic helices in green and yellow, and the unstructured connection region in gray (Hulko, Berndt et al. 2006).





Successful engineering the rotational motion in the HAMP region of the linker domain by GCN4 led to the hypothesis that the rotational motion is coherently mediated through the extended helix from HAMP to the coiled-coil and the DHP domain of the autokinase. This rotational motion is proposed to be regulated by the phenol binding at the linker domain and transmitted through the predicted  $\alpha 4$ . To explore the proposed rotational motion in the extended helix and the potential helical interface for signal transmission, I will fuse the GCN4 motif at different regions along the helix connecting the linker and the kinase domain to define the “ON” and “OFF” conformation. The helix registry of VirA and its homology with TM0853 also allows for examination of the molecular details of the *trans*-phosphorylation mechanism by direct amino acid insertion. At the end, a comparison of VirA’s *trans*-phosphorylation mechanism vs. TM0853’s *cis*-phosphorylation mechanism will be discussed.

### Materials and Methods

**Bacterial strains, plasmids, and reagents.** The bacterial strains and plasmids used in this chapter are listed in Table 2.1. *E. coli* strain XL1-Blue (Stratagene) was used for routine plasmid construction. The AS compound used for *vir* gene induction was purchased from Sigma-Aldrich Corp. IPTG used to induce protein expression and X-gal used in library screening was purchased from Research Products International Corp (rpi). The cloning reagents were purchased from New England Biolab and Promega.

**Table 2.1 Bacterial strains and plasmids used in Chapter 2**

Strains/plasmids	Relevant characteristics	Reference
<i>E. coli</i> strains		
XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac[F' proAB lacI<sup>q</sup>Z M15 Tn10 (Tc<sup>r</sup>)]</i>	Stratagene
<i>A. tumefaciens</i> strains		
A136	Strain C58 cured of pTi plasmid	(Watson, Currier et al. 1975)
A348-3	A136 containing pTiA6NC, $\Delta P_{virA}$ - <i>virA</i> , <i>virA</i> deletion, Km <sup>r</sup>	(Lee, Dudley et al. 1992)
Plasmids		
pYW15b	Broad-host-range expression vector, IncW, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pYW33	$P_{N25}$ -6xHis-LZ- <i>virA</i> ( <i>aa285-471</i> ) in pYW15b, Ap <sup>r</sup>	(Wang, Gao et al. 2002)
pYW39	$P_{N25}$ -6xHis- <i>virA</i> ( <i>aa285-829</i> )( <i>G665D</i> ) in pYW15a, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pYW48	$P_{virA}$ - <i>virA</i> ( <i>aa1-829</i> ) in pYW15b, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pSW209 $\Omega$	<i>virB::lacZ</i> , IncP, Spec <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pJZ4	$P_{virB}$ - <i>lacZ</i> in pMON596, IncP, Spec <sup>r</sup>	(Zhang, Boone et al. 2000)
pJZ6	IncW/ColE expression vector with $P_{N25}$ , Ap <sup>r</sup>	(Zhang, unpublished)
pRG109	$P_{N25}$ -His <sub>6</sub> - <i>virG</i> in pJZ4, Spec <sup>r</sup>	(Gao and Lynn 2005)

pRG150	<i>lacI<sup>f</sup></i> in pJZ4, Spec <sup>r</sup>	(Gao and Lynn 2007)
pRG178	<i>P<sub>N25</sub>-His<sub>6</sub>-LZ(4)-virA(aa426-711)(G665D)</i> in pYW15b, Ap <sup>r</sup>	(Gao and Lynn 2007)
pRG179	<i>P<sub>N25</sub>-His<sub>6</sub>-LZ(3)-virA(aa426-711)(G665D)</i> in pYW15b, Ap <sup>r</sup>	(Gao and Lynn 2007)
pRG180	<i>P<sub>N25</sub>-His<sub>6</sub>-LZ(0)-virA(aa426-711)(G665D)</i> in pYW15b, Ap <sup>r</sup>	(Gao and Lynn 2007)
pYL28	<i>P<sub>N25</sub>-His<sub>6</sub>-virA(aa285-829)(C435F)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL64	<i>P<sub>N25</sub>-virA(aa438-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL75	<i>P<sub>N25</sub>-virA(aa285-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL76	<i>P<sub>N25</sub>-virA(aa438-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL81	<i>P<sub>N25</sub>-virA(aa446-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL99	<i>P<sub>N25</sub>-virA(aa426-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL100	<i>P<sub>N25</sub>-virA(aa460-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL101	<i>P<sub>N25</sub>-virA(aa474-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL102	<i>P<sub>N25</sub>-virA(aa453-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL103	<i>P<sub>N25</sub>-virA(aa467-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL104	<i>P<sub>N25</sub>-virA(aa426-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL105	<i>P<sub>N25</sub>-LZ(0)-virA(aa438-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL106	<i>P<sub>N25</sub>-LZ(3)-virA(aa438-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL108	<i>P<sub>N25</sub>-virA(aa426-711)(C435F)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL111	<i>P<sub>N25</sub>-LZ(4)-virA(aa438-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL119	<i>P<sub>N25</sub>-LZ(1)-virA(aa438-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL120	<i>P<sub>N25</sub>-LZ(2)-virA(aa438-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL136	<i>P<sub>N25</sub>-virA(aa285-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL138	<i>P<sub>N25</sub>-virA(aa285-829)(Q427F)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL139	<i>P<sub>N25</sub>-virA(aa285-829)(Q427W)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL140	<i>P<sub>N25</sub>-virA(aa285-829)(C435K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL141	<i>P<sub>N25</sub>-virA(aa285-829)(E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL147	<i>P<sub>N25</sub>-virA(aa426-711)(Q427F)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL148	<i>P<sub>N25</sub>-virA(aa426-711)(Q427W)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL149	<i>P<sub>N25</sub>-virA(aa426-711)(E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL150	<i>P<sub>N25</sub>-virA(aa426-711)(C435K)</i> in pJZ6, Ap <sup>r</sup>	This chapter

pYL168	<i>P<sub>N25</sub>-virA(aa285-829)(E295Q/E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL169	<i>P<sub>N25</sub>-virA(aa285-829)(E299Q/E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL176	<i>P<sub>N25</sub>-virA(aa285-829)(E295Q)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL177	<i>P<sub>N25</sub>-virA(aa285-829)(E299Q)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL200	<i>P<sub>N25</sub>-LZ(4)-virA(aa450-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL201	<i>P<sub>N25</sub>-LZ(3)-virA(aa450-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL202	<i>P<sub>N25</sub>-LZ(0)-virA(aa450-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL203	<i>P<sub>N25</sub>-virA(aa285-829)(C435R)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL204	<i>P<sub>N25</sub>-virA(aa285-829)(E430Q/C435K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL205	<i>P<sub>N25</sub>-LZ(3)-virA(aa450-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL206	<i>P<sub>N25</sub>-LZ(0)-virA(aa450-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL207	<i>P<sub>N25</sub>-LZ(4)-virA(aa450-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL212	<i>P<sub>N25</sub>-virA(aa450-711)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL213	<i>P<sub>N25</sub>-virA(aa450-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL214	<i>P<sub>N25</sub>-LZ(-2)-virA(aa450-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL215	<i>P<sub>N25</sub>-LZ(-1)-virA(aa450-829)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL227	<i>P<sub>N25</sub>-virA(aa285-829)(E430Q)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL267	<i>P<sub>N25</sub>-LZ(1)-virA(aa426-711)(G665D)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL268	<i>P<sub>N25</sub>-LZ(2)-virA(aa426-711)(G665D)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL269	<i>P<sub>N25</sub>-LZ(-1)-virA(aa426-711)(G665D)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL270	<i>P<sub>N25</sub>-LZ(-2)-virA(aa426-711)(G665D)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL273	<i>P<sub>N25</sub>-virA(aa285-829)(D292N/E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL274	<i>P<sub>N25</sub>-virA(aa285-829)(E294Q/E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL275	<i>P<sub>N25</sub>-virA(aa285-829)(D292N)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL276	<i>P<sub>N25</sub>-virA(aa285-829)(E294Q)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL277	<i>P<sub>N25</sub>-virA(aa285-829)(K298A/E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL282	<i>P<sub>N25</sub>-virA(aa285-711)(I440-A-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL283	<i>P<sub>N25</sub>-virA(aa285-711)(C449-A-D450)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL284	<i>P<sub>N25</sub>-virA(aa285-711)(E457-A-H458)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL286	<i>P<sub>N25</sub>-virA(aa285-829)(K298A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL293	<i>P<sub>N25</sub>-virA(aa285-711)(I440-DA-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL294	<i>P<sub>N25</sub>-virA(aa285-711)(I440-DALK-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL295	<i>P<sub>N25</sub>-virA(aa285-711)(C449-DA-D450)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL296	<i>P<sub>N25</sub>-virA(aa285-711)(C449-DALK-D450)</i> in pJZ6, Ap <sup>r</sup>	This chapter

pYL297	<i>P<sub>N25</sub>-virA(aa285-711)(E457-DA-H458)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL298	<i>P<sub>N25</sub>-virA(aa285-711)(E457-DALK-H458)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL302	<i>P<sub>N25</sub>-virA(aa285-711)(I440-DAL-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL303	<i>P<sub>N25</sub>-virA(aa285-711)(Q427F)(I440-DAL-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL304	<i>P<sub>N25</sub>-virA(aa285-711)(E430K)(I440-DAL-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL305	<i>P<sub>N25</sub>-virA(aa285-711)(C435K)(I440-DAL-D441)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL306	<i>P<sub>N25</sub>-virA(aa285-711)(C449-DAL-D450)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL307	<i>P<sub>N25</sub>-virA(aa285-829)(K298E)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL308	<i>P<sub>N25</sub>-virA(aa285-829)(K298E/E430K)</i> in pJZ6, Ap <sup>r</sup>	This chapter

**Cloning, Transformation, and growth conditions.** *E. coli* strains were grown in Luria-Bertani (LB) medium with appropriate antibiotics at 37 °C, 200 rpm. *A. tumefaciens* strains were grown in LB medium or induction medium containing 1% glycerol at 28 °C, 225 rpm. Plasmid DNAs were purified using Zyppy Plasmid Miniprep kit from Zymo Research. DNA subcloning was carried out following the standard procedures. The PCR reaction was performed by the GeneAmp PCR System 9600 (Perkin Elmer). The PCR products or plasmids were digested with restriction enzymes and separated by 1-3% agarose TBE gels. The desired DNA fragments were purified with Perfectprep Gel Cleanup kit from Eppendorf. Ligation was carried out in a 10 µL volume at room temperature for 1 hour or 16 °C overnight. Transformations into *E. coli* strains were performed by heat-shock following the protocols provided by the competent cells (Stratagene). Transformations into *A. tumefaciens* were done by electroporation using Gene Pulser II (Bio-Rad) with the following parameter setting: 400 ohms resistance (Pulse Controller Unit), 2.5 kV, and 25 µF capacitance. After the transformation, 750 µL

of LB was added to the transformed cells, and the cells were incubated at 28 °C, 225 rpm for 2 hours prior to the plating on LB agar plates with appropriate antibiotics.

**Plasmid constructions.** The plasmids used in this chapter are listed in Table 2.1, and the primers used are listed in Table 2.2. The *Agrobacterium* expression vector used throughout this chapter is pJZ6, which drives the protein expression by the coliphage T5  $P_{N25}$  promoter. All of the *virA* variants or GCN4 fusions were introduced into pJZ6 with *Bam*HI and *Acc65I* sites. The *virA* truncations were generated by PCR from pYW48 with corresponding forward primers and the reverse primers LKKpnI or LKRA1, and introduced into pJZ6 to create pYL64, pYL75 (LK), pYL76, pYL81, pYL99, pYL100, pYL101, pYL102, pYL103, pYL104, pYL136 (LKR), pYL212, and pYL213. The GCN4 leucine zipper fusions LZ(n)-*virA*(450-829) were created by the two-step PCR method, in which the GCN4 part was amplified from pYW33 with YL95 and the corresponding reverse primers, and the *virA* part was amplified from pYW48 with the corresponding complementary forward primers and LKRA1. The amplified DNA was digested with *Bam*HI and *Acc65I*, and introduced into pJZ6 to generate pYL200, pYL201, pYL202, pYL214, and pYL215. pYL201, pYL202, and pYL200 served as the templates for amplification with YL95 and LKKpnI, followed by *Bam*HI and *Acc65I* digestion and ligation to pJZ6, to generate pYL205, pYL206, and pYL207. LZ(n)-*virA*(438-711) chimeras were generated by the same two-step PCR method, resulting in pYL105, pYL106, pYL111, pYL119, and pYL120. LZ(n)-*virA*(426-711)(*G665D*) fusions were also generated by the same two-step PCR method, except using pYW39 as the template for *virA*(*G665D*), and resulted in pYL267, pYL268, pYL269, and pYL270.

The coiled-coil amino acid insertion of LK was also generated by the two-step PCR method with LKR285, LKKpnI, and the corresponding complementary primers. The resulting constructs include the LK(440<sup>nn</sup>) variants pYL282, pYL293, pYL294, and pYL302; the LK(449<sup>nn</sup>) variants pYL283, pYL295, pYL296, and pYL306; and the LK(457<sup>nn</sup>) variants pYL284, pYL297, and pYL298. The LK(440<sup>+3</sup>) variants with the constitutively on mutations were generated by two-step PCR using pYL302 as the template for LK(440<sup>+3</sup>) and the primers LKR285, LKKpnI, and the corresponding complementary primers to introduce the mutations, resulted in pYL303, pYL304, and pYL305.

**Library screening of  $\alpha 4$  and mutagenesis.** Random mutagenesis library was introduced into aa426-437 of LKR(285-829) by the two-step PCR using the primers YL165-YL190 with NNN codon replacing each residue and amplified by LKR285 and LKRA1. The library was digested by *Bam*HI and *Acc*65I, and ligated into the *Bam*HI and *Acc*65I digested pJZ6, and directly transformed into *A. tumefaciens* strain A136 containing pRG109 by eletroporation. To select the constitutively active (ON) variants, the transformants of randomly mutated aa426-437 were screened on the non-inducing induction medium plates with X-gal. The blue colonies, which have the constitutive activities, were extracted and sequenced, and the phenotype was confirmed by site-directed mutagenesis. The LKR mutants were generated by the two-step PCR from pYW48 with LKR285, LKRA1, and the corresponding complementary primers to introduce the mutations. Followed by *Bam*HI and *Acc*65I digestion, the DNA fragment was introduced into pJZ6 to generate pYL138, pYL139, pYL140, pYL141, and pYL203.

pYL227, pYL286, and pYL307 were also generated by the same procedure. The  $\alpha 1$  mutations pYL176, pYL177, pYL275, and pYL276 were generated by the one-step PCR from pYW48 with primers YL238, YL239, YL373, YL374, and LKRA1, followed by *Bam*HI and *Acc*65I digestion and ligation with pJZ6.

426<sup>K</sup>(C435F) and 426<sup>K</sup>(C435K) were generated by PCR using pYL28 and pYL140 as the templates and YL134 and LKKpnI as the primers. After *Bam*HI and *Acc*65I digestion and the successive ligation into pJZ6, pYL108 and pYL150 was created. The other 426<sup>K</sup> variants with constitutive mutations were generated by the one-step PCR from their LKR constructs pYL138, pYL139, and pYL141, with primers YL214, YL215, YL216, and LKKpnI, and resulted in pYL147, pYL148, and pYL149. The  $\alpha 1$ -E430K double mutations were created by PCR using pYL141 as the template and the primers YL238, YL239, YL373, YL374, and LKRA1, followed by *Bam*HI and *Acc*65I digestion and pJZ6 ligation to generate pYL168, pYL169, pYL273, and pYL274. LKR(E430Q/C435K) was generated by the two-step PCR from pYW48 with complementary primers YL256 and YL257 to introduce both mutations, and amplified by LKR285 and LKRA1, resulted in pYL204. LKR(K298A/E430K) and LKR(K298E/E430K) were generated by the two-step PCR using pYL141 as the template and the corresponding complementary primers to introduce K298A and K2983E, respectively, resulted in pYL277 and pYL308. The nucleotide sequence of all the plasmid constructs were confirmed by DNA sequencing facility performed by Agencourt Genomic Service (now Beckman Coulter Genomics).



**Table 2.2 Primers used in Chapter 2**

Primer	Characteristic	Sequence (5'→3') ( <u>RE site</u> ; <i>mutation</i> , <b>insertion</b> )
LKR285	BamHI_285_for	CG <u>GGA TCC</u> GAT TGG TTA GCG CGG CGT
LKRA1	KpnI_stop_829_rev	GC <u>GGT ACC</u> GCA ACT CTA CGT CTT GAT
LKKpnI	KpnI_stop_711_rev	GG <u>GGT ACC</u> CTA ACG CGG TGC CTT ATT GCG
YL95	BamHI_LZ_for	GG <u>GGA TCC</u> GGA GGT TGC GGA GGT AAG
YL50	C435F_for	C ACC GCC <i>TTC</i> CTC TGT CAC
YL51	C435F_rev	GTG ACA GAG <i>GAA</i> GGC GGT G
YL106	BamHI_438_for	CC <u>GGA TCC</u> CAC TAT ATC GAT GTT CGG
YL110	BamHI_446_for	GG <u>GGA TCC</u> CAG ACC GAA TGC GAC GTT TTG G
YL134	BamHI_426_for	GG <u>GGA TCC</u> ATT CAG CTT CTT GAA CTC G
YL135	BamHI_453_for	GG <u>GGA TCC</u> GCC AGA CGA TTG GAG CAT G
YL136	BamHI_460_for	GG <u>GGA TCC</u> CAA CGC CTT GAG GCA GTT G
YL137	BamHI_467_for	GG <u>GGA TCC</u> ACA CTT GCC GGC GGA ATA G
YL138	BamHI_474_for	GG <u>GGA TCC</u> CAT GAA TTT AAT AAC ATT TTG
YL140	LZ(0)-438 <sup>K</sup> _for	G AAG AAG CTG GTC CAC TAT ATC GAT G
YL141	LZ(0)-438 <sup>K</sup> _rev	C ATC GAT ATA GTG GAC CAG CTT CTT C
YL142	LZ(3)-438 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC GCA CTG</b> CAC TAT ATC GAT G
YL143	LZ(3)-438 <sup>K</sup> _rev	C ATC GAT ATA GTG <b>CAG TGC GTC</b> GAC CAG CTT CTT C
YL144	LZ(4)-438 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC GCA CTG AAG</b> CAC TAT ATC GAT G
YL145	LZ(4)-438 <sup>K</sup> _rev	C ATC GAT ATA GTG <b>CTT CAG TGC GTC</b> GAC CAG CTT CTT C
YL146	LZ(1)-438 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC</b> CAC TAT ATC GAT G
YL147	LZ(1)-438 <sup>K</sup> _rev	C ATC GAT ATA GTG <b>GTC</b> GAC CAG CTT CTT C
YL148	LZ(2)-438 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC GCA</b> CAC TAT ATC GAT G
YL149	LZ(2)-438 <sup>K</sup> _rev	C ATC GAT ATA GTG <b>TGC GTC</b> GAC CAG CTT CTT C
YL165	L429X_for	CAG CTT <i>NNN</i> GAA CTC GCC ACC
YL166	L429X_rev	GGT GGC GAG TTC <i>NNN</i> AAG CTG
YL167	E430X_for	CAG CTT CTT <i>NNN</i> CTC GCC ACC
YL168	E430X_rev	GGT GGC GAG <i>NNN</i> AAG AAG CTG
YL169	L431X_for	CTT CTT GAA <i>NNN</i> GCC ACC GCC
YL170	L431X_rev	GGC GGT GGC <i>NNN</i> TTC AAG AAG
YL171	A432X_for	CTT GAA CTC <i>NNN</i> ACC GCC TGC
YL172	A432X_rev	GCA GGC GGT <i>NNN</i> GAG TTC AAG

YL175	A434X_for	CTC GCC ACC <i>NNN</i> TGC CTC TGT C
YL176	A434X_rev	G ACA GAG GCA <i>NNN</i> GGT GGC GAG
YL177	L436X_for	CC ACC GCC TGC <i>NNN</i> TGT CAC
YL178	L436X_rev	GTG ACA <i>NNN</i> GCA GGC GGT GG
YL179	C437X_for	GCC TGC CTC <i>NNN</i> CAC TAT ATC G
YL180	C437X_rev	C GAT ATA GTG <i>NNN</i> GAG GCA GGC
YL181	I426X_for	CGA CCT TGC CAA GGC GAA <i>NNN</i> CAG CTT CTT GAA CTC GCC
YL182	I426X_rev	GGC GAG TTC AAG AAG CTG <i>NNN</i> TTC GCC TTG GCA AGG TCG
YL183	Q427X_for	CCT TGC CAA GGC GAA ATT <i>NNN</i> CTT CTT GAA CTC GCC ACC
YL184	Q427X_rev	GGT GGC GAG TTC AAG AAG <i>NNN</i> AAT TTC GCC TTG GCA AGG
YL185	L428X_for	TGC CAA GGC GAA ATT CAG <i>NNN</i> CTT GAA CTC GCC ACC GCC
YL186	L428X_rev	GGC GGT GGC GAG TTC AAG <i>NNN</i> CTG AAT TTC GCC TTG GCA
YL187	T433X_for	CAG CTT CTT GAA CTC GCC <i>NNN</i> GCC TGC CTC TGT CAC TAT
YL188	T433X_rev	ATA GTG ACA GAG GCA GGC <i>NNN</i> GGC GAG TTC AAG AAG CTG
YL189	C435X_for	CTT GAA CTC GCC ACC GCC <i>NNN</i> CTC TGT CAC TAT ATC GAT
YL190	C435X_rev	ATC GAT ATA GTG ACA GAG <i>NNN</i> GGC GGT GGC GAG TTC AAG
YL191	Q427F_for	TGC CAA GGC GAA ATT <i>TTC</i> CTT CTT GAA CTC GCC
YL192	Q427F_rev	GGC GAG TTC AAG AAG <i>GAA</i> AAT TTC GCC TTG GCA
YL193	Q427W_for	CAA GGC GAA ATT <i>TGG</i> CTT CTT GAA CTC
YL194	Q427W_rev	GAG TTC AAG AAG <i>CCA</i> AAT TTC GCC TTG
YL197	C435K_for	GAA CTC GCC ACC GCC <i>AAA</i> CTC TGT CAC TAT ATC
YL198	C435K_rev	GAT ATA GTG ACA GAG <i>TTT</i> GGC GGT GGC GAG TTC
YL199	E430K_for	CAG CTT CTT <i>AAA</i> CTC GCC ACC
YL200	E430K_rev	GGT GGC GAG <i>TTT</i> AAG AAG CTG
YL214	BamHI_426_Q427F	GG <u>GGA TCC</u> ATT <i>TTC</i> CTT CTT GAA CTC GCC ACC
YL215	BamHI_426_Q427W	GG <u>GGA TCC</u> ATT <i>TGG</i> CTT CTT GAA CTC GCC ACC

YL216	BamHI_426_E430K	GG <u>GGA TCC</u> ATT CAG CTT CTT AAA CTC GCC ACC
YL238	BamHI_E295Q_for	CG <u>GGA TCC</u> GAT TGG TTA GCG CGG CGT TTA GAT TAC GAA CAG CTA
YL239	BamHI_E299Q_for	CG <u>GGA TCC</u> GAT TGG TTA GCG CGG CGT TTA GAT TAC GAA GAG CTA ATC AAA CAG ATC
YL250	LZ(4)-450 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC GCA CTG AAG GAC</b> GTT TTG GCC AGA
YL251	LZ(4)-450 <sup>K</sup> _rev	TCT GGC CAA AAC GTC <b>CTT CAG TGC GTC GAC</b> CAG CTT CTT C
YL252	LZ(3)-450 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC GCA CTG GAC GTT</b> TTG GCC AGA
YL253	LZ(3)-450 <sup>K</sup> _rev	TCT GGC CAA AAC GTC <b>CAG TGC GTC GAC CAG</b> CTT CTT C
YL254	LZ(0)-450 <sup>K</sup> _for	G AAG AAG CTG GTC GAC GTT TTG GCC AGA
YL255	LZ(0)-450 <sup>K</sup> _rev	TCT GGC CAA AAC GTC GAC CAG CTT CTT C
YL256	E430Q/C435K_for	CTT CTT CAA CTC GCC ACC GCC AAA CTC TGT
YL257	E430Q/C435K_rev	ACA GAG <i>TTT</i> GGC GGT GGC GAG <i>TTG</i> AAG AAG
YL258	C435R_for	GCC ACC GCC <i>CGC</i> CTC TGT CAC
YL259	C435R_rev	GTG ACA GAG <i>GCG</i> GGC GGT GGC
YL260	BamHI_450_for	GG <u>GGA TCC</u> GAC GTT TTG GCC AGA CGA
YL261	E430Q_for	ATT CAG CTT CTT CAA CTC GCC ACC GCC
YL262	E430Q_rev	GGC GGT GGC GAG <i>TTG</i> AAG AAG CTG AAT
YL263	LZ(-1)-450 <sup>K</sup> _for	CTG AAG AAG CTG GAC GTT TTG GCC AGA
YL264	LZ(-1)-450 <sup>K</sup> _rev	TCT GGC CAA AAC GTC CAG CTT CTT CAG
YL265	LZ(-2)-450 <sup>K</sup> _for	AGA CTG AAG AAG GAC GTT TTG GCC AGA
YL266	LZ(-2)-450 <sup>K</sup> _rev	TCT GGC CAA AAC GTC CTT CTT CAG TCT
YL365	LZ(1)-426 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC</b> ATT CAG CTT CTT GAA C
YL366	LZ(1)-426 <sup>K</sup> _rev	G TTC AAG AAG CTG AAT <b>GTC</b> GAC CAG CTT CTT C
YL367	LZ(2)-426 <sup>K</sup> _for	G AAG AAG CTG GTC <b>GAC GCA</b> ATT CAG CTT CTT GAA C
YL368	LZ(2)-426 <sup>K</sup> _rev	G TTC AAG AAG CTG AAT <b>TGC GTC</b> GAC CAG CTT CTT C
YL369	LZ(-1)-426 <sup>K</sup> _for	CTG AAG AAG CTG ATT CAG CTT CTT GAA C
YL370	LZ(-1)-426 <sup>K</sup> _rev	G TTC AAG AAG CTG AAT CAG CTT CTT CAG
YL371	LZ(-2)-426 <sup>K</sup> _for	AGA CTG AAG AAG ATT CAG CTT CTT GAA C
YL372	LZ(-2)-426 <sup>K</sup> _rev	G TTC AAG AAG CTG AAT CTT CTT CAG TCT
YL373	BamHI_D292N_for	CG <u>GGA TCC</u> GAT TGG TTA GCG CGG CGT TTA AAT TAC GAA GAG CTA

YL374	BamHI_E294Q_for	CG <b>GGA TCC</b> GAT TGG TTA GCG CGG CGT TTA GAT TAC CAA GAG CTA
YL381	K298A_for	GAG CTA ATC GCA GAG ATC GGA GTA
YL382	K298A_rev	TAC TCC GAT CTC <i>TGC</i> GAT TAG CTC
YL383	virA440 <sup>+1</sup> _for	CTC TGT CAC TAT ATC <b>GCT</b> GAT GTT CGG CGT AAG C
YL384	virA440 <sup>+1</sup> _rev	G CTT ACG CCG AAC ATC <b>AGC</b> GAT ATA GTG ACA GAG
YL385	virA449 <sup>+1</sup> _for	GT AAG CAG ACC GAA TGC <b>GCT</b> GAC GTT TTG GCC AG
YL386	virA449 <sup>+1</sup> _rev	CT GGC CAA AAC GTC <b>AGC</b> GCA TTC GGT CTG CTT AC
YL387	virA457 <sup>+1</sup> _for	G GCC AGA CGA TTG GAG <b>GCT</b> CAT GCG CAA CGC CTT G
YL388	virA457 <sup>+1</sup> _rev	C AAG GCG TTG CGC ATG <b>AGC</b> CTC CAA TCG TCT GGC C
YL399	virA440 <sup>+2</sup> _for	GT CAC TAT ATC <b>GAC GCT</b> GAT GTT CGG CGT AAG
YL400	virA440 <sup>+2</sup> _rev	CTT ACG CCG AAC ATC <b>AGC GTC</b> GAT ATA GTG AC
YL401	virA440 <sup>+4</sup> _for	CTC TGT CAC TAT ATC <b>GAC GCT CTG AAG</b> GAT GTT CGG CGT AAG
YL402	virA440 <sup>+4</sup> _rev	CTT ACG CCG AAC ATC <b>CTT CAG AGC GTC</b> GAT ATA GTG ACA GAG
YL403	virA449 <sup>+2</sup> _for	GT AAG CAG ACC GAA TGC <b>GAC GCT</b> GAC GTT TTG GCC AG
YL404	virA449 <sup>+2</sup> _rev	CT GGC CAA AAC GTC <b>AGC GTC</b> GCA TTC GGT CTG CTT AC
YL405	virA449 <sup>+4</sup> _for	GT AAG CAG ACC GAA TGC <b>GAC GCT CTG AAG</b> GAC GTT TTG GCC AG
YL406	virA449 <sup>+4</sup> _rev	CT GGC CAA AAC GTC <b>CTT CAG AGC GTC</b> GCA TTC GGT CTG CTT AC
YL407	virA457 <sup>+2</sup> _for	G GCC AGA CGA TTG GAG <b>GAC GCT</b> CAT GCG CAA CGC CTT G
YL408	virA457 <sup>+2</sup> _rev	C AAG GCG TTG CGC ATG <b>AGC GTC</b> CTC CAA TCG TCT GGC C
YL409	virA457 <sup>+4</sup> _for	G GCC AGA CGA TTG GAG <b>GAC GCT CTG AAG</b> CAT GCG CAA CGC CTT G
YL410	virA457 <sup>+4</sup> _rev	C AAG GCG TTG CGC ATG <b>CTT CAG AGC GTC</b> CTC CAA TCG TCT GGC C

YL411	K298E_for	GAG CTA ATC GAA GAG ATC GGA GTA
YL412	K298E_rev	TAC TCC GAT CTC TTC GAT TAG CTC
YL415	virA440 <sup>+3</sup> _for	CTC TGT CAC TAT ATC <b>GAC GCT CTG</b> GAT GTT CGG CGT AAG
YL416	virA440 <sup>+3</sup> _rev	CTT ACG CCG AAC ATC <b>CAG AGC GTC</b> GAT ATA GTG ACA GAG
YL417	virA449 <sup>+3</sup> _for	GT AAG CAG ACC GAA TGC <b>GAC GCT CTG</b> GAC GTT TTG GCC AG
YL418	virA449 <sup>+3</sup> _rev	CT GGC CAA AAC GTC <b>CAG AGC GTC</b> GCA TTC GGT CTG CTT AC

---

**$\beta$ -galactosidase assays for *vir* gene induction.** The LZ(n)-426<sup>K(G665D)</sup> variants were transformed into *A. tumefaciens* strain A348-3 containing pRG150, which has *lacI<sup>f</sup>* to induce the chimera expression only during IPTG induction. The *A. tumefaciens* strains were grown in LB medium with appropriate antibiotics at 28 °C to an OD<sub>600</sub> of 0.4-0.6. The cells were pelleted by centrifugation at 4 °C, 7000 x g, for 10 min. The pellet was washed with PBS, and diluted to OD<sub>600</sub> ~ 0.1 into tubes containing a total of 1 mL induction medium with 200  $\mu$ M IPTG, and cultured at 28 °C, 225 rpm for 15 hrs.  $\beta$ -galactosidase activity was determined as previously described (Miller 1972), and the reading of optical densities at 600 nm and 415 nm was performed by EL800 microplate reader (BIO-TEK Instruments).

Except for the LZ(n)-426<sup>K(G665D)</sup> variants, all of the other *virA* variants were transformed into *A. tumefaciens* strain A136 containing pRG109, which carries *P<sub>virB</sub>-lacZ* and *P<sub>N25</sub>-virG*, for *vir* gene expression. The cells were grown and pelleted by the same procedure, and diluted to OD<sub>600</sub> ~ 0.1 into tubes containing a total of 1 mL induction media with or without 300  $\mu$ M AS, as indicated, and cultured at 28 °C, 225 rpm for 15

hrs. The  $\beta$ -galactosidase activity was determined by the same method as described, from the reading of the optical densities at 600 nm and 415 nm.

**Immunoblot analysis.** *A. tumefaciens* strains were grown in 50 mL LB medium with appropriate antibiotics at 28 °C overnight. The cells were harvested by centrifugation at 4 °C, 7000 x g, for 10 min. The pelleted cells were washed with PBS and lysed on ice by sonication, performed by the Ultrasonic Processor (Cole Parmer) with the following parameters: ON pulse: 1 sec; OFF pulse: 2 sec; total pulse: 1 min. The clear lysates were obtained by centrifugation at 4 °C, 9000 x g, for 10 min, and analyzed by 10% SDS-PAGE followed by electro-blotting onto nitrocellulose membrane using the mini trans-blot transfer apparatus (Bio-Rad). The membrane was blocked with 3% BSA in TBS, and probed with anti-VirA polyclonal antibody (Binns et al, unpublished) at 1:200 dilutions. Visualization was achieved using the goat anti-rabbit antibody conjugated with alkaline phosphatase (Amersham) at 1:1000 dilutions, followed by the 1-step NBT/BCIP development (Pierce).

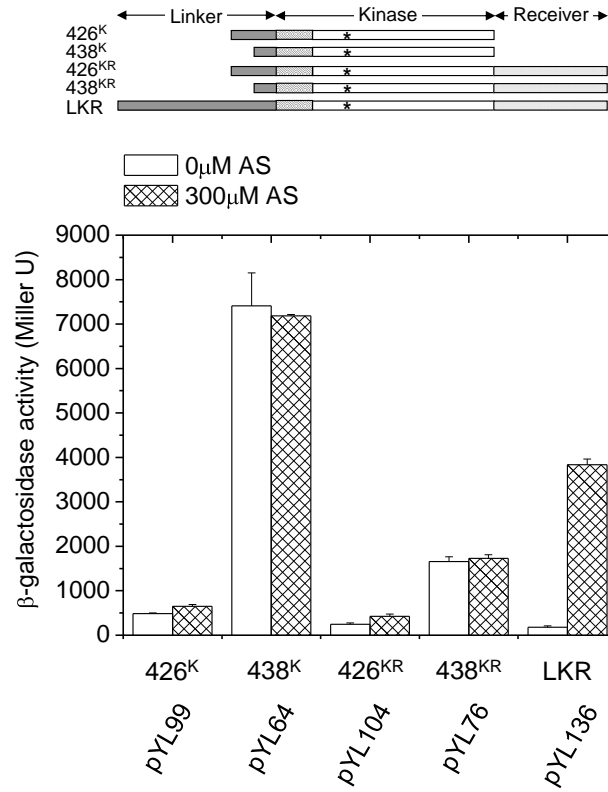
## Results

### GCN4 fusions

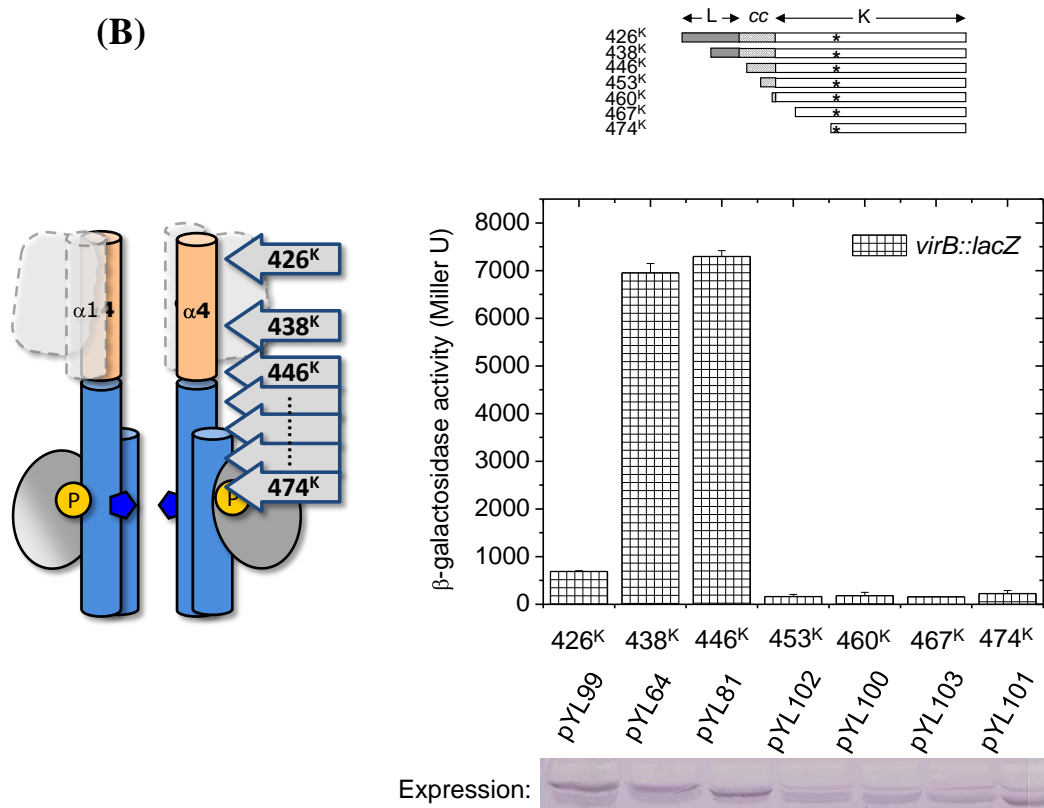
Secondary structure predictions of VirA LK (linker-kinase) suggested that aa440-462 formed the coiled-coil structure (Fig. 2.1). As an initial test of this hypothesis, 438<sup>K</sup> and 438<sup>KR</sup>, which preserve the entire coiled-coil region of kinase, were constructed and analyzed for their  $\beta$ -galactosidase expression. When compared with 426<sup>K</sup> and 426<sup>KR</sup> (Fig. 2.4A)(Gao and Lynn 2005), both 438<sup>K</sup> and 438<sup>KR</sup> significantly elevated activity,

**Fig. 2.4 aa426-437 represses VirA kinase activity.** *A. tumefaciens* strain A136 carrying pRG109 and the indicated plasmids were assayed for *vir* gene expression. **A)**  $\beta$ -galactosidase activity of K and KR starting at aa426 or aa438, and analyzed the activity in the absence (open bar) or presence (hatched bar) of 300  $\mu$ M AS. **B)** Different N-terminal extended kinase were assayed for  $\beta$ -galactosidase activity in the absence of inducers. The design of truncations is illustrated at the left. 426<sup>K</sup> starts from the  $\alpha$ 4 of the linker domain (orange), and 474<sup>K</sup> starts right at the conserved His474. The linker domain except  $\alpha$ 4 is shown in dashed line. Schematic representation of each construct is shown above the activity chart with the conserved His474 denoted in asterisk. Protein expression of each constructs was analyzed by Western blot, probed by anti-VirA antibody.

(A)



(B)





consistent with the HAMP sequence serving a repressive function, and narrowed the repressive role to 11 amino acids (aa426-437). Systematic truncations of the coiled-coil region were generated to further analyze the coiled-coil as well as defining the region of functional kinase (Fig. 2.4B). 446<sup>K</sup>, which represents the conserved kinase region according to the alignment of VirA and TM0853 (Fig. 2.2), had similar level of activity as of 438<sup>K</sup>. Further coiled-coil truncation (453<sup>K</sup>), a complete deletion (460<sup>K</sup>), or removal of a portion of the kinase (467<sup>K</sup> and 474<sup>K</sup>) were all inactive, and the decrease in *in vivo* protein expression (Fig. 2.4B) indicated the inactivation was likely due to destabilization of the proteins.

Previous attempts to probe the ratcheting motion within the HAMP helices of the linker domain suggested that the kinase activity was regulated by changes in the registry of heptad repeats along the coiled-coil (Gao and Lynn 2007). To determine whether the ratcheting motion extended to the kinase, the thermodynamically stable, parallel GCN4 coiled-coil of *Saccharomyces cerevisiae* (Fig. 2.5) was fused at aa450, roughly half-way between the active (446<sup>K</sup>) and inactive (453<sup>K</sup>) truncations (Fig. 2.6A). The 450<sup>K</sup> truncation showed weak activity *in vivo* (Fig. 2.6B), similar to 453<sup>K</sup>, and the coiled-coil heptads prediction placed the conserved His474 in an *e* position (Fig. 2.6A). The in register LZ(0)-450<sup>K</sup> fusion, which also places His474 at *e*, has 20-fold more *in vivo* activity. A three amino acid insertion, LZ(3)-450<sup>K</sup>, would move His474 to the *a* position (Fig. 2.6A), and this chimera shows twice the activity of LZ(0)-450<sup>K</sup> (Fig. 2.6B). A four amino acid insertion, LZ(4)-450<sup>K</sup>, which would position His474 at *b*, shows the same level of kinase activity as LZ(0)-450<sup>K</sup>. The ratcheting effect was even more apparent when the repressive receiver (R) domain was retained (Fig. 2.6C). The *in vivo* protein

expression was enhanced in all GCN4-450<sup>KR</sup> fusions (Fig. 2.6C), indicating the fusion helps stabilize the protein. LZ(3)-450<sup>KR</sup> displayed the highest *in vivo* activity when compared to 450<sup>KR</sup>. LZ(0)-450<sup>KR</sup> and LZ(4)-450<sup>KR</sup> showed a reduced activity even though expression levels were similar. Rotating His474 to the *c* and *d* position at the opposite face of the coiled-coil, as LZ(-1)-450<sup>KR</sup> and LZ(-2)-450<sup>KR</sup>, completely turned off the kinase activity (Fig. 2.6D).

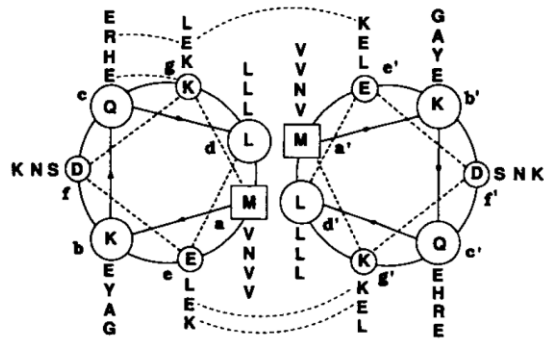
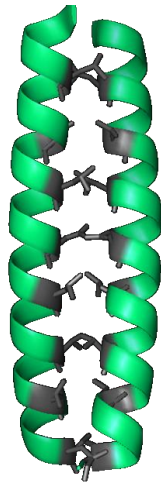
The GCN4 leucine zipper was also fused to 438<sup>K</sup>, which contains the entire coiled-coil and displayed the highest native kinase activity (Fig. 2.7). Similar to GCN4-450<sup>KR</sup> fusions, GCN4-438<sup>K</sup> fusions also increased the *in vivo* protein expression, compared to 438<sup>K</sup>. LZ(2)-438<sup>K</sup>, which would position His474 at the *e* position, slightly enhanced the kinase activity. Moving His474 to the *d* and *f* position, LZ(1)-438<sup>K</sup> and LZ(3)-438<sup>K</sup>, reduced the activity, and the *c* and *g* positions, LZ(0)-438<sup>K</sup> and LZ(4)-438<sup>K</sup>, showed even less activity. The protein expression of all the GCN4-438<sup>K</sup> fusions were the same (Fig. 2.7C), and taken together, these chimeras made the same ratcheting prediction.

The GCN4-450 and GCN4-438 fusions both displayed ratcheted activity, which is consistent with the hypothesis that the rotational motion changes the proximity between the DHp and the CA domain for phosphorylation, and the distance between the two domains is modulated by the degree of rotation of the helix. The rotation is triggered by the HAMP-like helices of the linker domain upon phenol perception and transmitted through the proposed  $\alpha 4$  of the linker (Wang, Gao et al. 2002; Gao and Lynn 2007). The kinase truncation (Fig. 2.4B) indicates that the native functional kinase, which starts at around aa438, is repressed in activity by the  $\alpha 4$  of the linker domain (aa426-437). Therefore, a repressive dimerization interface should exist between  $\alpha 4$ - $\alpha 4'$ , and the same

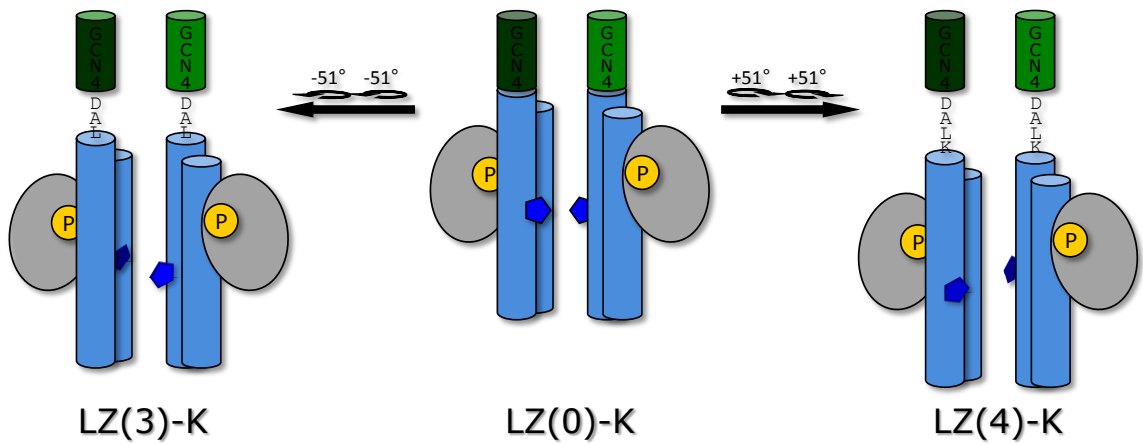
**Fig. 2.5 The yeast transcription factor GCN4.** **A)** GCN4 amino acid sequence and the heptad repeats formation. The parallel, coiled-coil structure of GCN4 is shown by its crystal structure (PDBID: 2ZTA), and the hydrophobic residues in *ad* positions aligned at the dimer interface are shown in ball-and stick. The helical wheel representation of GCN4 shows the position of the amino acids within the coiled-coil (O'Shea, Klemm et al. 1991). **B)** GCN4 fusions mediate His474 rotation in the VirA kinase. The degree of rotation is controlled by different adapters: LZ(4)-fusion creates a +51° rotation, while LZ(3)-fusion creates a -51° (+309°) rotation.

(A)

*abcdefghijklmnopabcdefghijklmnopabcdefghijklmnop*  
**GCN4** : GGCGGKVKQLEDKVEELLSKNYHLENEVARLKKLV

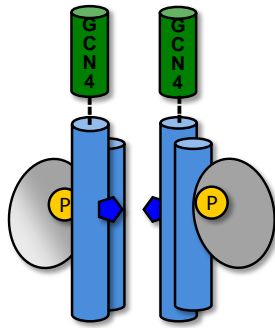


(B)



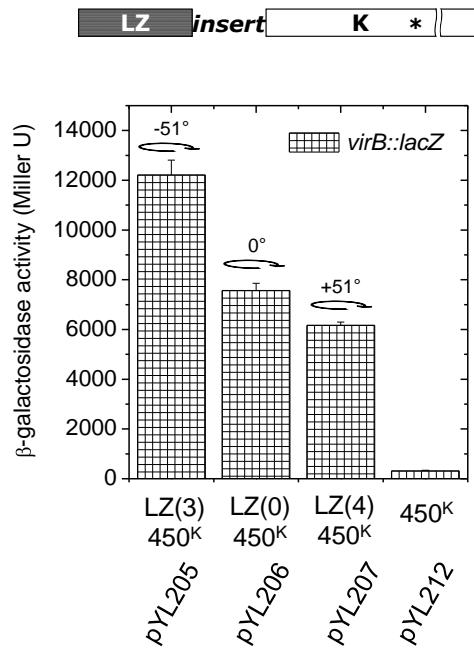
**Fig. 2.6 GCN4-fused 450<sup>K</sup> and 450<sup>KR</sup>.** **A)** Experimental design of the GCN4 fusions. The heptad repeats from *a* to *g* were built off from GCN4 registry and the adapters. GCN4 enforces the hydrophobic *ad* interface, and shifts the heptads of the kinase coiled-coil according to different adapters. The predicted position of His474 in each fusion is shown at the end of the sequence. *A. tumefaciens* strain A136 carrying pRG109 and the indicated **B)** GCN4-450<sup>K</sup> fusions and **C)** GCN4-450<sup>KR</sup> fusions were assayed for *vir* gene expression without inducers. The degree of rotation created by each fusion is shown in the graph. The 0 degree rotation is defined at LZ(0)-450 as it positions His474 at the native *e* position, same as the COILS prediction (Fig. 2.1). Protein expression of the constructs was analyzed by Western blot, probed by anti-VirA antibody. **D)** Helical wheel representation of the coiled-coil dimer. The predicted positions of His474 in each GCN4 fusions are as indicated. The active fusions are labeled as the “ON” conformation, whereas the fusions without activity are the “OFF” conformations.

(A)

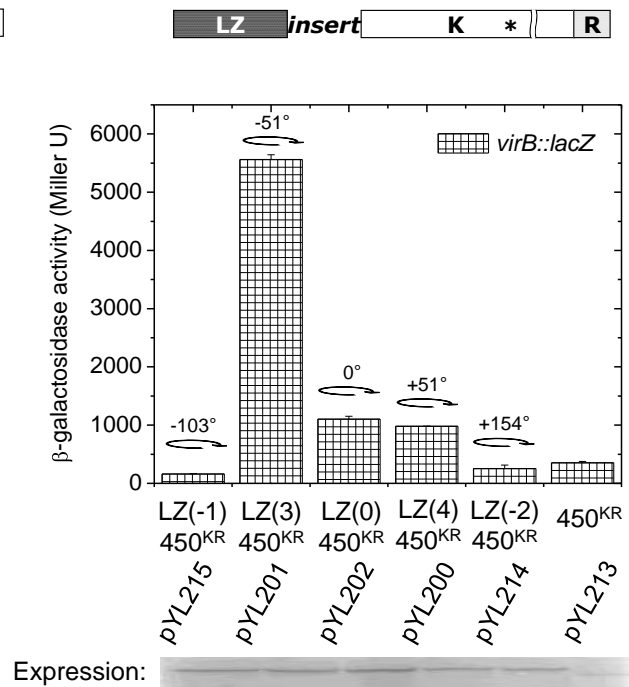


	LZ	insert	K	*	R
LZ (4) -450 <sup>K/KR</sup>		<i>cdefgabc</i>	<i>cdefgabcd</i>	<i>b</i>	
LZ (3) -450 <sup>K/KR</sup>		<i>cdefgabcd</i>	<i>efgabc</i>	<i>a</i>	
LZ (0) -450 <sup>K/KR</sup>		<i>cdefga</i>	<i>bcdefg</i>	<i>e</i>	
LZ (-1) -450 <sup>K/KR</sup>		<i>cdefg</i>	<i>abcdef</i>	<i>d</i>	
LZ (-2) -450 <sup>K/KR</sup>		<i>cdef</i>	<i>gabcde</i>	<i>c</i>	
450 <sup>K/KR</sup>			<i>bcdefg</i>	<i>e</i>	
					<i>DVLARR--H<sub>474</sub>--</i>

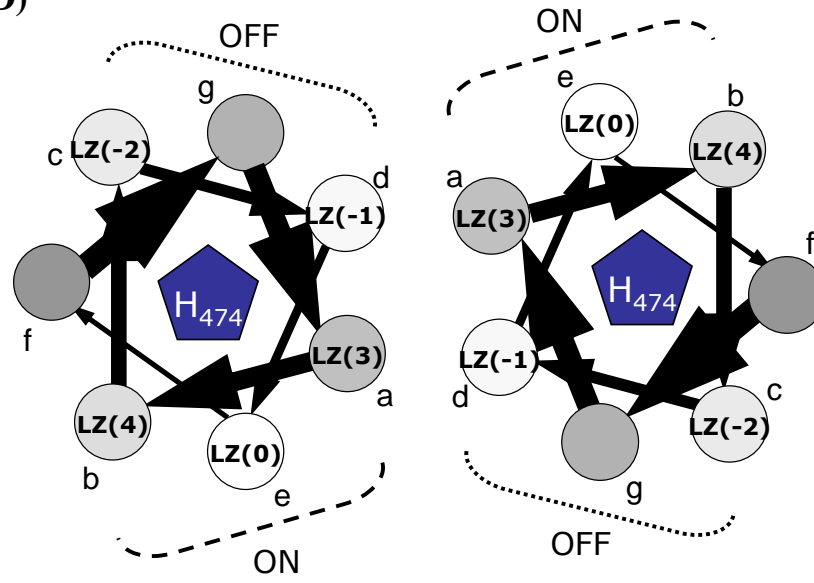
(B)



(C)



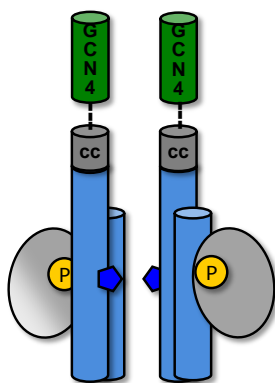
(D)



**Fig. 2.7 GCN4-fused 438<sup>K</sup>.** **A)** Experimental design of the GCN4-438<sup>K</sup> fusions. The predicted position of His474 in each fusion is shown at the end of the sequence. **B)** Helical wheel representation of the coiled-coil dimer of the GCN4-438<sup>K</sup> fusions. The positions of His474 in each GCN4 fusions are as indicated. **C)** *A. tumefaciens* strain A136 carrying pRG109 and indicated GCN4-438<sup>K</sup> fusions were assayed for *vir* gene expression without inducers. The degree of rotation created by each fusion is shown in the graph. The 0 degree rotation is defined at LZ(2)-438<sup>K</sup>, which has the highest *in vivo* activity. Protein expression of each constructs was analyzed by Western blot, probed by anti-VirA antibody.

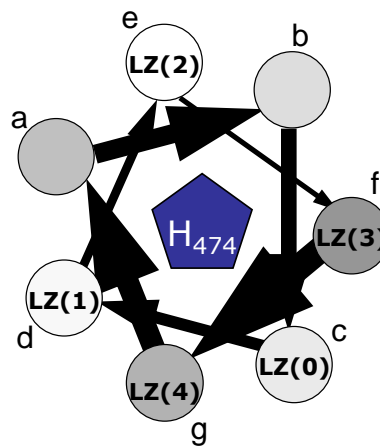
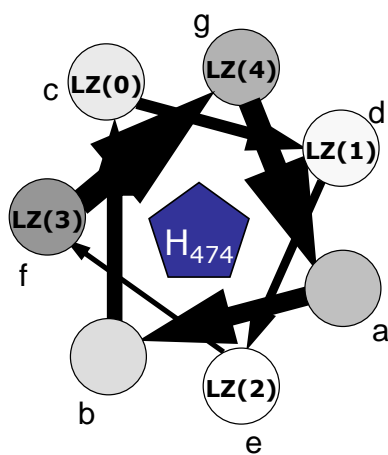


(A)

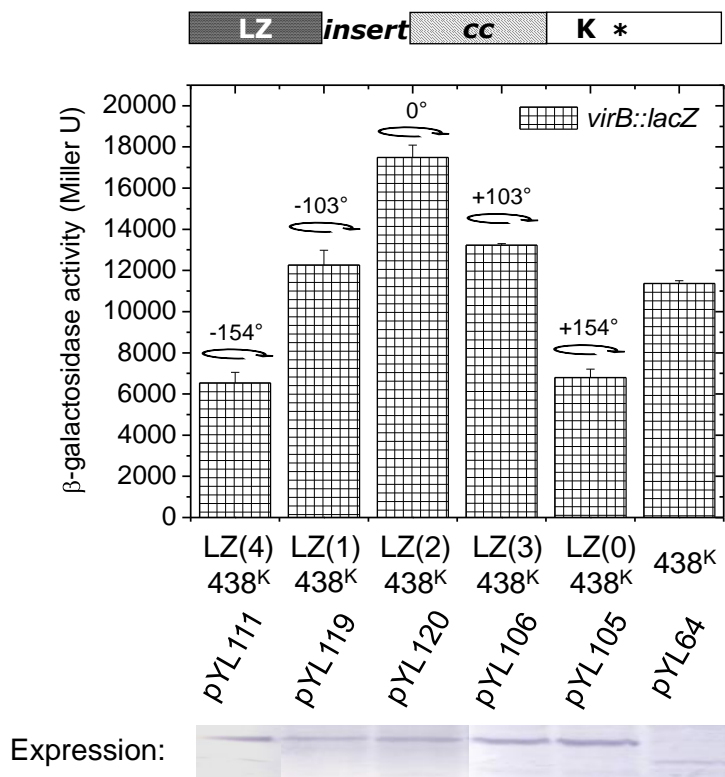
LZ (4) -438<sup>K</sup>LZ (3) -438<sup>K</sup>LZ (2) -438<sup>K</sup>LZ (1) -438<sup>K</sup>LZ (0) -438<sup>K</sup>438<sup>K</sup>

LZ	insert	cc	K	*
	<i>cdefgabc</i>	<i>cdefgabc</i>		<i>g</i>
LZ (4)	--RLKKLVDALKHYIDVRRK---		H <sub>474</sub>	--
	<i>cdefgabcd</i>	<i>efgabcde</i>		<i>f</i>
LZ (3)	--RLKKLVDAL-HYIDVRRK---		H <sub>474</sub>	--
	<i>cdefgabc</i>	<i>defgabcd</i>		<i>e</i>
LZ (2)	--RLKKLVDA--HYIDVRRK---		H <sub>474</sub>	--
	<i>cdefgab</i>	<i>cdefgabc</i>		<i>d</i>
LZ (1)	--RLKKLVD---HYIDVRRK---		H <sub>474</sub>	--
	<i>cdefga</i>	<i>bcdefgab</i>		<i>c</i>
LZ (0)	--RLKKLV----HYIDVRRK---		H <sub>474</sub>	--
		<i>defgabcd</i>		<i>e</i>
438 <sup>K</sup>	-----			H <sub>474</sub>

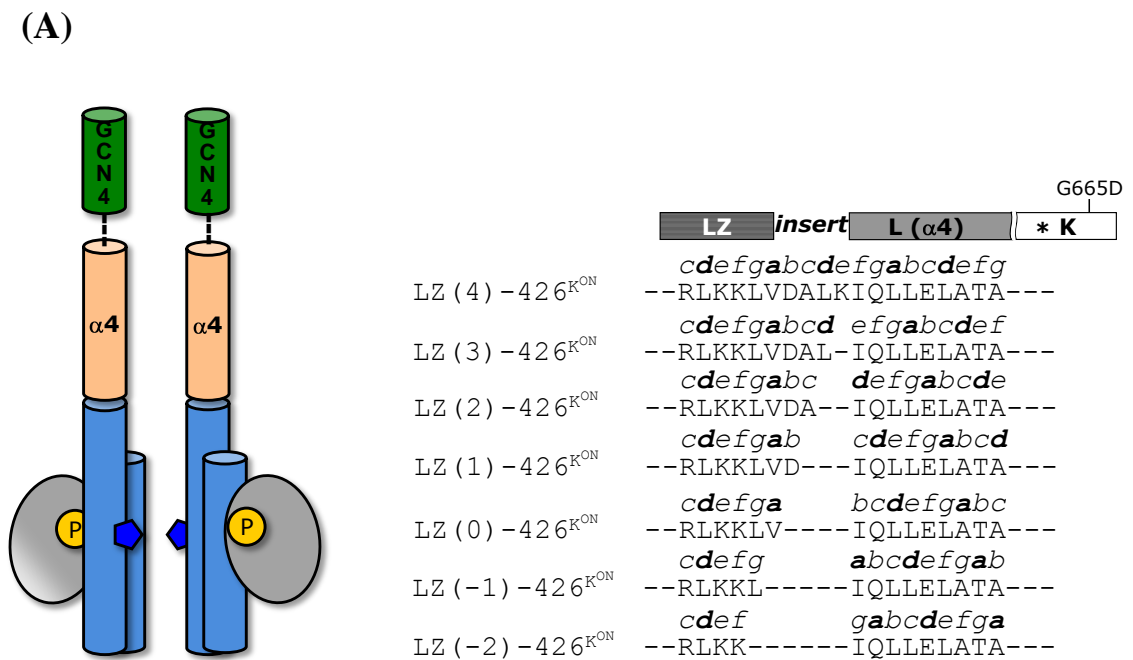
(B)



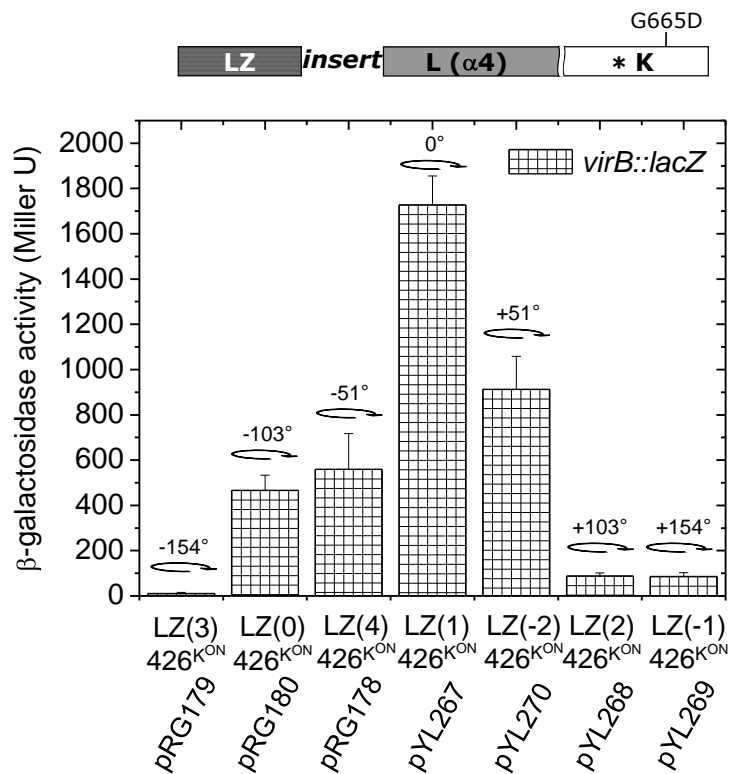
(C)



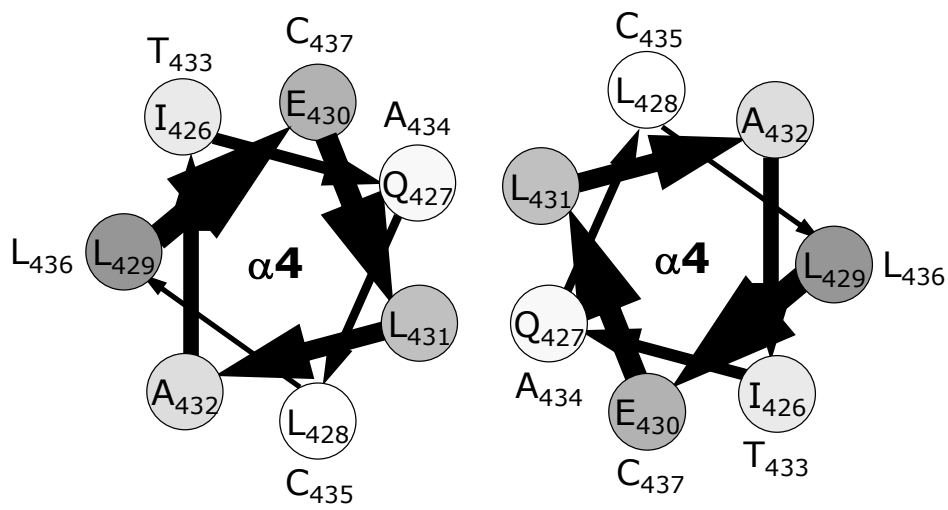
**Fig. 2.8 GCN4-fused 426<sup>K(G665D)</sup>.** **A)** Experimental design and the heptad repeats registry of the GCN4-426<sup>K(G665D)</sup>(426<sup>KON</sup>) fusions. **B)** *A. tumefaciens* strain A348-3 carrying pRG150 and indicated GCN4-426<sup>KON</sup> fusions were assayed for *vir* gene expression without phenolic inducer and supplemented with 200  $\mu$ M IPTG. The degree of rotation created by each fusion is shown in the graph. The 0 degree rotation is defined at LZ(1)- 426<sup>KON</sup>, which has the highest *in vivo* activity. **C)** Helical wheel representation of aa426-437 (linker  $\alpha$ 4) of the LZ(1)-426<sup>KON</sup> fusion.



(B)



(C)



ratcheting motion should occur in  $\alpha4$ - $\alpha4'$  to mediate the ON/OFF conformation, and populates through the helices for kinase activation. To analyze the dimer interface of  $\alpha4$ - $\alpha4'$  and the coherent rotational motion, a complete heptads of GCN4-426<sup>K</sup> fusions were created (Fig. 2.8). Since the wild-type 426<sup>K</sup> has low basal activity, a constitutive mutation G665D was incorporated (426<sup>KON</sup>). As seen in Fig. 2.8B, 426<sup>KON</sup> showed a clear ratcheting activity as observed in LZ-450<sup>K</sup> and LZ-438<sup>K</sup> chimeras. The kinase activity was strictly depending on the degree of rotation, with LZ(1)-426<sup>KON</sup> displayed the highest activity, and successively decreased when rotating away from this conformation. The  $\alpha4$ - $\alpha4'$  dimer interface of LZ(1)-426<sup>KON</sup> is shown in Fig. 2.8C

#### **Library screening of constitutively active mutants in aa426-437 (linker $\alpha4$ )**

The ability of linker  $\alpha4$  to repress the kinase activity and relief of the repression by the rotational motion suggests that the  $\alpha4$ - $\alpha4'$  interface has an intermediate coiled-coil stability. This simple model predicts that stabilizing mutations within linker  $\alpha4$  could be found that give signal independent activity by merely stabilizing the active dimer interface. To test this model, aa426-437 of LKR(285-829) was randomly mutagenized, and the transformants were screened on induction plates without phenolic inducer and search for active variants (Fig. 2.9A). This approach identified six constitutively active mutations with three substitutions at Cys435, two at Gln427, and one at Glu430. Among them, E430K and C435K maintained some phenol inducibility in the presence of 300  $\mu$ M AS, while the other four mutants, Q427F, Q427W, C435R, and C435F, showed a strong constitutive phenotype independent of phenol (Fig. 2.9B).

In order to evaluate whether the constitutive phenotype of the mutations selected from the library was resulted from  $\alpha 4$ - $\alpha 4'$  interaction, these constitutive mutations were moved to 426<sup>K</sup>. The removal of the rest of the linker domain (aa285-425), which presumably constitutes the phenol binding site, leaves only the presence of the  $\alpha 4$  coiled-coil for a direct examination of the mutations for coiled-coil interaction, independent of phenolic inducers. As seen in Fig. 2.10, while wild-type 426<sup>K</sup> has a repressed activity, all the hydrophobic constitutive variants (Q427F, Q427W, C435F) enhanced kinase activity, which may result from stabilizing the  $\alpha 4$  coiled-coil dimer interface at the “ON” conformation, as shown in Fig. 2.8C. One of the charged constitutive 426<sup>K</sup> variant, C435K, also had an enhanced activity. The location of C435 is close to E430 in our model (Fig. 2.8C), suggests that those two residues can form an inter-subunit salt-bridge to stabilize the coiled-coil at the ON conformation. To test this idea directly, the double mutant, LKR(E430Q/C435K) was generated. As seen in Fig. 2.11, the double mutation removed the constitutive phenotype of C435K as well as the phenol inducibility, consistent with disruption of the inter-subunit salt-bridge formation. Taken together, these data suggest the constitutive phenotype of the  $\alpha 4$  mutations arise from either stabilizing the coiled-coil interface through hydrophobic complementarities or through inter-subunit salt-bridge formation.

The low basal activity of E430K in 426<sup>K</sup> (Fig. 2.10) suggests that its constitutive phenotype in LKR is unlikely a result of  $\alpha 4$ - $\alpha 4'$  stabilization. The GAF model of the linker domain and the HAMP domain predict the formation of a four-helix bundle composed of two  $\alpha 4$  and two  $\alpha 1$  helices.  $\alpha 1$  contains several charged residues at the helix surface (Fig. 2.12A), and therefore, the constitutive phenotype of E430K could

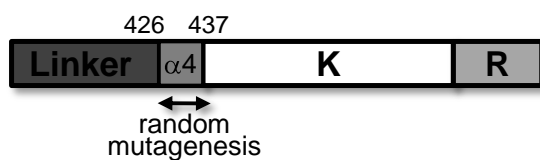
possibly result from the  $\alpha 1$ - $\alpha 4$  charge interaction. Double mutations of E430K and the neutralized charged residues in  $\alpha 1$  were generated to test this idea. As shown in Fig. 2.12B, neutralizing the negatively charged residues in  $\alpha 1$  along with E430K altered the level of the constitutive activities. However, the most significant change was observed in the double mutation of E430K and K298E (Fig. 2.12C). The single substitution of the K298 residue with either a neutral residue or a negatively charged residue abolished kinase activity and phenol inducibility, but when a compensating charge was built between  $\alpha 1$  and  $\alpha 4$  (K298E/E430K), the kinase activity and the phenol inducibility were regenerated. This result enforces the idea of salt-bridge formation between  $\alpha 1$ - $\alpha 4$  and its critical role in mediating signal transmission.

### **Kinase coiled-coil insertion**

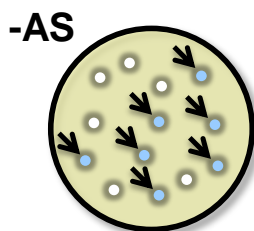
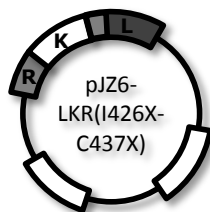
The above experiments further supported the idea that the *in vivo* kinase activity of VirA could be regulated by small motions, either by the GCN4-mediated conformational change or linker  $\alpha 4$  mutations that bias specific helical conformation. This model was further extended to examine VirA's phosphorylation mechanism. The conserved His474 resides in the dimerization domain of the kinase, and the kinase activation requires both successful phosphorylation from the ATP-binding domain and the subsequent phosphoryl-transfer to VirG. Brencic et al. used hetero-dimer formation to show that His474 in VirA kinase is phosphorylated between subunits of the dimer, the so-called *trans*-phosphorylation mechanism (Brencic, Xia et al. 2004). Previous *in vivo* phosphorylation of VirA using  $^{32}\text{P}$  further indicated that His474 accumulates a basal level of phosphate without the phenolic inducer (Mukhopadhyay, Gao et al. 2004).

**Fig. 2.9 Library screening for constitutively active (ON) mutation in linker  $\alpha 4$  (aa426-437).** **A)** Library was generated by randomly mutagenized aa426-437 in LKR(285-829) with NNN-encoding primers for each amino acid. Constitutively active variants were selected by screening the library on non-inducing medium plate for blue colonies (ON phenotype). **B)** *A. tumefaciens* strain A136 carrying pRG109 and the constitutively on mutants were assayed for *vir* gene expression with or without 300  $\mu$ M AS.

(A)

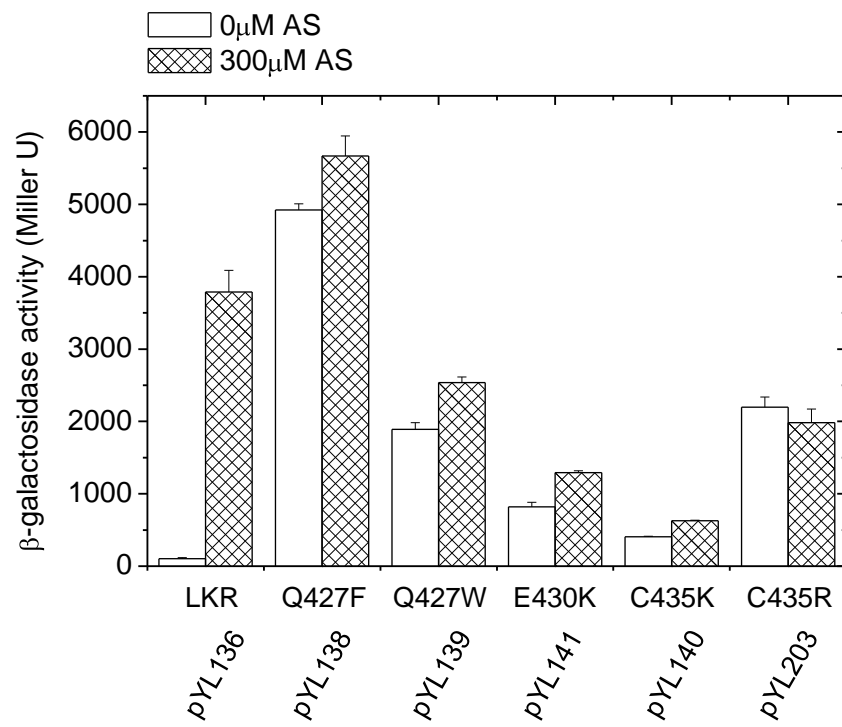


Ex. I426X: 5' --GCCAAGGCGAANNNCAGCTTCTTG--3'  
3' --CGGTTCCGCTTNNNGTCGAAGAAC--5'

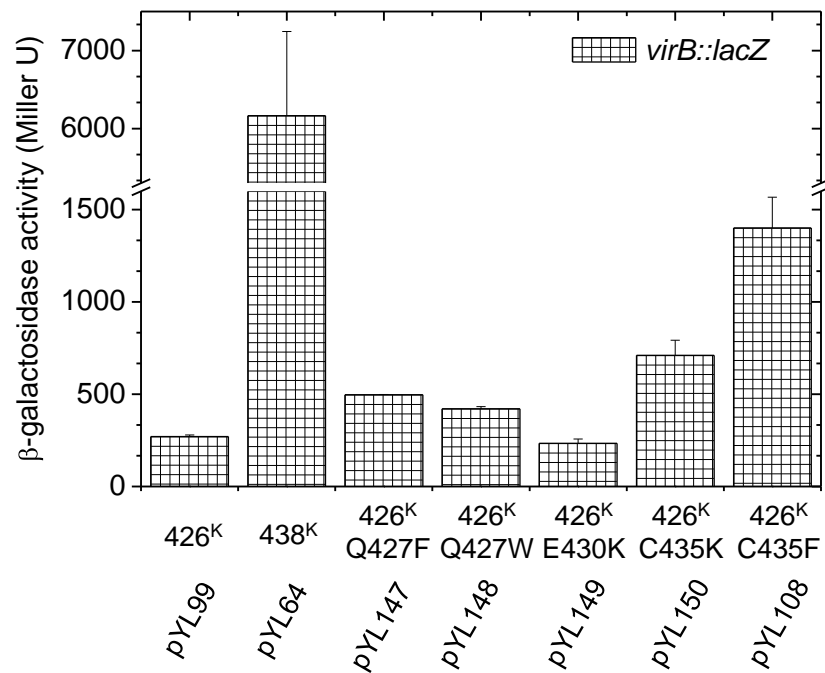


Non-inducing medium plate:  
no phenolic inducer  
Blue colony: constitutive ON

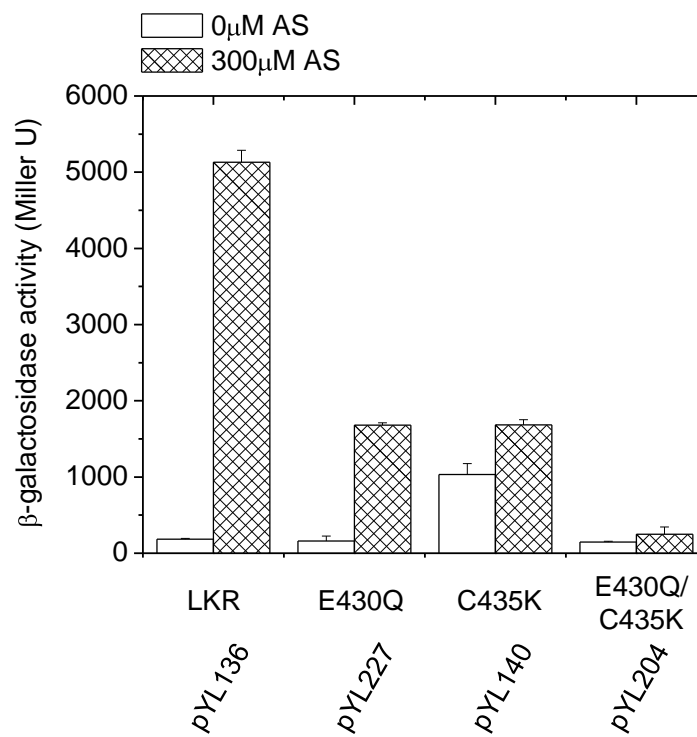


**(B)**

**Fig. 2.10**  $\beta$ -galactosidase activity of the constitutive on mutants in 426<sup>K</sup>. A. *tumefaciens* strain A136 carrying pRG109 and the indicated constitutive on mutations in 426<sup>K</sup> were assayed for *vir* gene expression without phenolic inducer. 438<sup>K</sup> was used as an additional activity control.

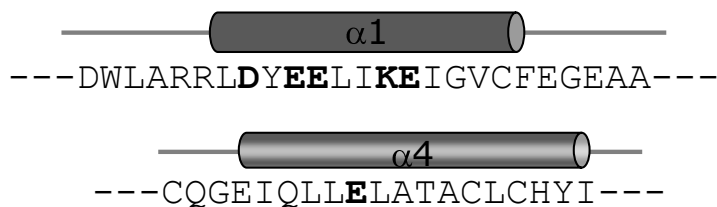


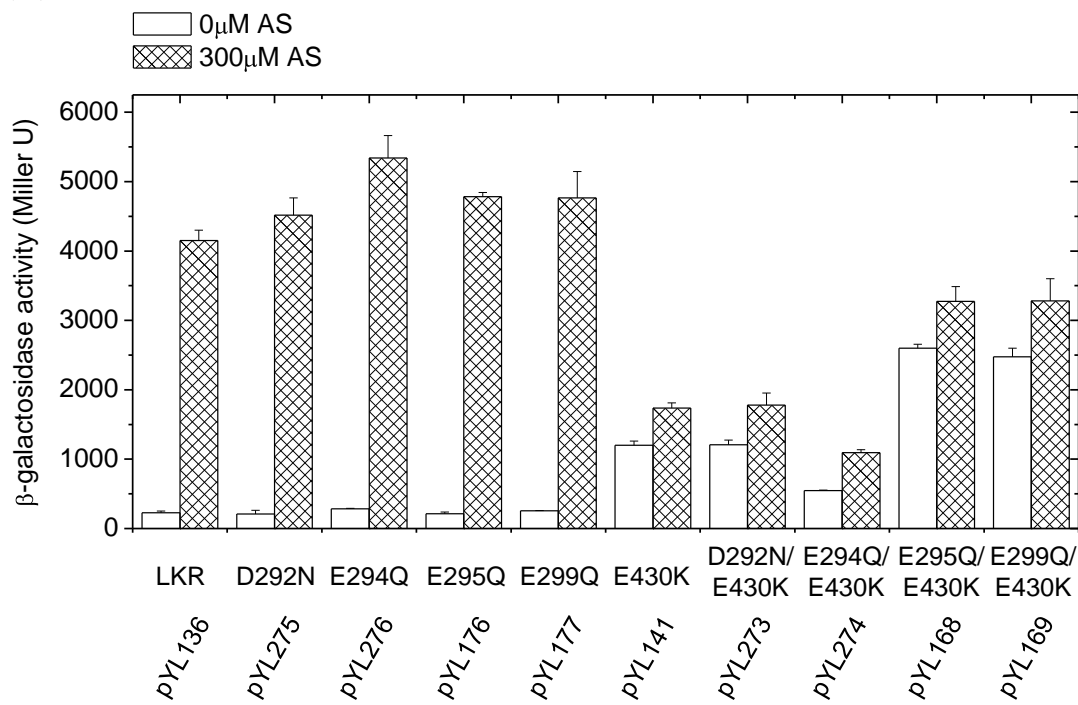
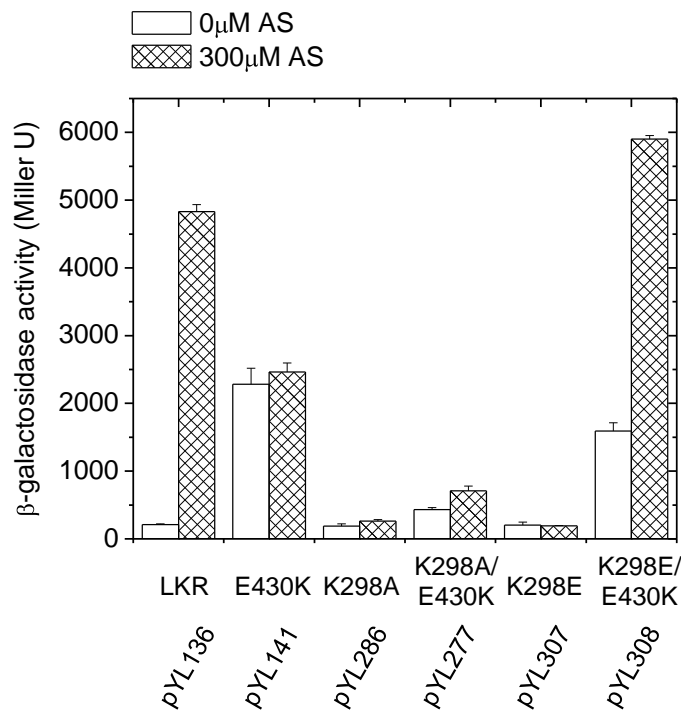
**Fig. 2.11 The constitutive on mutant C435K resulted from  $\alpha 4$ - $\alpha 4'$  salt-bridge formation.** *A. tumefaciens* strain A136 carrying pRG109 and the  $\alpha 4$  mutations in LKR were assayed for *vir* gene expression with or without 300  $\mu$ M AS.



**Fig. 2.12 The constitutive on mutation E430K resulted from  $\alpha 1$ - $\alpha 4$  salt-bridge formation.** **A)** Amino acid sequence of linker  $\alpha 1$  and  $\alpha 4$ . The charged residues are in bold. **B)** *A. tumefaciens* strain A136 carrying pRG109 and indicated LKR variants of double mutations of E430K and the negatively charged residues in  $\alpha 1$  were assayed with or without 300  $\mu$ M AS for *vir* gene expression, and **C)** double mutation of E430K and the positively charged K298 in  $\alpha 1$ .

(A)

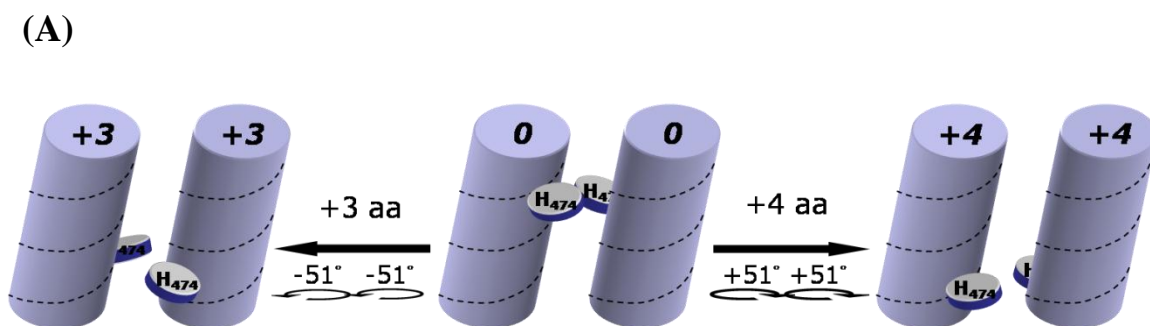


**(B)****(C)**

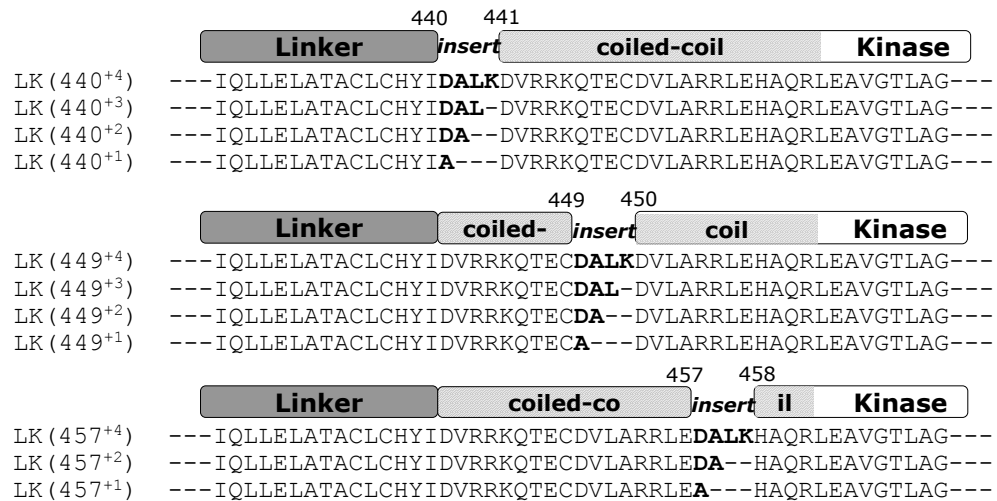
Phosphorylation and phosphoryl-transfer to VirG only increase upon signal sensing. Therefore, it seems that His474 is located at a position that could partially accept phosphate from the ATP-binding domain from the other subunit in an uninduced state, and, upon signal activation, the rotational motion as proposed above could move His474 to a position to increase phosphorylation as well as exposing it for phosphoryl-transfer to VirG. The *T. maritima* TM0853, homologous to *A. tumefaciens* VirA (Fig. 2.2), adopts a different *cis*-phosphorylation mechanism (Casino, Rubio et al. 2009). I hypothesized that if the phosphorylation is triggered by the proposed coiled-coil rotation, and if the direction of rotation can dictate the *trans*- or *cis*-phosphorylation mechanism, then it should now be possible to analyze these motions by direct amino acid insertion in the coiled-coil.

To probe the direction of active rotation in the helices of VirA, amino acids were directly inserted into the coiled-coil region of the kinase. Similar to the GCN4 fusions, assuming the intact linker domain provides the stabilization force of GCN4, a +3 insertion would extend the coiled-coil by one turn, and create a  $-51^\circ$  ( $+310^\circ$ ) rotation (Fig. 2.13A). Similarly, a +4 insertion would also extend the coiled-coil by one turn, and create a  $+51^\circ$  rotation. As seen in Fig. 2.13B and C, directly inserting amino acids at the beginning of the kinase coiled-coil ( $440^{+n}$ ) or in the middle of the coiled-coil ( $449^{+n}$ ) both ratcheted the kinase activity similar to the GCN4 chimeras. Amino acid insertion toward the end of the coiled-coil ( $457^{+n}$ ) was deleterious to the kinase activity. Moreover, the  $+51^\circ$  rotation ( $440^{+4}$ ,  $449^{+4}$ ) enhanced kinase activity, while the  $-51^\circ$  rotation ( $440^{+3}$ ,  $449^{+3}$ ) dramatically decreased kinase activity, meaning that  $+51^\circ$  seems to be the turn-on

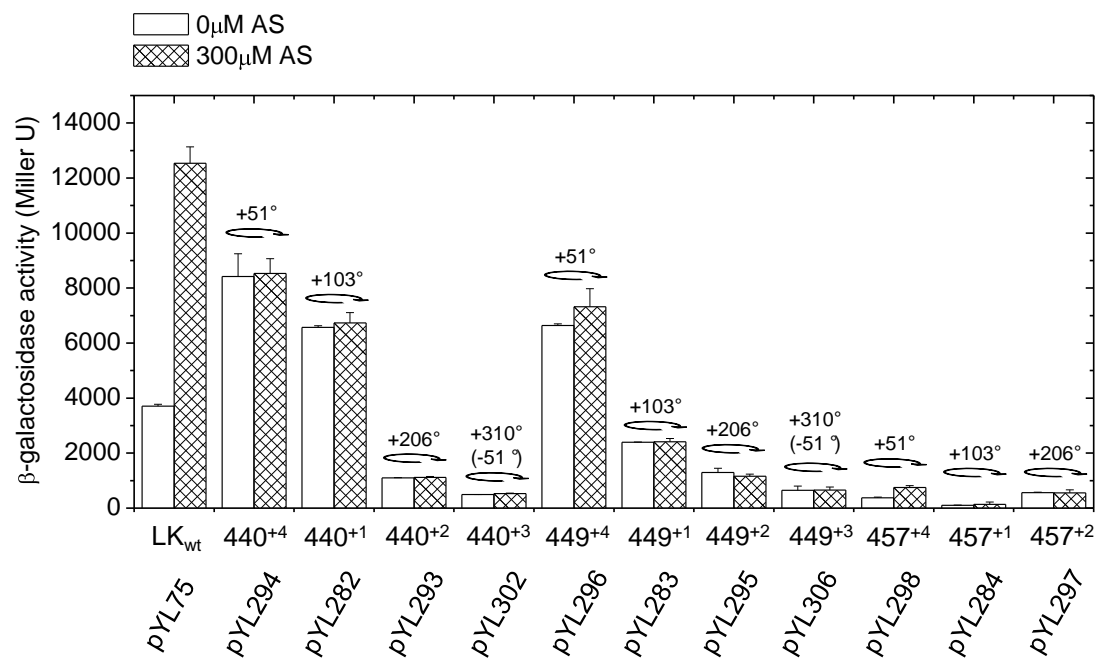
**Fig. 2.13 Direct amino acid insertion in kinase coiled-coil.** **A)** Schematic illustration of how His474 moves along the helix coil according to the amino acid insertion at the N-terminus. **B)** Experimental design of the direct amino acid insertions. Three regions of insertions were created: at the beginning of the kinase coiled-coil (440<sup>+n</sup>), in the middle of the coiled-coil (449<sup>+n</sup>), and at the end of the coiled-coil (457<sup>+n</sup>). The inserted amino acids are shown in bold. **C)** *A. tumefaciens* strain A136 carrying pRG109 and the LK constructs with the indicated insertions were assayed for *vir* gene expression with or without 300  $\mu$ M AS. The degree of rotation created by each constructs is shown in the graph.



(B)

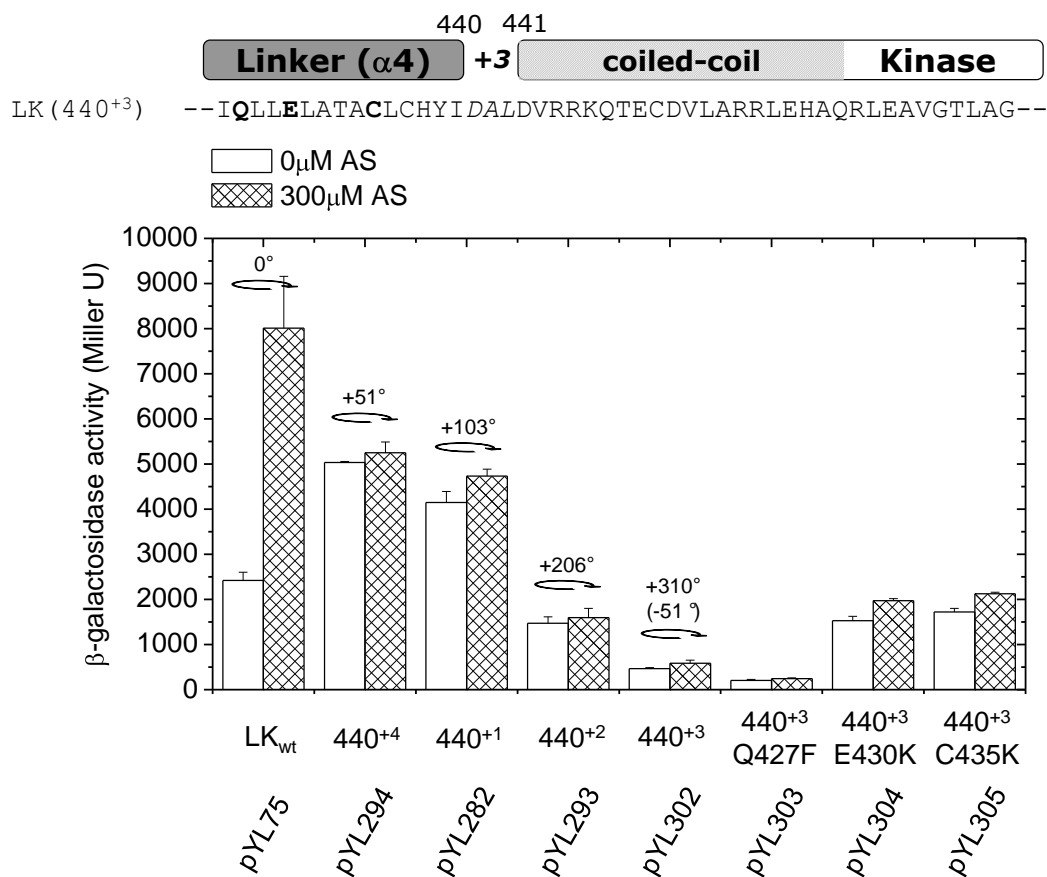


(C)





**Fig. 2.14 The linker  $\alpha 4$  constitutive “ON” mutations with the “OFF” kinase coiled-coil insertions.** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression with or without 300  $\mu$ M AS. LK(440<sup>+1/+2/+3/+4</sup>) constructs were the same as in Fig. 2.13C. The “OFF” insertion LK(440<sup>+3</sup>) were constructed with the linker  $\alpha 4$  constitutive “ON” mutations (Fig. 2.9B), shown in bold in the scheme.



rotation, and  $-51^\circ$  is the turn-off rotation. The constitutive mutations identified in the linker  $\alpha 4$  (Fig. 2.9) were incorporated with the off variant 440<sup>+3</sup> to determine whether the kinase activity can be rescued by coiled-coil stabilization. While E430K and C435K both rescued the off phenotype and part of the phenol inducibility, the Q427F mutation was not able to rescue the 440<sup>+3</sup> off phenotype (Fig. 2.14). These data establish that those constitutive mutations are able to overcome the insertions in the coiled-coil.

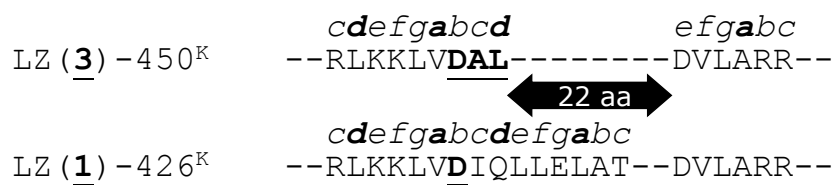
## Discussion

Many of the signaling cascades inside the cell involve a conserved catalytic core, coupled with diverse signal sensing modules. The flexibility and variety of signal perception converge at a point on the catalytic domain for one signal output. In this chapter, I have attempted to find this convergence point between the signal sensing domains and the histidine autokinase in *A. tumefaciens* VirA histidine kinase. A coiled-coil structure was identified at the signal sensing and kinase core interface, and was examined for general signal transmission. The stability of the coiled-coil motifs have been exploited in many different protein functions, ranging from regulating protein oligomerization, DNA binding, and signal transductions (Burkhard, Stetefeld et al. 2001; Lupas, Gruber et al. 2005). The COILS prediction algorithms suggested that VirA aa440-463 forms coiled-coil, and the experimental data showed that the  $\alpha 4$  of linker significantly repressed the kinase activity to the level of the uninduced LK (Gao and Lynn 2005). This observation is consistent with the argument that the kinase is constitutive, and the function of the regulatory domain is to repress activity until the signals are perceived (McCullen and Binns 2006).

VirA kinase coiled-coil is not predicted to have a single strong association interface, and this observation led to the proposed model for intermediate stability and activation through biasing the “ON” conformation of the coiled-coil. Among the molecular motions proposed for the parallel coiled-coil architecture, the rotational motion was initially examined via the fusion of GCN4 domains. The basic idea was to use the single registering stability of GCN4 with different adapters to shift the registry of the VirA kinase coiled-coil. Using GCN4 fusions, the smallest degree of rotation that could be mimicked is  $51^\circ$ , consistent with the knob-into-holes form of the coiled-coil (Cochran and Kim 1996). Nevertheless, significant kinase activity changes were mediated in all LZ-450<sup>K</sup>, LZ-438<sup>K</sup>, and LZ-426<sup>KON</sup> fusions. Moreover, from the LZ-450<sup>K</sup> fusions, the active position for the conserved histidine was narrowed to *a*, *e*, and *b* position of the heptad repeats (Fig. 2.6D), consistent with movement for *trans*-phosphorylation between subunits. The two ON conformations in GCN4-450<sup>K</sup> and GCN4-426<sup>K</sup> fusions, LZ(3)-450<sup>K</sup> and LZ(1)-426<sup>K</sup>, have a length difference of 22 amino acids, which generates one residue out of registry from the heptads (Fig. 2.15). This extra residue may be a regulation node for signal transmission, by forming a bulge in the coiled-coil, as observed in the structure of *Bacillus subtilis* histidine kinase DesK (Albanesi, Martin et al. 2009), or to form a stutter registry, which can also result in different coiled-coil conformations for signal transmission (Airola, Watts et al. 2010; Stewart and Chen 2010).

As powerful as the insight from coiled-coil analyses has been, the structural model for VirA suggests that the interface from the linker domain is processed through a dimer-dimer, or a four-helix bundle architecture. The repression mediated by  $\alpha 4$ , in that sense, may be resulting from a biasing of the partially “OFF” interface between  $\alpha 4$ - $\alpha 4'$ .

**Fig. 2.15** The two most active conformations of GCN4-450<sup>K</sup> and GCN4-426<sup>K</sup>. The amino acid sequence at the fusion point of LZ(3)-450<sup>K</sup> and LZ(1)-426<sup>K</sup> are shown with the adapter sequence in bold and underlined, and the heptad registry indicated from *a* to *g*. The length difference between the two forms is 22 residues.



The constitutive mutations found in the linker  $\alpha 4$  helix were most consistent with  $\alpha 4$ - $\alpha 4'$  interactions. Both hydrophobic (Q427F, Q427W, C435F) and electrostatic interaction (E430K, C435K, C435R) were found to stabilize the interface. Furthermore, the potential salt-bridge formation by K298-E430 extends the signal transmission region from the kinase coiled-coil to the four-helix bundle of the HAMP region of the linker domain. Other possible motions that couple with the critical rotational motion for kinase activation may also be possible. For example, the “piston-like” motion proposed for the charged residues in  $\alpha 1$  could modulate the energy barrier for  $\alpha 4$  rotation (Gao and Lynn 2007), and these ideas will be discussed in Chapter 4.

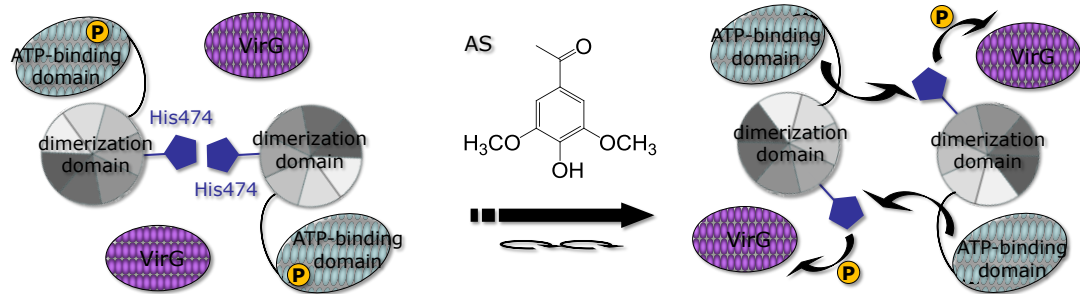
Further, the ratcheting regulation could control the kinase activity by the level of histidine phosphorylation or by the subsequent phosphoryl-transfer efficiency. Previous *in vivo* phosphorylation of the LZ(4/3/0)-426<sup>K</sup> fusions suggest the former mechanism is more likely (Gao and Lynn 2007). The structure of TM0853, a homologous HK of VirA, provides a model to test the possible mechanism for *trans*- vs. *cis*- phosphorylation. From the alignment of TM0853 and VirA, TM0853 has an additional residue between Gly466 and Thr467 of VirA (Fig. 2.2). This extra residue would shift the conserved histidine of TM0853 (His260) by  $+103^\circ$  relative to the position of His474 in VirA (Fig 2.16) if they have similar heptad repeats registry. The amino acid insertion (Fig. 2.13) suggested a clockwise rotation occurs for VirA activation. If VirA adopts a similar structure as that of TM0853, the clockwise rotation should bring the His474 of VirA closer to the ATP-binding domain of the other subunit for *trans*-phosphorylation, now consistent with the current data. On the other hand, His260 in TM0853, which locates  $+103^\circ$  relative to His474 of VirA because of the additional one-residue, will be turned by the same

rotational motion from an exposed surface to the ATP-binding domain of the same subunit, requiring the *cis*-phosphorylation mechanism. The two different mechanisms are illustrated in Fig. 2.16. Following this observation, a systematic analysis of the coiled-coil registry from other HKs should predict whether they adopt the *trans*- or *cis*-phosphorylation mechanism.

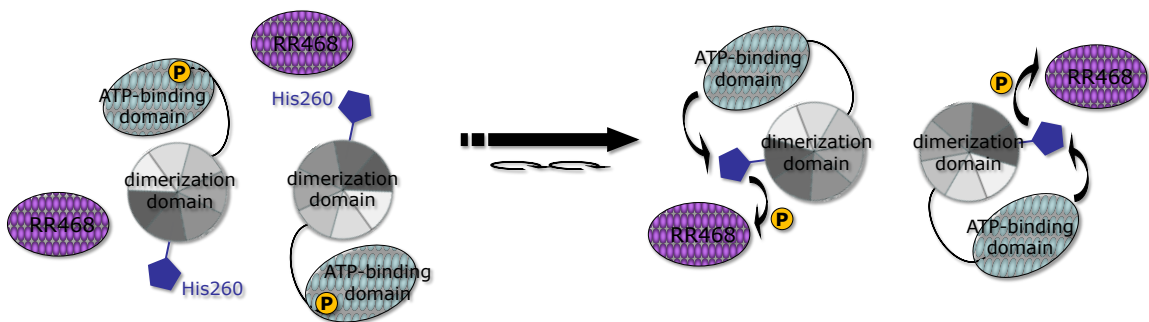
The dynamic conformational change that activates the kinase domain of VirA is quite remarkable. How these long-range conformational changes are initiated by small molecule binding remains unclear. Moreover, the actual mechanism of integrating the occupancy of multiple different signal molecules by these long-range motions remained a central question in my future studies.

**Fig. 2.16 Proposed phosphorylation model of VirA and TM0853.** In VirA, the conserved His474 is predicted to be buried in the dimerization interface. Upon sensing the inducing signal, the proposed clockwise rotation moves His474 to close proximity of the ATP-binding domain at the other subunit for *trans*-phosphorylation. In TM0853, the extra residue in the coiled-coil of kinase shifts the registry of the conserved His260 by one residue, which creates a +103° displacement of His260 relative to VirA's His474. Therefore, the same proposed rotation upon signal sensing will move the conserved His260 in TM0853 toward the ATP-binding domain at the same subunit for *cis*-phosphorylation.

### In VirA:



### In TM0853:





## **CHAPTER 3**

### **GENETIC, COMPUTATIONAL, AND BIOPHYSICAL CHARACTERIZATION OF PHENOL PERCEPTION OF THE LINKER DOMAIN**

#### **Introduction**

Chapter 2 described how the linker domain, as a GAF motor, mounts onto the VirA kinase core to transmit signal perception to activate phosphorylation. The  $\alpha 4$  of the linker domain, directly preceding the coiled-coil of the kinase core, is proposed to mediate coiled-coil rotation to activate the kinase. The signal transmission is mediated by critical phenol sensing at the linker domain. The predicted GAF fold of the linker domain predicts specific phenol perception by the linker, including a prediction of the phenol binding site, and these predictions have allowed for specific physical and experimental constraints to be developed. I will use these data to further extend the structural model for helical ratcheting as a mechanism of kinase control.

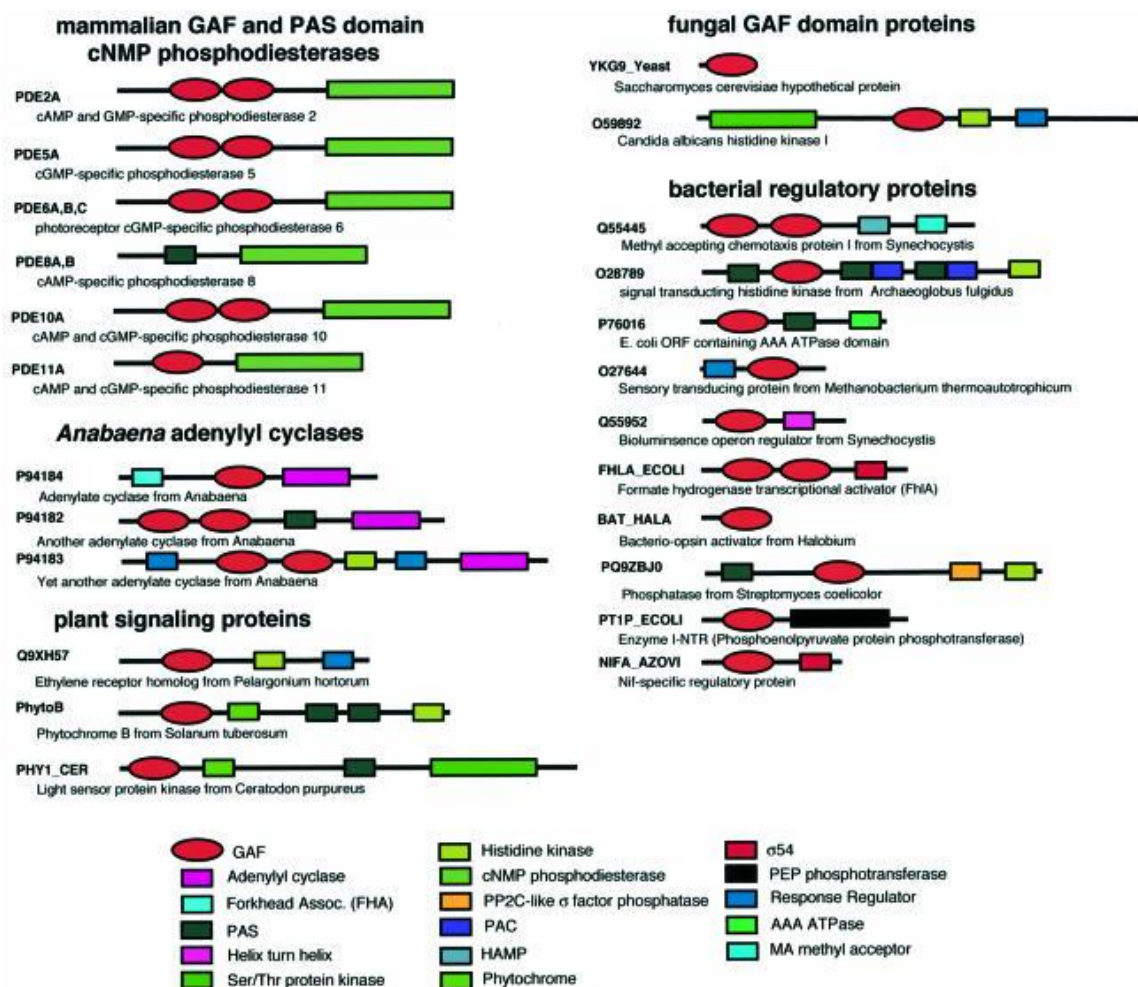
#### **The GAF domain**

The GAF domain was first identified as a homolog domain observed in plant and cyanobacterial phytochromes and as a cGMP-stimulated phosphodiesterase in vertebrates (Aravind and Ponting 1997). Since then, it has been largely assigned to mediate small molecule binding and regulation of various protein activities in all three kingdoms of life. There are now more than 10,000 GAF sequences identified in the Pfam database involved in more than 1000 different architectures (Finn, Mistry et al. 2010). Shown in Fig. 3.1 are several examples of the GAF domain architecture. Many GAF domains are coordinated

with histidine kinases and Ser/Thr protein kinases as well as regulating DNA binding. The majority of the GAF domains are used in secondary messenger proteins, such as the phosphodiesterase (PDE) and adenylyl cyclase, which bind cGMP or cAMP to allosterically regulate the functional activity.

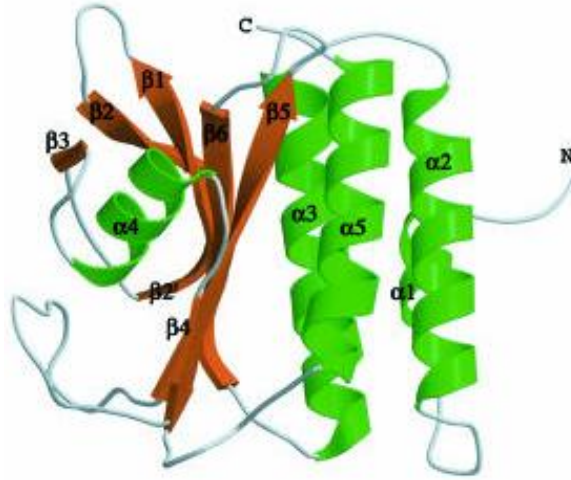
There are now 108 GAF-like structures available in the Pfam database. Shown in Fig. 3.2A is the first solved GAF domain structure, YKG9, from *Saccharomyces cerevisiae* (Ho, Burden et al. 2000). This structure, as well as other identified GAF structures, reveal a similar fold with central  $\beta$ -sheets flanked by a helix bundle region on one side and a more unstructured region at the other (Martinez, Beavo et al. 2002). Some of the GAF domains form dimers through hydrophobic interactions of the helices, but dimer formation does not seem to be required for activity. The GAF domains can coordinate with other regulatory domains for better regulation, and some GAF domains have tandem registry, such as the GAF-A/B domain observed in adenylyl cyclase (Fig. 3.2B)(Martinez, Bruder et al. 2005). Some GAF domains are observed covalently linked with other regulatory domains to provide multiple regulation points in response to additional signals; most commonly seen coordinating PAS domains. One of the structures obtained was the entire photoreceptor domain, adapting the PAS-GAF-phytochrome fold, in *Pseudomonas aeruginosa* (Fig. 3.2C). This entire photo-regulatory domain is mounted onto a histidine kinase output, and the similar topology of the three regulatory domains may suggest a common evolutionary origin between them (Yang, Kuk et al. 2008).

**Fig. 3.1 Representative GAF domain architectures in different organisms.** The GAF domain architecture observed in mammal, plant, bacteria, and fungus are illustrated. Representations of the functional and regulatory groups are indicated below. The GAF domains are drawn as red ovals (Ho, Burden et al. 2000).

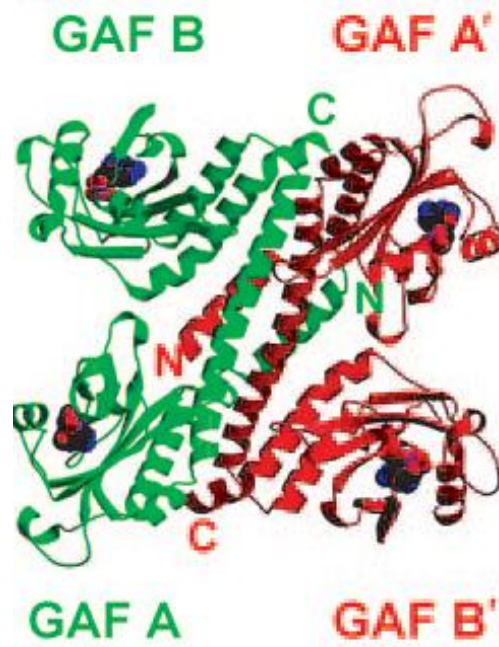


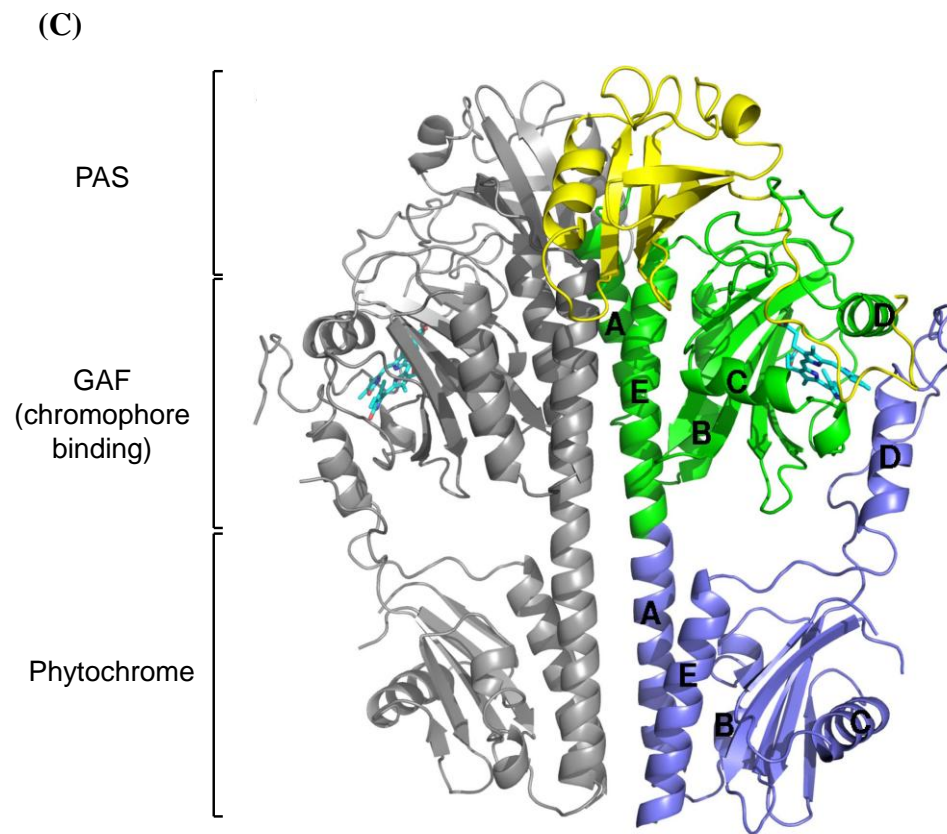
**Fig. 3.2 Crystal structure of GAF domains.** **A)** The crystal structure of YKG9, a homolog of *Saccharomyces cerevisiae* PDE5-GAFa. The structure was crystallized as a dimer, and shown here is the monomer structure with secondary structures labeled (Ho, Burden et al. 2000). **B)** The crystal structure of the tandem GAF domain of *cybB2* adenylyl cyclase in cyanobacterium *Anabaena* (Martinez, Bruder et al. 2005). The tandem GAF domain forms a dimer, and each monomer contains a cAMP binding site, shown in CPK color. **C)** The crystal structure of the fully photoactive photosensory core domain from *Pseudomonas aeruginosa* bacteriophytochrome. The entire photosensory domain forms a dimer, and is composed of the tandem linkage of a PAS (yellow), a GAF (green), and a phytochrome (blue) domain (Yang, Kuk et al. 2008). The chromophore shown in ball-and-stick (cyan) is bound to the GAF domain.

(A)



(B)



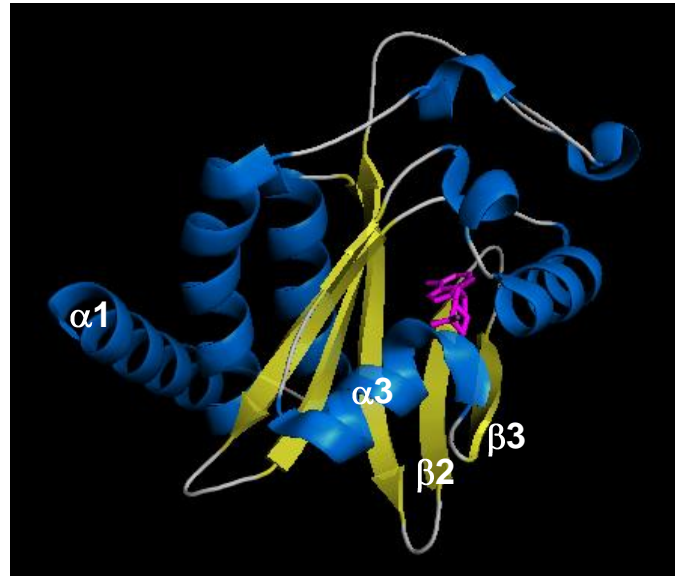


Similar to PAS domains, GAF domains can also accommodate structurally diverse ligands. This ability may be due to the less-structured region positioned on the other side of the  $\beta$ -sheets, a region which generally has the lowest conservation. Besides cyclic nucleotides, it binds the heme group in the NO/ O<sub>2</sub> responding two-component system DevS/DevR (Sardiwal, Kendall et al. 2005; Lee, Cho et al. 2008), specific chromophores in systems that respond to light (Cornilescu, Ulijasz et al. 2008). It also binds small molecules, such as in the nitrogen fixation protein NifA in *Azotobacter vinelandii* that binds to 2-oxoglutarate to regulate DNA binding affinity (Little and Dixon 2003), and the N-terminal domain of CodY that binds to branched chain amino acids and regulates *Bacillus subtilis* transition to stationary phase under nutrition depletion condition (Levdikov, Blagova et al. 2006). Fig. 3.3 shows two examples of ligand binding in GAF domains, and in both cases, some secondary structural features at the ligand binding site were observed. The cAMP binding in the GAF-B domain of *cydB2* adenylyl cyclase is located between  $\alpha 3$  and the  $\beta$ -sheets, and the  $\beta 2$ - $\beta 3$  loop was found to define substrate specificity (Fig. 3.3A). Fig. 3.3B shows the isoleucine binding site of the GAF domain of CodY, which lies between  $\alpha 3$ ,  $\alpha 4$ , and the  $\beta$ -sheets. According to ligand binding in those GAF structures, the linker domain of VirA may similarly bind to phenols in the region between the  $\beta$ -sheets and  $\alpha 3$ , the  $\alpha 3$ - $\beta 3$  loop, and the  $\beta 3$ - $\beta 4$  loop regions.

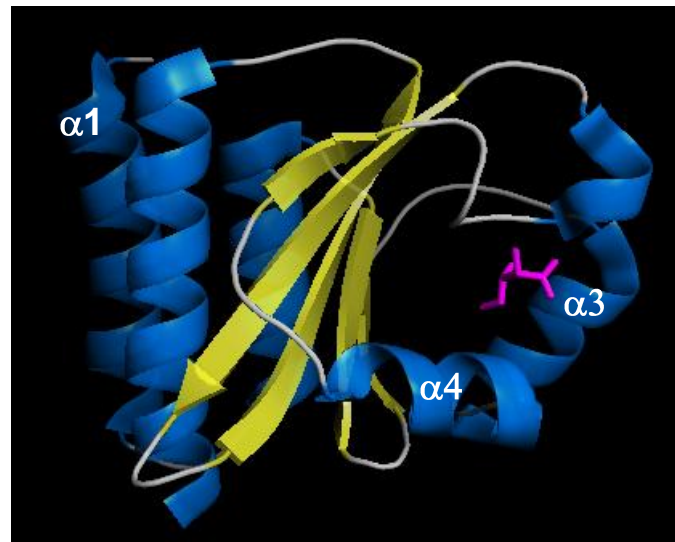
**Fig. 3.3 Crystal structure of GAF domains binding with their ligands.** The GAF structure is colored in blue for  $\alpha$ -helix and yellow for  $\beta$ -sheet; the ligand is shown in ball-and-stick and colored in magenta. **A)** The GAF-B domain of *cyaB2* adenylyl cyclase in *Anabaena* (PDBID: 1YKD). cAMP binds to the region between  $\alpha 3$  and the  $\beta$ -sheets. The  $\beta 2$ - $\beta 3$  loop, locating at the other side of the structure, constitutes the ligand specificity. **B)** The structure of the N-terminal domain of CodY in *Bacillus subtilis* (PDBID: 2B18). This domain folds into a GAF-like structure and binds to branched amino acids. Shown in the figure is an isoleucine binds at the pocket inside  $\alpha 3$ ,  $\alpha 4$ , and the  $\beta$ strands.



(A)



(B)



### **Phenol specificity of different *Agrobacterium* strains and Y293T**

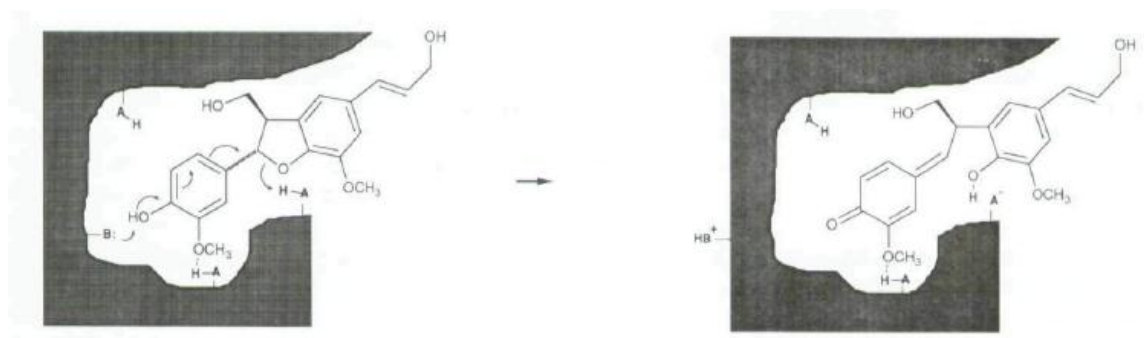
The structural features of the inducing phenols, as described in Chapter 1, are the conserved hydroxyl group, one or two *o*-methoxy groups, and a wide range of *para*-substituents ranging from –H to large phenyl-containing group (Spencer and Towers 1988; Joubert, Beaupere et al. 2002). Based on these characteristics, a phenolic receptor and a “proton transfer” inducing mechanism have been proposed (Fig. 3.4)(Duban, Lee et al. 1993; Lee, Tzeng et al. 1993). In this model, a hydrophobic cleft binds the aromatic structure of the phenol with a basic residue close to the hydroxyl group for proton transfer. Two acidic residues could further facilitate phenol binding by forming hydrogen bonds with the *o*-methoxies, and a large, flexible region that can accommodate structurally variable *para*-substituents would also contain an acidic residue for proton transfer. The proton transfer between the phenol and the receptor was proposed to cause a conformational change of the protein for kinase activation. In order to validate this model, both computational and genetic approaches have been used to search for the possible phenol binding site in this chapter.

Although most *Agrobacterium* strains contain the phenol specificity described above, including our experimentally examined A6, a strain discovered in South Korea, KU12, adapts a very different phenol specificity (Lee, Jin et al. 1995; Lee, Jin et al. 1996). In KU12, the phenols with *o*-methoxy groups are less potent than those without these substituents. For example, 4-hydroxyacetophenone (HAP), which is normally not an inducer for strain A6 and C58, has a higher inducibility in KU12 than acetocyringone (AS). The specificity may be modulated by a different binding site or a different signal

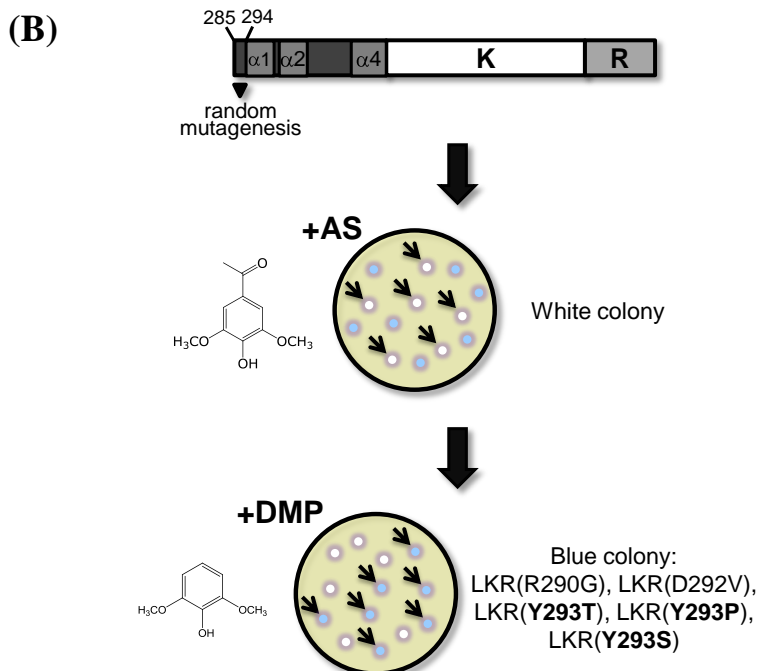
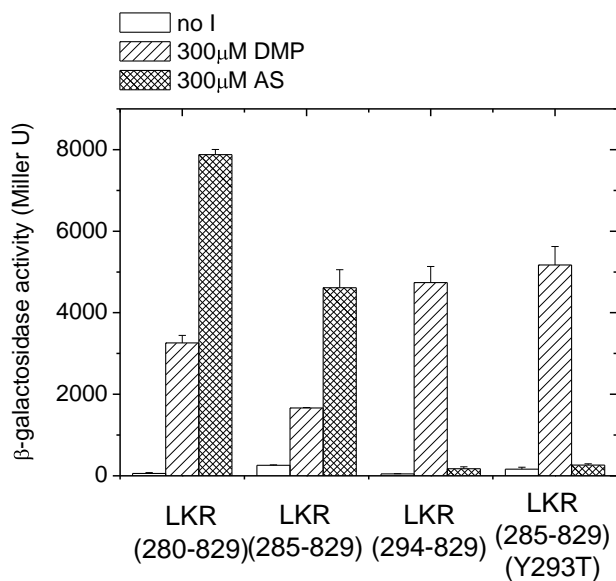
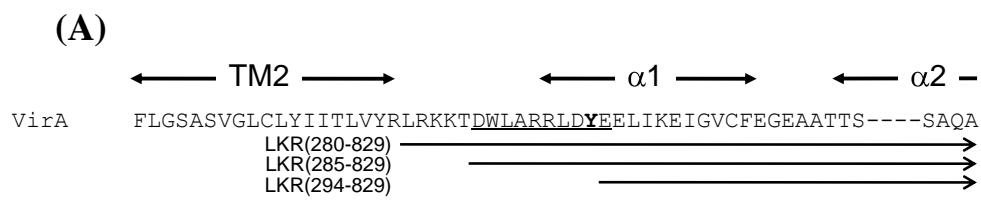
transmission mechanism. The full-length VirA amino acid sequence of KU12 has only 45% identity to A6, while that between C58 and A6 is 73% (Lee, Ha et al. 1998). The sequence identity of the VirA linker domain between KU12 and A6 is down to 40%, while that between C58 and A6 remains 73%. Therefore, the different phenol specificity may be the result of the evolved linker domain sequence, and how these differences can further determine the host range of *Agrobacterium* will be discussed in Chapter 4.

Regardless of the potential phenol binding site suggested by the GAF model, an interesting result of phenol specificity regulation was observed in VirA mutants (Fang, 2007). When the amino acids after the membrane spanning helix and before the predicted GAF fold of the linker domain were removed, resulting in LKR(294-829), a different phenol sensitivity was observed when compared with LKR(285-829). In this construct, the generic AS molecule was non-inducing, while the simple 2,6-dimethoxyphenol (DMP) was a potent inducer (Fig. 3.5A). It seems aa285-294 was regulating the perception of the *para*-acetyl group on phenol. To further identify the specific residue regulating such specificity, a random mutagenesis library was generated for aa285-294 and screened for the AS-non-inducible and 2,6-DMP-inducible variants (Fig. 3.5B). Among the mutations identified, Y293T displayed the most dramatic phenol specificity change (Fig. 3.5A). Even though there is no natural *Agrobacterium* strain identified with different phenol specificity towards *para*-substitutions, mutations were rationally designed based on these results. I end this chapter with several preliminary analyses of the purified linker domain to validate the GAF model.

**Fig. 3.4 Proposed phenol binding site in VirA and the “proton transfer” model.** The binding site is shown with the phenolic inducer dehydrodiconiferyl alcohol (DCA), which is bound at a hydrophobic site of the protein. The binding site also contains several hydrogen bond donor and acceptor for the proton transfer reaction. After proton transfer, the receptor undergoes a conformational change that induces kinase activity (Duban, Lee et al. 1993).



**Fig. 3.5 Phenol specificity and the library design.** **A)**  $\beta$ -galactosidase activity of different LKR constructs. The VirA amino acid sequence following TM2 is shown with the predicted linker GAF structures. The region for the random mutagenesis library is underlined, and the Y293 residue is shown in bold.  $\beta$ -galactosidase activity was analyzed in the absence of inducer or presence of 300  $\mu$ M 2,6-DMP or AS. **B)** The design of the sequential library screening for phenol specificity switch. LKR with aa285-294 randomly mutagenized was primarily screened on induction medium plate with 100  $\mu$ M AS. The white colonies were extracted, and screened again on induction medium plate with 100  $\mu$ M 2,6-DMP. The blue colonies, which represent the variants changed in phenol specificity, were selected from the second screening.  $\beta$ -galactosidase activity of one of the identified mutant, LKR(285-829)(Y293T), is shown in **(A)**.



## Materials and Methods

**SAM-T06 model.** The automated method SAM-T06 is used to generate structure models of the VirA linker domain of strain A6 and strain KU12 (Karplus, Barrett et al. 1998; Karplus, Karchin et al. 2003). The inquiry sequence of the A6 linker domain is aa280-449, and aa281-456 for strain KU12.

**Docking the phenols by Autodock.** Docking of phenolic compounds into the linker structure models were performed by Autodock 4.2 (Morris, Goodsell et al. 1998) and analyzed by AutoDockTools (ADT), developed by the Scripps Research Institute. The PDBQT files of the receptor and the ligand were prepared by ADT for docking. The linker (receptor) structure was set as rigid, and the phenol (ligand) was flexible with free rotatable bonds. AS has 4 rotatable bonds, 2,6-DMP has 3, and HAP has 2. The grid maps and the coordinates were prepared prior to docking by Autogrid with the parameters listed in Fig. 3.6A. The Genetic Algorithm was used as the conformational search algorithm, and the parameter setting is shown in Fig. 3.6B. The default values were used as the docking parameters. In each docking trial, 10 conformations with the lowest free energy were probed, and analyzed by ADT.

**Fig. 3.6 Parameters used in Autodock.** Autodock 4.2 was used to search for stable conformations of the phenolic compounds in linker. Two linker structures were used as the docking receptor: one from *Agrobacterium* strain A6 and the other from strain KU12. The parameters defining the grid box in the linker structure are listed in **A**), and the Genetic Algorithm parameters used for docking are shown in **B**).

**(A)**

	A6	KU12
Total grid points	197945	197945
# of points in x	60	60
# of points in y	54	54
# of points in z	58	58
Spacing (Å)	0.75	0.75
Center of grid box	(-3.435, -0.802, -1.625)	(-0.347, -0.464, 0.444)

**(B)**

**Genetic Algorithm Parameters**

Number of GA Runs: 10

Population Size: 150

Maximum Number of evals: short ▼ 250000

Maximum Number of generations: 27000

Maximum Number of top individuals that automatically survive: 1

Rate of Gene Mutation: 0.02

Rate of Crossover: 0.8

GA Crossover mode: twopt ▼

Mean of Cauchy distribution for gene mutation: 0.0

Variance of Cauchy distribution for gene mutation: 1.0

Number of generations for picking worst individual: 10

Accept Close



**Bacterial strains, plasmids, and reagents.** The bacterial strains and plasmids used in this chapter are listed in Table 3.1. *E. coli* strain XL1-Blue (Stratagene) was used for routine plasmid construction, and strain M15[pREP4] (Qiagen) was used for protein over-expression. The phenolic compounds used for *vir* gene induction were purchased from Sigma-Aldrich Corp, and their chemical structures are listed in Table 3.3. The IPTG used to induce protein expression was purchased from Research Products International Corp. The chemical cross-link reagent dimethyl suberimidate (DMS) was purchased from Pierce. The cloning reagents were purchased from New England Biolab and Promega.

**Table 3.1 Bacterial strains and plasmids used in Chapter 3**

Strains/plasmids	Relevant characteristics	Reference
<i>E. coli</i> strains		
XL1-Blue	<i>recA1 endA1 gyrA96 thi1 hsdR17 supE44 relA1 lac[F' proAB lacI<sup>q</sup>Z M15 Tn10 (Tc<sup>r</sup>)]</i>	Stratagene
M15[pREP4]	Derived from <i>E. Coli</i> strain K12 and have the phenotype NaI <sup>S</sup> , Str <sup>S</sup> , Rif <sup>S</sup> , Thi <sup>-</sup> , Lac <sup>-</sup> , Ara <sup>+</sup> , Gal <sup>+</sup> , Mtl <sup>-</sup> , F <sup>-</sup> , RecA <sup>+</sup> , Uvr <sup>+</sup> , Lon <sup>+</sup>	Qiagen
<i>A. tumefaciens</i> strains		
A136	Strain C58 cured of pTi plasmid	(Watson, Currier et al. 1975)
Plasmids		
pQE30	<i>P</i> <sub>N25</sub> -His <sub>6</sub> -MCS-STOP, pBR322 <i>ori</i> , Ap <sup>r</sup>	Qiagen
pYW15b	Broad-host-range expression vector, IncW, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pYW48	<i>P</i> <sub>virA</sub> - <i>virA(aal-829)</i> in pYW15b, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pSW209Ω	<i>virB::lacZ</i> , IncP, Spec <sup>r</sup>	(Wang,

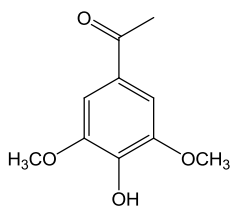
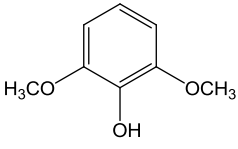
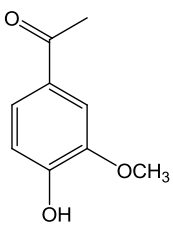
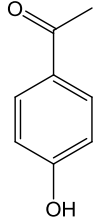
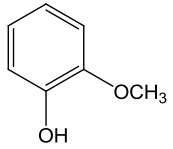
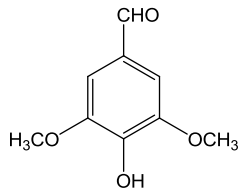
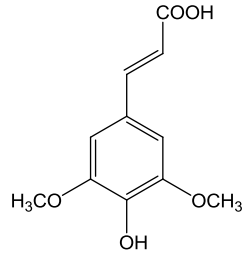
		Mukhopadhyay et al. 2000)
pJZ4	<i>P<sub>virB</sub>-lacZ</i> in pMON596, IncP, Spec <sup>r</sup>	(Zhang, Boone et al. 2000)
pJZ6	IncW/ColE expression vector with <i>P<sub>N25</sub></i> , Ap <sup>r</sup>	(Zhang, un- published)
pRG15	<i>P<sub>N25</sub>-His<sub>6</sub>-virA(aa285-471)</i> in pBS, Ap <sup>r</sup>	(Wang, Gao et al. 2002)
pRG109	<i>P<sub>N25</sub>-His<sub>6</sub>-virG</i> in pJZ4, Spec <sup>r</sup>	(Gao and Lynn 2005)
pQF282	<i>P<sub>N25</sub>-virA(aa285-829)(Y293T)</i> in pJZ6, Ap <sup>r</sup>	(Fang 2007)
pQF375	<i>P<sub>N25</sub>-virA(aa280-829)</i> in pJZ6, Ap <sup>r</sup>	(Fang 2007)
pQF417	<i>P<sub>N25</sub>-virA(aa294-829)</i> in pJZ6, Ap <sup>r</sup>	(Fang 2007)
pYL3	<i>P<sub>N25</sub>-virA(aa712-829)</i> in pYW15b, Ap <sup>r</sup>	This chapter
pYL67	<i>virA(aa280-455)</i> in pQE30, Ap <sup>r</sup>	This chapter
pYL69	<i>P<sub>N25</sub>-virA(aa285-444)</i> in pQE30, Ap <sup>r</sup>	This chapter
pYL75	<i>P<sub>N25</sub>-virA(aa285-711)</i> in pJZ6, Ap <sup>r</sup>	Chapter 2
pYL136	<i>P<sub>N25</sub>-virA(aa285-829)</i> in pJZ6, Ap <sup>r</sup>	Chapter 2
pYL160	<i>P<sub>N25</sub>-virA(aa285-829)(W341A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL161	<i>P<sub>N25</sub>-virA(aa285-829)(W341F)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL162	<i>P<sub>N25</sub>-virA(aa285-829)(W355A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL163	<i>P<sub>N25</sub>-virA(aa285-829)(W355F)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL173	<i>P<sub>N25</sub>-virA(aa285-829)(F376A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL221	<i>P<sub>N25</sub>-virA(aa285-829)(F346A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL223	<i>P<sub>N25</sub>-virA(aa285-829)(F346Y)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL224	<i>virA(aa280-455)(W341A)</i> in pQE30, Ap <sup>r</sup>	This chapter
pYL233	<i>P<sub>N25</sub>-virA(aa285-829)(L431R)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL237	<i>virA(aa280-455)(Y414W)</i> in pQE30, Ap <sup>r</sup>	This chapter
pYL238	<i>P<sub>N25</sub>-virA(aa285-829)(F324A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL239	<i>P<sub>N25</sub>-virA(aa285-829)(F325A)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL246	<i>P<sub>N25</sub>-virA(aa285-829)(E430W)</i> in pJZ6, Ap <sup>r</sup>	This chapter
pYL286	<i>P<sub>N25</sub>-virA(aa285-829)(K298A)</i> in pJZ6, Ap <sup>r</sup>	Chapter 2
pYL299	<i>P<sub>N25</sub>-virA(aa285-711)(W341A)</i> in pJZ6, Ap <sup>r</sup>	This chapter

pYL300	$P_{N25}$ - <i>virA(aa285-711)</i> (W355A) in pJZ6, Ap <sup>r</sup>	This chapter
pYL313	$P_{N25}$ - <i>virA(aa285-711)</i> (F346A) in pJZ6, Ap <sup>r</sup>	This chapter
pYL315	$P_{N25}$ - <i>virA(aa285-711)</i> (K298A) in pJZ6, Ap <sup>r</sup>	This chapter

**Table 3.2 Primers used in Chapter 3**

Primer	Characteristic	Sequence (5' → 3') ( <u>RE site</u> ; <i>mutation</i> )
LKR285	BamHI_285_for	CG <u>GGA TCC</u> GAT TGG TTA GCG CGG CGT
LKRA1	KpnI_stop_829_rev	GC <u>GGT ACC</u> GCA ACT CTA CGT CTT GAT
LKKpnI	KpnI_stop_711_rev	GG <u>GGT ACC</u> CTA ACG CGG TGC CTT ATT GCG
EcoRIS	EcoRI_T5_for	GCA <u>GAA TTC</u> ATT AAA GAG GAG AA
YL5	SacI_712_for	GCT <u>GAG CTC</u> GGA AAC GGG GAG ATT GTG GC
YL6	KpnI_stop_829_rev	GCT <u>GGT ACC</u> TGA CAC GTC GCA ACT CTA CGT C
YL103	KpnI_stop_444_rev	CC <u>GGT ACC</u> TCA ACG CCG AAC ATC GAT ATA G
YL104	KpnI_stop_455_rev	CC <u>GGT ACC</u> TCA TCG TCT GGC CAA AAC GTC
YL219	W341A_for	CAT GAC CGT AGA <i>GCG</i> GCT GTC GAA ACA
YL220	W341A_rev	TGT TTC GAC AGC <i>CGC</i> TCT ACG GTC ATG
YL221	W341F_for	CAT GAC CGT AGA <i>TTC</i> GCT GTC GAA ACA
YL222	W341F_rev	TGT TTC GAC AGC <i>GAA</i> TCT ACG GTC ATG
YL223	W355A_for	C CCA AAA CCT GTG <i>GCG</i> GAC GAC AGC GTG
YL224	W355A_rev	CAC GCT GTC GTC <i>CGC</i> CAC AGG TTT TGG G
YL225	W355F_for	C CCA AAA CCT GTG <i>TTC</i> GAC GAC AGC GTG
YL226	W355F_rev	CAC GCT GTC GTC <i>GAA</i> CAC AGG TTT TGG G
YL234	F376A_for	CGG GCG ACG GTA <i>GCC</i> CGC ATC ATA TCG
YL235	F376A_rev	CGA TAT GAT GCG <i>GGC</i> TAC CGT CGC CCG
F346AF	F346A_for	CT GTC GAA ACA <i>GCC</i> GGT GCG AAA CAC CC
F346AR	F346A_rev	GG GTG TTT CGC ACC <i>GGC</i> TGT TTC GAC AG
F346YF	F346Y_for	CT GTC GAA ACA <i>TAC</i> GGT GCG AAA CAC CC
F346YR	F346Y_rev	GG GTG TTT CGC ACC <i>GTA</i> TGT TTC GAC AG
YL295	Y414W_for	GT TCA CTG GGT <i>TGG</i> CAA AGC TAT CGC
YL296	Y414W_rev	GCG ATA GCT TTG <i>CCA</i> ACC CAG TGA AC
YL301	L431R_for	G CTT CTT GAA <i>CGC</i> GCC ACC G
YL302	L431R_rev	C GGT GGC <i>GCG</i> TTC AAG AAG C
YL307	F324A_for	CAG CGC <i>GCC</i> TTT GAT GCC GAT AC
YL308	F324A_rev	GT ATC GGC ATC AAA <i>GGC</i> GCG CTG
YL309	F325A_for	CAG CGC TTC <i>GCT</i> GAT GCC GAT AC
YL310	F325A_rev	GT ATC GGC ATC <i>AGC</i> GAA GCG CTG
YL333	E430W_for	CAG CTT CTT <i>TGG</i> CTC GCC ACC GCC
YL334	E430W_rev	GGC GGT GGC GAG <i>CCA</i> AAG AAG CTG

**Table 3.3 Chemical structures of phenolic inducers used in Chapter 3**

Full name	Short name	Chemical structure
Acetosyringone	AS	
2,6-dimethoxyphenol	DMP	
Acetovanillone	AV	
4-hydroxyacetophenone	HAP	
Guaiacol	GA	
Syringaldehyde	SAE	
Sinapinic acid	SPA	

**Cloning, Transformation, and growth conditions.** Same as in Chapter 2.

**Plasmid constructions.** The plasmids used in this chapter are listed in Table 3.1. Site-directed *virA* mutants were generated by the same two-step PCR method described in Chapter 2 with the primers listed in Table 3.2. The LKR mutants were generated by PCR from pYW48 with LKR285, LKRA1, and the corresponding complementary primers to introduce the mutations. Following *Bam*HI and *Acc*65I digestion, the DNA fragment was introduced into *Bam*HI and *Acc*65I digested pJZ6, generating pYL160, pYL161, pYL162, pYL163, pYL173, pYL221, pYL223, pYL233, pYL238, pYL239, and pYL246. The LK mutants W341A, W355A, F346A, and K298A were generated by PCR using pYL160, pYL162, pYL221, and pYL286 as the templates with LKR285 and LKKpnI, and introduced into pJZ6 with *Bam*HI and *Acc*65I sites, resulted in pYL299, pYL300, pYL313, and pYL315.

N-terminal His<sub>6</sub>-tagged receiver domain (*virA*712-829) was generated by PCR from pYW48 with primers YL5 and YL6, followed by *Sac*I and *Kpn*I digestion, and ligated into *Sac*I and *Kpn*I digested pYW15b to generate pYL3. His<sub>6</sub>-tagged linker domain (*virA*280-455) was generated by PCR from pQF375 with primers EcoRIS and YL104, followed by *Bam*HI and *Kpn*I digestion and ligated into *Bam*HI and *Kpn*I digested pQE30 to generate pYL67. His<sub>6</sub>-tagged linker (*virA*285-444) was generated by PCR from pYW48 with primers LKR285 and YL103, followed by *Bam*HI and *Acc*65I digestion and ligated into *Bam*HI and *Acc*65I digested pQE30 to generate pYL69. *virA*(280-455)(W341A) and *virA*(280-455)(Y414W) were generated by the two-step PCR using pYL67 as the template and EcoRIS, YL104, and the corresponding complementary

primers to introduce the mutations. Followed by *EcoRI* and *Acc65I* digestion, the DNA fragment was introduced into pQE30 to generate pYL224 and pYL237. The nucleotide sequence of all the plasmid constructs were confirmed by DNA sequencing facility performed by Agencourt Genomic Service (now Beckman Coulter Genomics).

**$\beta$ -galactosidase assays for *vir* gene induction.** All of the VirA variants were transformed into *A. tumefaciens* strain A136 containing pRG109, which carries *P<sub>virB</sub>-lacZ* and *P<sub>N25</sub>-virG*, for *vir* gene expression. The *A. tumefaciens* strains were grown and pelleted with the same procedure as described in Chapter 2, and diluted to OD<sub>600</sub> ~ 0.1 into tubes containing a total of 1 mL induction medium with or without the phenolic compounds at the indicated concentration, and cultured at 28 °C, 225 rpm for 15 hrs. The  $\beta$ -galactosidase activity was determined by the same method as described in Chapter 2, from the reading of the optical densities at 600 nm and 415 nm.

**Immunoblot analysis.** Same as in Chapter 2.

**Protein expression and purification.** *E. coli* M15[pREP4] strain carrying the indicated linker or receiver constructs were grown in LB medium at 37 °C, 200 rpm overnight. The overnight culture was diluted 1:100 into fresh LB medium, and subcultured at 37 °C, 200 rpm to an OD<sub>600</sub> of 0.6-0.7. The cell culture was cooled down to 20 °C, and 0.05 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) was added to induce protein overexpression. Cells were grown at 20 °C, 200 rpm, for an additional 14-18 hours, and harvested by centrifugation at 4 °C, 6000 rpm, for 10 min. The cell pellet was washed with 0.9% NaCl,

re-suspended in lysis buffer, and lysed on ice by sonication (ON pulse: 1 sec; OFF pulse: 2 sec; total pulse: 2 min). The clear lysate was obtained by centrifugation at 4 °C, 9000 x g, for 30 min. The His<sub>6</sub>-tagged protein was purified with Ni-NTA affinity column following the protocol in *The Qiaexpressionist* (Qiagen). The purified protein was exchanged into PBS buffer, pH 6.4, by the 10K Amicon Ultra (Millipore). All the buffers for purification and protein storage contain an additional 10% glycerol to enhance protein stability.

**Circular Dichroism.** 40 μL purified protein was placed into a quartz cuvette with 0.1 mm pathlength. Each spectrum was obtained by the average of three successive scans from 260 nm to 190 nm at a scanning rate of 100 nm/min with a resolution of 0.2 nm, using a Jasco-J180 spectropolarimeter. The ellipticity  $[\theta]_{\text{obs}}$  was converted to the mean-residue-ellipticity ( $[\theta]$ , deg cm<sup>2</sup> dmol<sup>-1</sup>) by the equation  $[\theta] = ([\theta]_{\text{obs}} \times 100 \times Mr) / (c \times l \times N_A)$ , in which  $Mr$  is the molecular weight of the protein in Dalton,  $c$  is the protein concentration in mg/mL,  $l$  is the pathlength of the cuvette in cm, and  $N_A$  is the number of the residues of the protein.

**Tryptophan fluorescence.** The fluorescence spectrum was measured by HORIBA Jobin Yvon FluoroMax-3. 600 μL of the purified protein was placed into a quartz cuvette with 0.2 cm pathlength. The sample was excited at 295 nm (slit: 5 nm) and the emission spectrum was acquired by scanning from 310 nm to 400 nm (slit: 5 nm) at a scanning rate of 1 nm/sec.

**Chemical cross-link.** The purified linker and receiver domains were cross-linked with dimethyl suberimidate (DMS) at the ratio of 1:20 and 1:50 (m/v). The reaction was performed at room temperature for 1 hour before quenched by the addition of SDS-PAGE sample buffer. The samples were boiled for 10 minutes, followed by 15% SDS-PAGE analysis and electro-blotting onto nitrocellulose membrane. The membrane was blocked with 3% BSA in TBS, and probed with the monoclonal anti-polyhistidine antibody (Sigma). Visualization was achieved using the goat anti-mouse antibody conjugated with alkaline phosphatase (Pierce) at 1:1000 dilutions, followed by the 1-step NBT/BCIP development (Pierce).

## Results

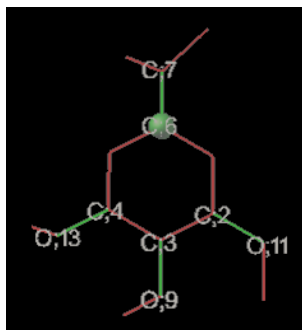
### Computational model of phenol binding in GAF

The GAF model of the linker domain, as based on the homology model, suggested that the ligands bind at the less-structured region of the protein. To further probe the possible phenol binding site, Autodock 4.2, which uses a semiempirical free energy force field to evaluate the conformations during docking simulations, was used to dock AS across the entire linker domain (Fig. 3.7). The first trial successfully located 10 conformations with the lowest free energy, and the results were compiled in Fig. 3.7C. The most stable conformation has the binding energy of -3.93 kcal/mol. Among the 10 conformations, three are close to the  $\alpha 3$ - $\beta 3$  region and one in the  $\beta 3$ - $\beta 4$  loop, all close to the predicted GAF sites. Two sites were sandwiched in the region between the  $\beta$ -sheets and  $\alpha 4$ , two conformations were close to the  $\beta$ -sheets and  $\alpha 2$ , and the other two were peripheral to the loops.

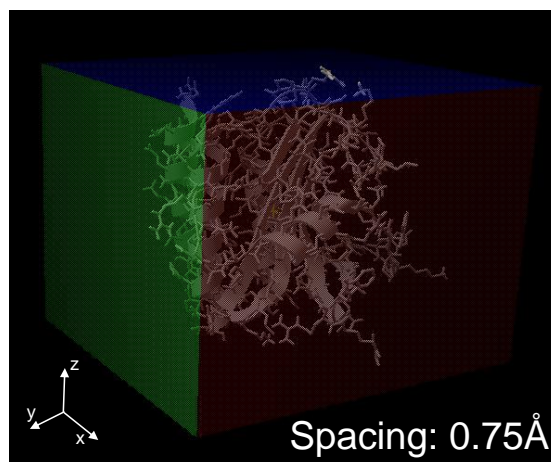


**Fig. 3.7 Computational docking of phenolic compounds into the modeled linker structure.** **A)** AS structure is shown with the root atom for grid points docking (C6) as a green ball. The rotatable bonds are in green and non-rotatable bonds are in red. **B)** The grid box defining the linker region docked by AS. The x-axis is red, y-axis is green, and z-axis is blue. The “blind docking” covers the entire linker domain, with the spacing between each grid point set at 0.75 Å. **C)** The docking result of AS → linker. 10 most stabilized conformations were compiled and shown in green. The linker structure is colored the same as in Fig. 1.6B with secondary structures labeled. Two different views are shown, one from the side and the other from the top of the linker domain.

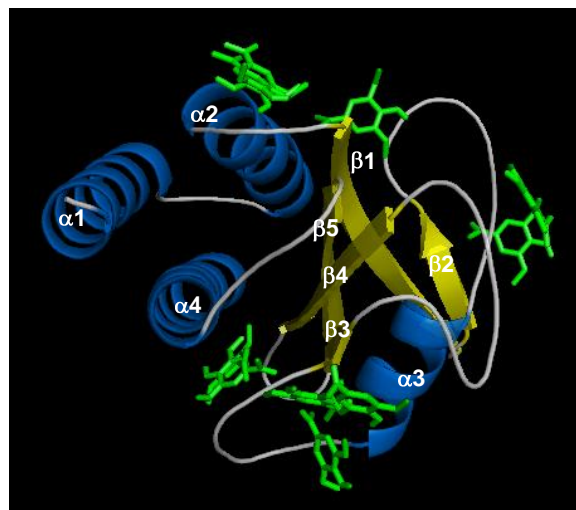
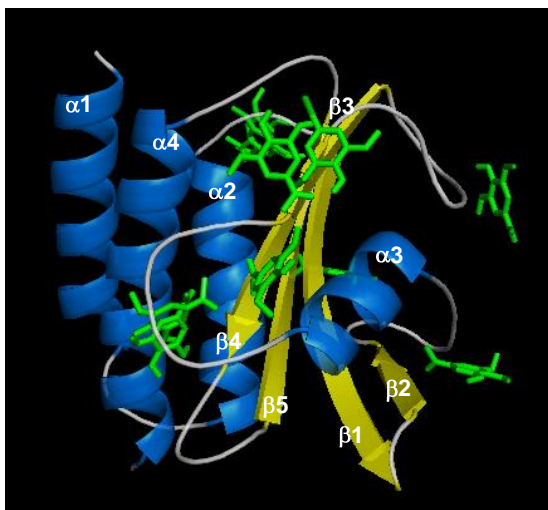
(A)



(B)



(C)



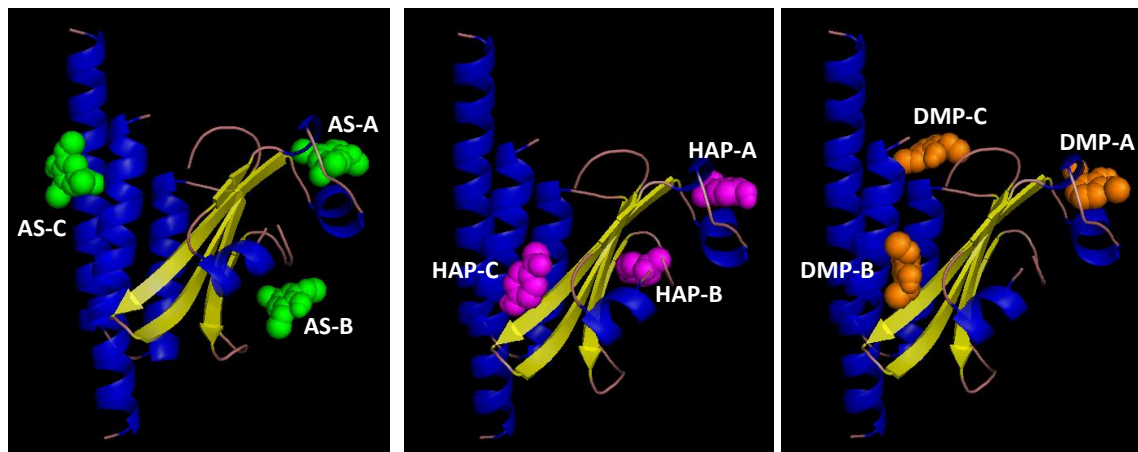
The AS docking sites of the first trial were mostly peripheral to the linker structure, and many of them did not overlap. This could be due to a bias of our threaded structure model as the docking results depend greatly on the receptor structure. In order to evaluate the variation caused from the protein structure, the docking was also performed on the linker model generated by the automated method, SAM-T06, using the sequence of VirA<sub>A6</sub>(280-449) as the input, when GAF structure 3P01 was automatically detected as the threading template. As seen in Fig. 3.8A and C, the SAM-T06 generated structure has similar secondary architecture and folding to the previous structural model, except for the prediction of two additional helices located between  $\beta$ 3 and  $\beta$ 4. Docking AS to the SAM-T06 generated structure had a generally lower binding energy, with the most stable conformation being -4.18 kcal/mol. Moreover, most of the AS conformations overlap with some others, resulting into several focused binding sites. As seen in Fig. 3.8A, three major binding sites are close to the newly predicted helices between  $\beta$ 3 and  $\beta$ 4 (AS-A), close to  $\alpha$ 3 (AS-B), and at the helix bundle region (AS-C). The amino acids within 5 Å of each AS conformation are shown in Fig. 3.8C. This SAM-T06 generated structure was further used to dock with two other phenolics HAP and 2,6-DMP. As seen in Fig. 3.8A, the docking of HAP resulted in a single common binding site of AS (HAP-A), one in the GAF-predicted binding pocket (HAP-B), and one site between the  $\beta$ -sheets and the helix bundle (HAP-C). The docking of 2,6-DMP also resulted in binding sites similar to AS and HAP molecules (DMP-A), a site between the  $\beta$ -sheets and helix bundle (DMP-B), and a site close to  $\alpha$ 2 (DMP-C). The amino acids located within 5 Å of each phenolic conformation are shown in Fig. 3.8C, and taken together, the docking of phenols

to the SAM-T06 generated linker structure resulted in more stable and focused sites. Some of the sites are conserved among the three different phenols (AS-A, HAP-A, and DMP-A), some non-conserved but close to the GAF predicted region (AS-B, HAP-B), and some non-conserved sites are located on the other side of the  $\beta$ -sheets.

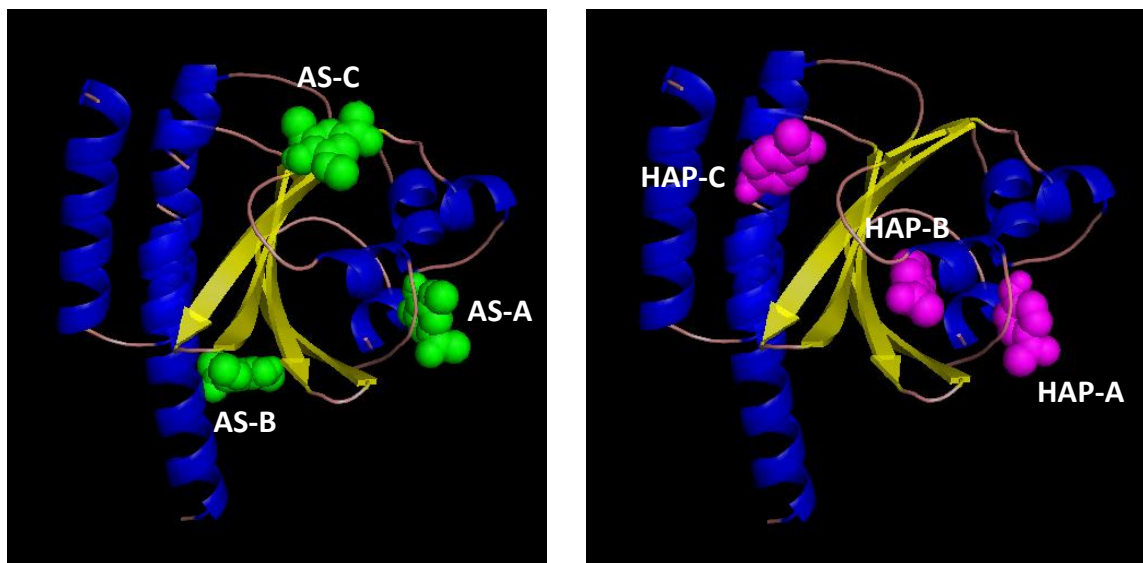
Initial computational analysis of the phenol binding of the *Agrobacterium* strain KU12 was also performed with its linker structure model generated by the SAM-T06 method, using amino acid sequence VirA<sub>KU12</sub>(281-456) as input. The GAF structure 3P01 was also identified from the automated method and used as the threading template, and the resulting structure model was used for phenol docking. As seen in Fig. 3.8B, even though the same template was used for threading linker<sub>A6</sub> and linker<sub>KU12</sub>, the resulting structure and folding is slightly different, possibly due to only 40% sequence identity between the two. Instead of having a long  $\alpha$ 1, linker<sub>KU12</sub> has the  $\beta$ -sheets starting more upstream and an additional helix predicted between  $\beta$ 2 and  $\beta$ 3. The docking of AS and HAP into the linker<sub>KU12</sub> structure also resulted in several stable and focused sites, as seen in Fig. 3.8B. Docking AS into linker<sub>KU12</sub> has the most sites located in the helix-loop region (AS-A), one site at the bottom of the  $\beta$ -sheets (AS-B), and one peripheral to  $\beta$ 3 (AS-C). Docking HAP into linker<sub>KU12</sub> has one conserved site as seen for AS (HAP-A), one into the GAF structure (HAP-B), and one between the  $\beta$ -sheets and the helices (HAP-C). The amino acids within 5 Å around each phenol in linker<sub>KU12</sub> are shown in Fig. 3.8C.

**Fig. 3.8 The docking of phenols to the linker structure generated by SAM-T06 method.** **A)** Three major conformations of docking AS, HAP, and DMP individually into the A6 linker structure generated by SAM-T06. The phenols are shown in CPK and colored in green for AS, magenta for HAP, and orange for DMP. **B)** Three major conformations of docking AS and HAP into the KU12 linker structure generated by SAM-T06. **C)** The amino acid sequence of the linker domain of strain A6 and KU12, shown with the SAM-T06 predicted structure. The previous threaded linker structure of A6 (Fig. 1.6B) is also shown for comparison. Each docking trial resulted in 10 stable conformations, numbering from 1 to 10 from the lowest to highest binding energy. The amino acids located 5 Å around each phenol conformation was colored corresponding to **A)**. The isoleucine-bound GAF structure 2B18 is shown above with amino acids 5 Å around the isoleucine ligand colored in cyan. The amino acids subjected to mutagenesis are shown in bold.

(A)



(B)



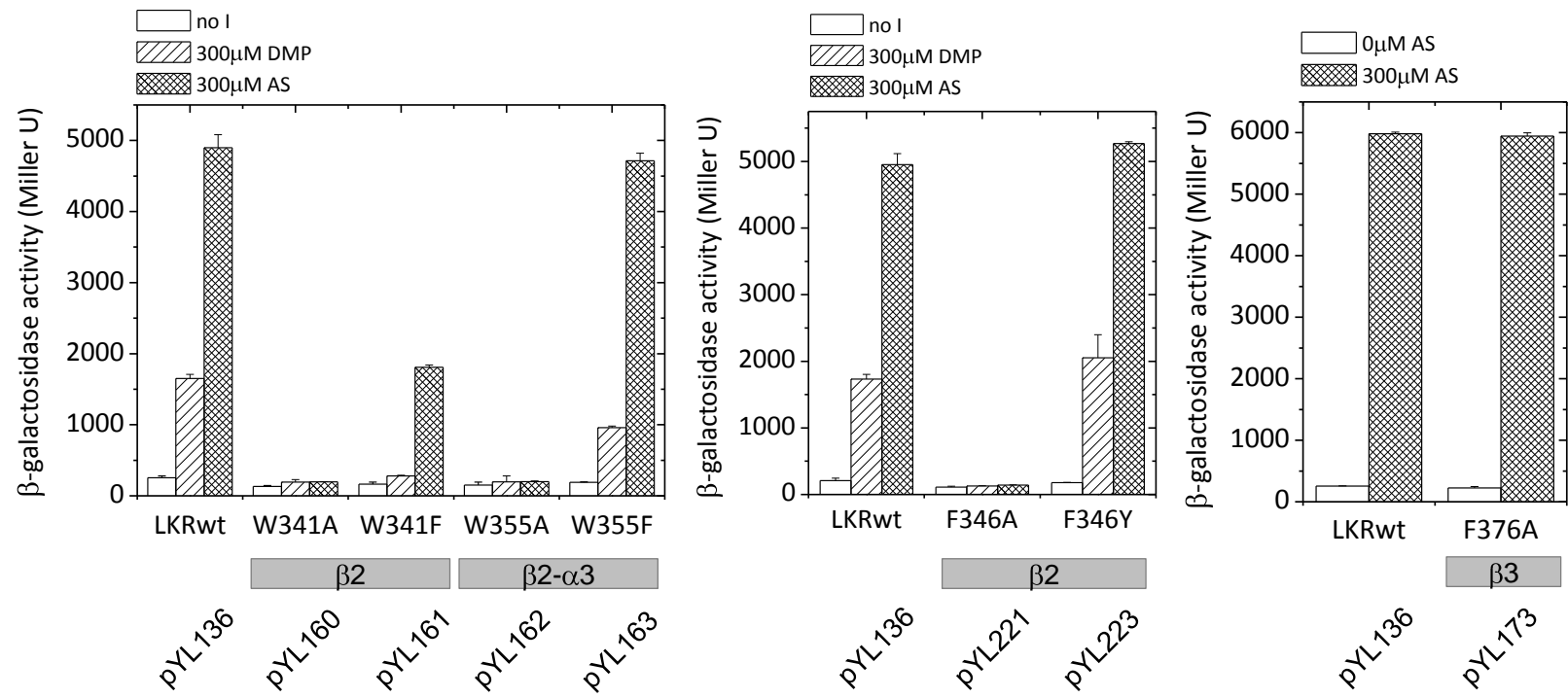


### Genetic analysis of phenol perception

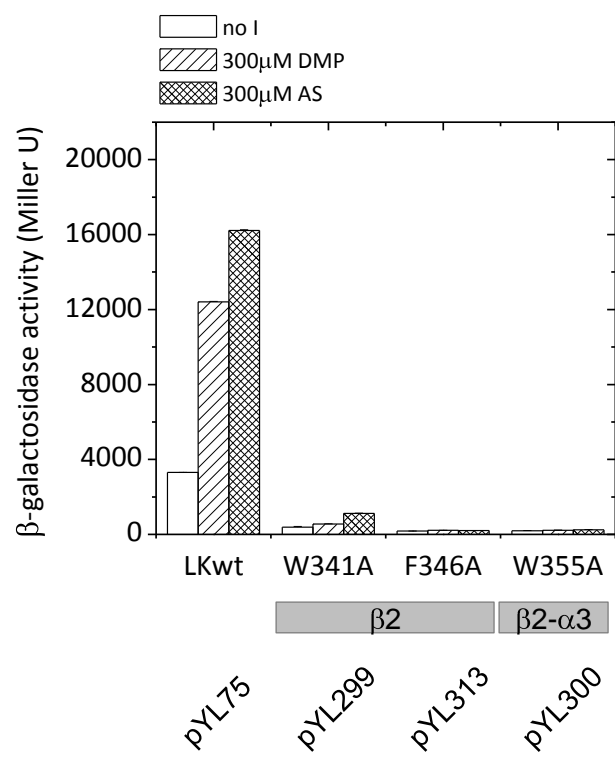
The computational approach by Autodock indicated several conserved phenol binding sites in the linker, many of which are consistent with the GAF prediction. The majority of the residues within 5 Å around the stable phenol sites are in the sequence region between  $\beta 2$  and  $\beta 4$ . This predicted phenol binding site (Fig. 3.4) involves a hydrophobic cleft that could stabilize phenol binding, and this hydrophobic core should be conserved among different linker structures. The sequence alignment of different *Agrobacterium* strains (Fig. 3.9A) indicates several conserved hydrophobic residues, and the aromatic ones were initially selected for analyzing the *in vivo* phenol inducibility in LKR by alanine scanning. W341, which is not fully conserved and is slightly outside the predicted phenol binding region, was also included in the scan. As seen in Fig. 3.9B, except for F376A, which has an unchanged phenol inducibility *in vivo*, W341A, F346A, and W355A all abolished phenol inducibility. Changing those residues to other aromatic ones restored activity. One possible reason that those residues showed a null phenotype is that the alanine mutation may be detrimental for protein folding. To test this idea, the mutations were moved to LK, which has a basal  $\beta$ -galactosidase activity in the un-induced state, which can serve as an indicator of proper protein folding. As seen in Fig. 3.9C, while W341A restored the phenol inducibility with decreased maximal activity, the activity of both F346A and W355A were barely detected *in vivo*, indicating the alanine substitution at F346A and W355A may have significantly altered the protein structure, and resulted into a non-functional conformation.





**(B)**

(C)



### Regulation of phenol specificity at the helix region

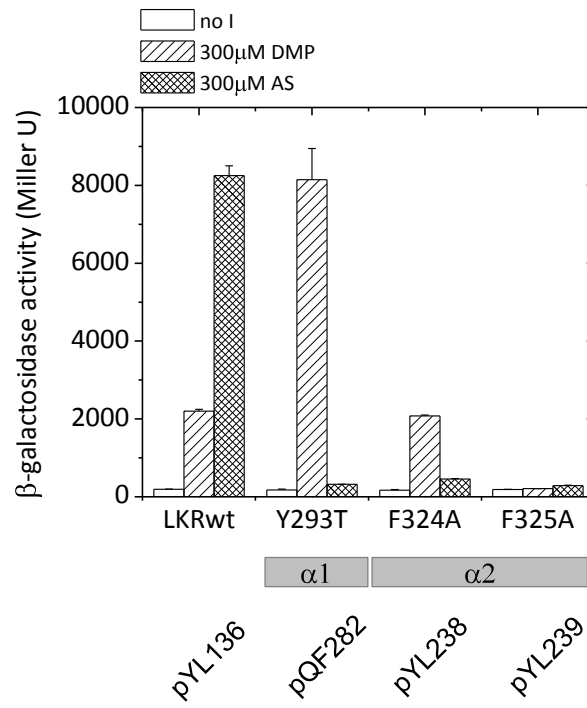
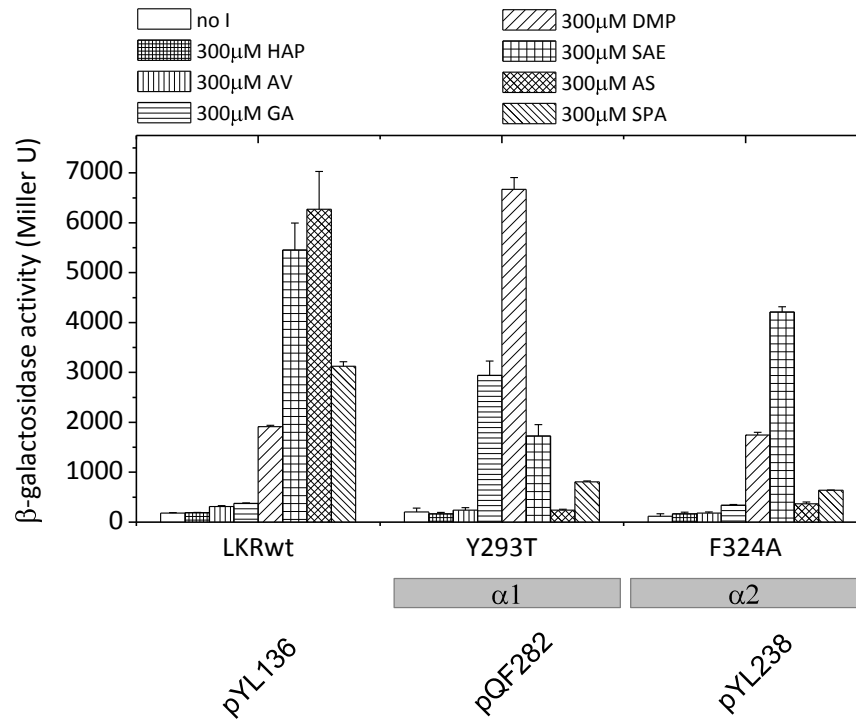
While the GAF prediction and computational modeling both identified the phenol binding site in the less-structured region, Y293T in the VirA truncation LKR(285-829) remarkably changes the phenol specificity, indicating that the phenol specificity can be regulated by the helical residues located on the other side of the  $\beta$ -sheets. While enzyme activity can be altered by distant mutations, for example in *Pseudomonas fluorescens*, the enantioselectivity of the esterase can be significantly enhanced by mutations 12-14 Å away from the active site (Horsman, Liu et al. 2003), the effect requires tight coupling through the structural scaffold. Distant mutations may modulate substrate specificity or enzyme activity in several ways, but most involve long-range protein conformational changes by distant hydrophobic forces, charge distributions, or hydrogen bond formation (Nechushtai, Lammert et al. 2011).

To further analyze the phenol specificity observed in Y293T, the linker structure model and the helix bundle conformation built in Chapter 2 now allow specific mutations to be evaluated. Previous saturation mutagenesis results at Y293 indicated that Y293T, Y293P, Y293S, and Y293A all had a similar specificity switch regarding *para*-substitutions, and Y293F, Y293L and Y293H had a similar phenol recognition profile (Fang 2007). Taken together, these results suggest a general hydrophobic environment around this residue may be critical for *para*-substituent perception. According to the structural model of the linker domain, there are two other hydrophobic residues located close to Y293, F324 and F325, which reside in  $\alpha$ 2, a helix positioned anti-parallel to the  $\alpha$ 1 helix. Each of these two phenylalanine residues were mutated to alanine in LKR and

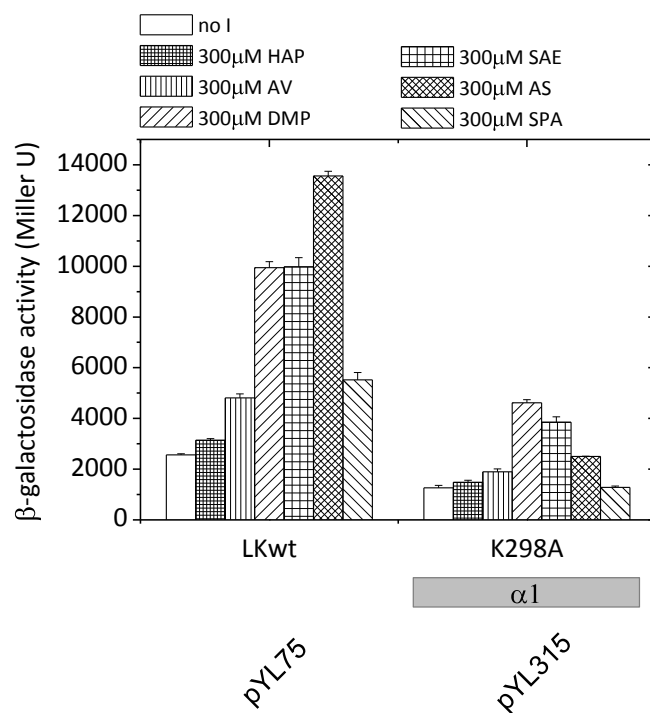
examined for phenol specificity (Fig. 3.10A). The overall phenol inducibility was significantly reduced in F325A, similar to the phenotype observed in W341A, F346A, and W355A, while for F324A, the phenol specificity towards AS and 2,6-DMP changed, similar to Y293T. A more detailed phenol recognition profile of F324A was analyzed with phenols containing no *o*-methoxy (HAP), one *o*-methoxy (AV), two *o*-methoxy groups but no *para*-substituent (DMP), 1-C *para*-substituent (SAE), 2-C *para*-substituent (AS), and 3-C charged *para*-substituent (SPA) (Table 3.3, Fig. 3.10B). The phenol specificity of F324A towards the *o*-methoxies did not seem to change, similar to Y293T, and while the most potent phenolic inducer for wild-type LKR is the two-carbon (2-C) AS, for Y293T it is the 0-C DMP, and for F324A it was 1-C SAE.

The regulation of phenol specificity was also observed in K298 and E430 while analyzing the potential salt-bridge formation in Chapter 2 (Fig. 2.12C). Removing the positive charge from K298 generally reduced phenol inducibility, but when constructed in LK, the phenol specificity was moved towards the 0-C DMP (Fig. 3.10C). Similar specificity changes were also observed when mutating E430 to the more hydrophobic tryptophan residue (Fig. 3.10D). Taken together, it seems the hydrophobic association between  $\alpha 1$  and  $\alpha 2$ , and the charge distribution between  $\alpha 1$  and  $\alpha 4$  both regulate *para*-substituent perception. Moreover, another surprising finding in specificity regulation was in L431R, which was originally designed to analyze whether it mimics the constitutively on phenotype of C435R. The constitutive phenotype was not observed in L431R, but instead it displayed a significant change in phenol specificity (Fig. 3.10D). Since this substitution is predicted to be at the  $\alpha 4$ - $\alpha 4'$  dimer interface (Fig. 2.8C), the addition of a positive charge at this position significantly reduced the inducibility of most of the

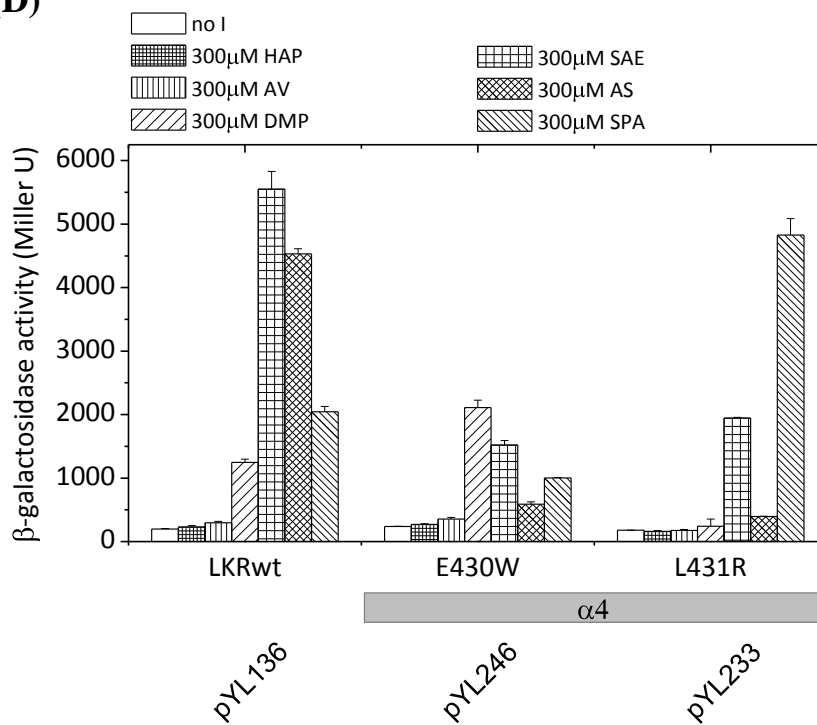
**Fig. 3.10 Engineering phenol specificity at the helices.** **A)** *A. tumefaciens* strain A136 carrying pRG109 and Y293T or the  $\alpha 2$  mutants in LKR were assayed for *vir* gene expression with 300  $\mu$ M 2,6-DMP or AS. **B)** LKR(Y293T) and LKR(F324A) were assayed for *vir* gene expression with 300  $\mu$ M indicated phenolic compounds. **C)** *A. tumefaciens* strain A136 carrying pRG109 and LK(K298A) was assayed for *vir* gene expression with 300  $\mu$ M indicated phenolic compounds. **D)** *A. tumefaciens* strain A136 carrying pRG109 and  $\alpha 4$  mutants in LKR were assayed for *vir* gene expression with 300  $\mu$ M indicated phenolic compounds. The secondary structure where each residue is located is indicated in the gray bars.

**(A)****(B)**

(C)



(D)





phenolics while increasing the perception of sinapinic acid (SPA), with its large and negatively charged side chain.

### ***In vitro* characterization of the GAF-like linker domain**

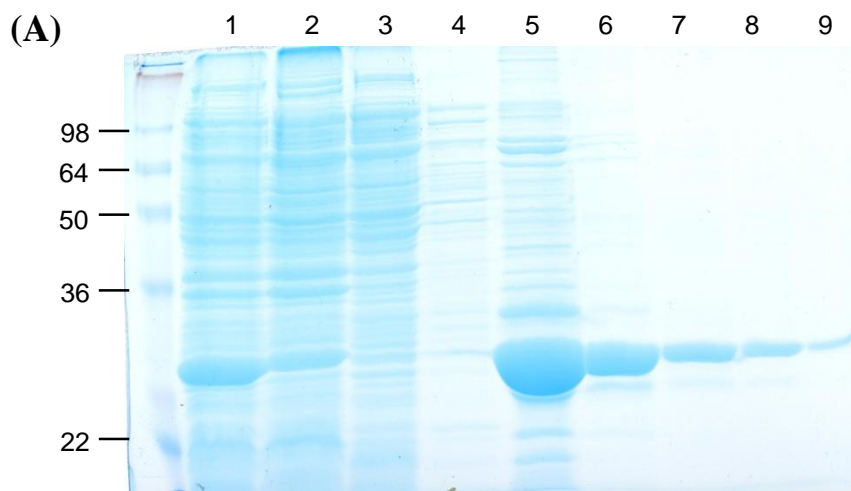
Previous analyses of phenol perception by the linker domain showed several properties, including the potential phenol binding site and specificity regulation, which could be predicted and examined using the GAF model. However, an understanding of the actual phenol binding event will require physical constraints. Therefore, the linker domain was constructed with an N-terminal His<sub>6</sub>-tag for purification and further characterization. As seen in Fig. 3.11A, the His<sub>6</sub>-tagged linker domain was successfully expressed and purified by conventional affinity chromatography with a yield of ~8 mg per liter of LB culture. Several mutants were successfully produced with the same procedure. To further characterize the structure of the linker, the percentage of secondary structure was calculated from the circular dichroism spectrum (Fig. 3.11B). The secondary structure components of the linker domain calculated by three different methods was similar to the GAF structure 1F5M, the template used to obtain the linker structure model. One of the aromatic mutants W341A was also characterized (Fig. 3.12), and although W341A had a lower solubility as compared to the wild-type linker, its secondary structure components were similar to the wild-type.

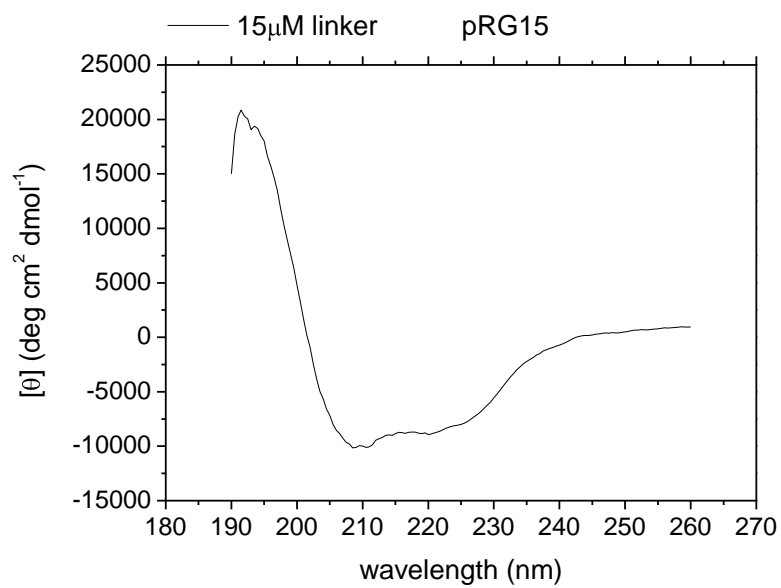
To further validate the linker structural model, intrinsic tryptophan fluorescence was used to analyze folding. The fluorescence of the tryptophan residues in the protein report on their environment, with a tryptophan in a more hydrophobic environment having a lower  $\lambda_{\max}$  due to better energy transfer, and a more surface exposed tryptophan

having a higher  $\lambda_{\max}$ . The excitation wavelength was set to 295 nm to avoid absorption and fluorescence from the tyrosines and phenylalanines. Two residues in the linker domain representing different environment, based on the structure model, were selected for fluorescence analysis (Fig. 3.13A). As seen in Fig. 3.13B, while the  $\lambda_{\max}$  of the wild-type linker fluorescence is 350 nm, the W341A and the Y414W mutants both showed a blue-shift, to 340 nm and 346 nm, respectively, indicating that W341 is located at a more exposed surface, while Y414 is more buried, both consistent with the structure model.

Besides the spectroscopic characterization, previous results indicated a potential interaction with the receiver domain (Fang, 2007), chemical cross-linking methods were therefore used to examine interactions with the receiver domain. The receiver (aa712-829) was constructed with an N-terminal His<sub>6</sub>-tag and was successfully expressed and purified. As seen in Fig. 3.14, using the amine specific homo-bifunctional imidoester cross-linker DMS (dimethyl suberimidate), the linker domain (aa285-471) was detected to be covalently linked as a homodimer (lane 7 and 8) as well as forming the heterodimer with the receiver domain (lane 5 and 6). A truncated linker domain with a shortened coiled-coil (aa285-444) was also constructed and analyzed by chemical cross-linking. However, the shortened linker (aa285-444) does not form a homodimer under these conditions with DMS (lane 3 and 4), nor was it detected to form a covalent heterodimer with the receiver domain (lane 1 and 2). The R domain by itself did form homodimers and higher oligomers (lane 9 and 10).

**Fig. 3.11 *In vitro* characterization of the linker domain.** **A)** Purification of the linker domain. N-terminal His<sub>6</sub>-tagged VirA(285-471) was over-expressed by 0.2 mM IPTG induction for 18 hours at 16 °C. The protein was purified with Ni-NTA affinity resin following the protocol in Materials and Methods. Lane 1: whole cell; lane 2: soluble lysate; lane 3: flow-through of Ni-NTA; lane 4: wash-off; lane 5-9: eluted fractions. **B)** Circular dichroism spectrum of 15 μM purified linker and the calculated percentage of secondary structure. The percentage of secondary structure of the linker was calculated by three programs: CONTILL, CDSSTR, and SELCON3 (Yang, Wu et al. 1986; Johnson 1999). The secondary structure of the GAF domain 1F5M is also listed.

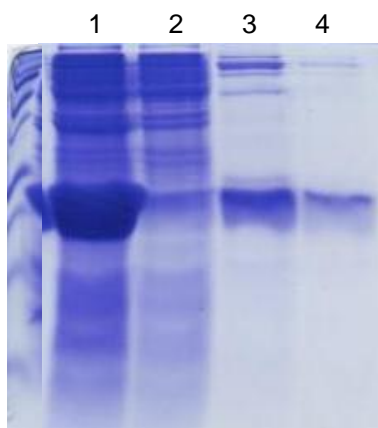


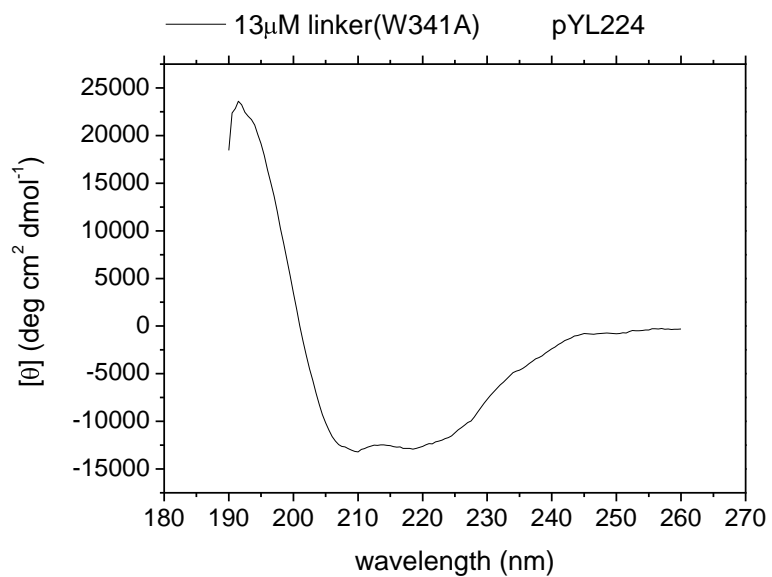
**(B)**

% Secondary structure of linker				
	Helices	Sheets	Turn	Unordered
<b>CONTILL</b>	33.0	21.3	18.4	27.2
<b>CDSSTR</b>	34.0	19.5	19.0	27.2
<b>SELCON3</b>	31.9	15.9	21.7	28.8
<b>1F5M</b>	37	23	--	--

**Fig. 3.12 Linker(W341A).** **A)** His<sub>6</sub>-VirA(280-455)(W341A) was over-expressed by 0.05 mM IPTG induction for 18 hours at 20 °C. The protein was purified with Ni-NTA affinity resin same to the wild-type linker. Lane 1: whole cell; lane 2: soluble lysate; lane 3-4: eluted fractions. **B)** Circular dichroism spectrum of 13 μM linker(W341A) and the calculated percentage of secondary structure. The percentage of secondary structure of linker(W341A) was calculated by the same programs used by the wild-type linker and compared to the average percentage of secondary structure of the wild-type linker.

**(A)**

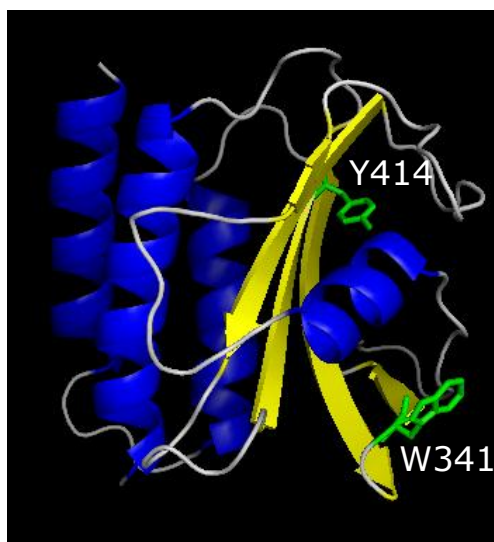


**(B)**

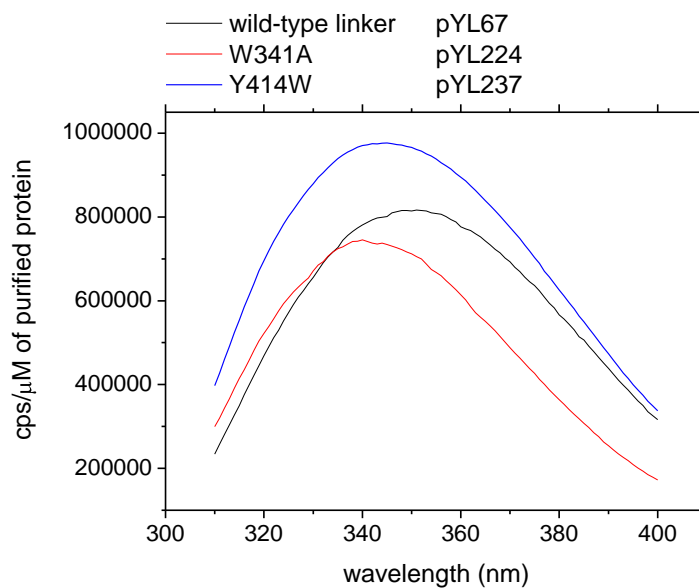
% Secondary structure of Linker(W341A)				
	Helices	Sheets	Turn	Unordered
<b>CONTILL</b>	38.4	15.9	20.2	25.6
<b>CDSSTR</b>	41.4	14.1	17.3	26.9
<b>SELCON3</b>	38.4	13.4	19.9	27.3
<b>Wild-type L</b>	33.0	18.9	19.7	27.7

**Fig. 3.13 Intrinsic tryptophan fluorescence of the wild-type and mutant linkers. A)** W341 and Y414, shown in ball-and-stick, in the linker structure model. **B)** Tryptophan fluorescence of the wild-type linker (black), linker(W341A) (red), and linker(Y414W) (blue). The intensity of each spectrum was normalized to their protein concentration.

(A)



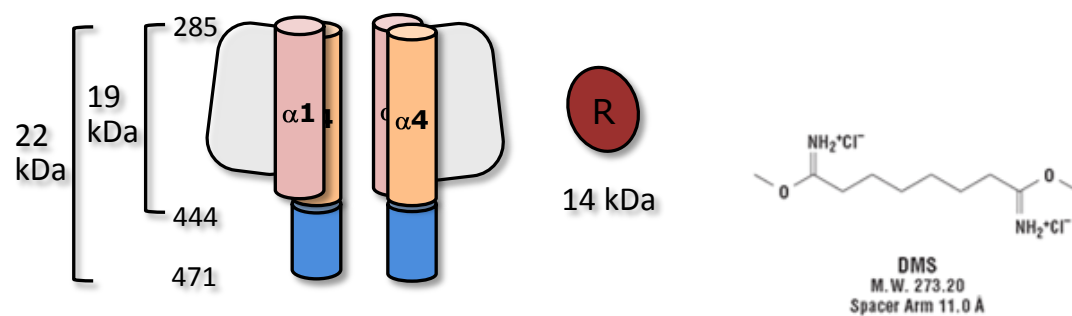
(B)



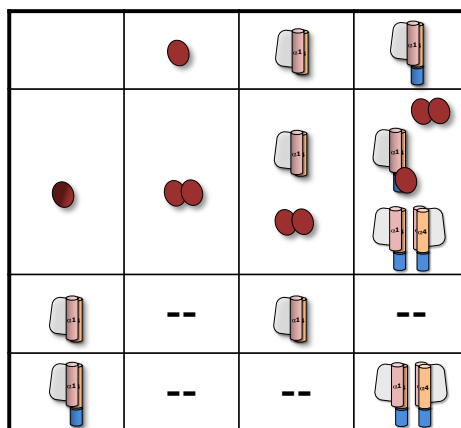
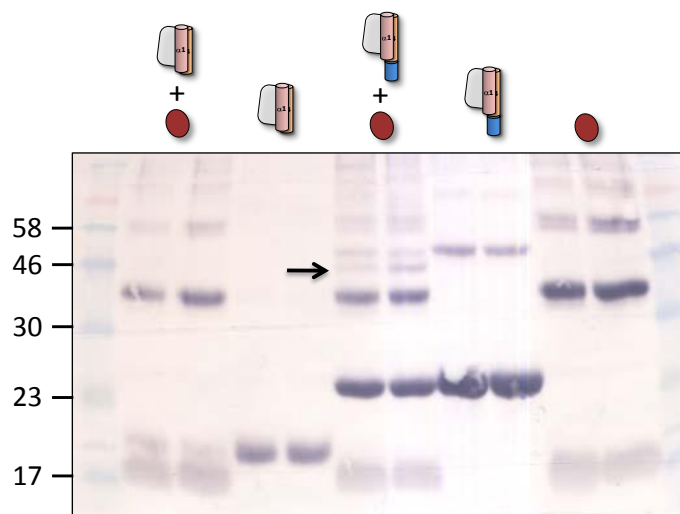
**Fig. 3.14 Chemical cross-link of the VirA linker and receiver domains.** **A)** Schematic representation of the two linker constructs (aa285-444; aa285-471), the receiver (aa712-829), and their corresponding molecular weight. The chemical structure of the cross-linker DMS is also shown. **B)** Western blotting by anti-His<sub>6</sub> antibody to probe both the linker and the receiver was used to detect the results of chemical cross-linking. Two molar ratios of protein:DMS (1:20, 1:50) were used, and shown in parallel. Lane 1, 2: linker(285-444)+R; lane 3, 4: linker(285-444) alone; lane 5, 6: linker(285-471)+R; lane 7, 8: linker(285-471) alone; lane 9, 10: R alone. The cross-linked heterodimer of linker(285-471) and receiver is marked with an arrow. The schematic representation of the cross-linking result is shown in the table.



(A)



(B)



## Discussion

In this chapter, I initiated the characterization of the GAF fold and phenol perception by the linker domain. With the aid of computational modeling, the phenolic compounds were initially docked into the linker structure model and evaluated. The results supported the GAF fold and predicted potential residues involved in signal perception and transmission. However, computational docking relies largely on a constructed structure model and the free energy calculation at each docking grid point involves a flexible ligand and a rigid receptor (protein) conformation. Any deviation of the receptor conformation in the micro-environment can result in large free energy variations. Moreover, the structure model we constructed (Fig. 1.6B) was based on a manual alignment of predicted secondary structure of the linker with other known GAF domains (Fig. 1.6A) using the Swiss-Model Workspace to fill-in the gaps. This structure model may well represent the global folding and predict the residue distribution of the linker domain, but it did not seem to be a good receptor for docking as the stable phenolic conformations probed by Autodock were spread all over the linker structure.

The linker structure generated by the automated SAM-T06 method appears to be more suitable for docking. The stable phenol conformations generally have lower free energy and were focused at several sites. Additional analysis of the possible phenol binding site by the computational approach was carried out using the linker structure of strain KU12, constructed by SAM-T06, and docked with the phenols. Even though the sequence identity of the linker domain from strain A6 and KU12 is only 40%, the SAM-T06 predicts only slight overall differences, and the docking of phenols identified conserved regions. As seen in Fig. 3.8C, the majority of the linker<sub>A6</sub> residues around the

probed phenol binding sites reside between  $\alpha 3$  and  $\beta 4$ , which was conserved with the docking results of the linker<sub>KU12</sub> structure. Another binding region also indicated by computational docking is located between  $\beta 2$  and  $\alpha 3$ , and both regions were further probed via genetic analyses. Improvements in the computational docking may be achieved using a modified structure model with more constraints, or by applying a more sophisticated docking method that can assign some residues of the receptor to be more flexible. “Blind docking” requires a larger spacing of grid points in order to cover the entire linker surface, so if the potential binding site can be further narrowed down to a certain region, the grid point spacing can be lowered to obtain a more accurate free energy calculations for the conformational search.

I selected the aromatic residues between  $\beta 2$  and  $\beta 4$  for the initial genetic analysis of phenol binding because they potentially constitute the hydrophobic core for phenol binding (Fig. 3.4) and are conserved among ten *Agrobacterium* strains (Fig. 3.9A). However, except for F376, whose hydrophobic side chain does not seem to be required for phenol perception, and W341, which is located outside the predicted binding region and was still inducible when changed to alanine in LK, alanine substitution seems to be detrimental for the overall activity of F346A and W355A, and gave a null phenotype in the LK constructs. Whether the conservation of those residues is for proper protein folding or direct phenol binding can be further characterized by selecting different mutations, for example leucine. Other conserved hydrophobic residues in the region between  $\beta 2$  and  $\beta 4$  can also be analyzed, and future mutagenesis can also be specifically designed based on phenol specificity differences and amino acid sequence alignments between strain A6 and KU12.

Additional insights into phenol specificity were provided by the discovery of Y293T. Extending the random mutagenesis library of aa285-294 (Fig. 3.5)(Fang 2007), the mutations identified in this chapter extended *para*-substituent perception across the helix bundle, formed by  $\alpha 1$ ,  $\alpha 2$ , and  $\alpha 4$  (Fig. 3.10). The phenol specificity of F324A shifted from the 2-C AS to the 1-C SAE, moving in the same direction as of Y293T (from the 2-C AS to the 0-C DMP), suggesting a hydrophobic surface between  $\alpha 1$  and  $\alpha 2$  regulates the perception transmission of different sizes of *para*-substitutions. Similar specificity changes were also observed by modulating the potential salt-bridge formation between  $\alpha 1$ - $\alpha 4$ , by K298 and E430, as seen by the specificity change of K298A and E430W from the 2-C AS to the 0-C DMP. The *para*-substituent perception specificity was shifted in the other direction in L431R, whose hydrophobic-to-positive-charge change at the  $\alpha 4$ - $\alpha 4'$  interface led to the perception of phenols with larger, negatively charged side chains. More phenolic compounds with different *para*-substitutions could be used to further confirm these observations, but it seems the global conformation of the six-helix bundle of the linker domain predominantly regulates the *para*-substitution perception.

Mutations located distal to the enzyme active site can modify the enzyme activity remotely (Liebeton, Zonta et al. 2000; Horsman, Liu et al. 2003; Jackson, Baldwin et al. 2004; Nechushtai, Lammert et al. 2011). Some mutations induce specific secondary structure formation, some merely render a different protein conformation, and the conformational change is transmitted to the substrate binding site to alter the enzyme activity. According to the phenol binding model of the linker domain proposed in Fig. 3.4, the perception of the *para*-substitution is the most flexible, but the distant regulation of

the mutations at the helix bundle, and the conformational change transmitted across the  $\beta$ -sheets that mediate the *para*-substitution perception raises very interesting questions about mechanism.

Both the computational and genetic analyses still can not exclude the possibility that the phenol binding site is not following the GAF prediction. Therefore, a more direct method is critically needed to characterize the phenol binding site in the linker. Using biophysical methods to analyze the purified protein binding with the ligand *in vitro* might clearly reveal the ligand binding site. Preliminary analyses of the purified linker domain included some structural and functional characterization. The secondary structure of the linker domain was detected by circular dichroism, which is consistent with its GAF fold. The spectroscopic analysis using intrinsic tryptophan fluorescence as a reporter also showed the solvent environment of W341 and Y414 to be consistent with the GAF structure model. A crude functional test of the linker domain, based on its possible interaction with the receiver domain, was also positive. The covalently cross-linked product was observed from the receiver and the linker with the extended coiled-coil (aa285-471), which suggests the coiled-coil served as the linker-receiver interface. Whether the shorter linker(285-444) forms homodimers, and whether it interacts with the receiver domain could not be detected by DMS and needs further analysis. Nevertheless, the preliminary results of the purified linker(285-471) support the GAF fold.

However, the experimental limitation of the purified linker domain and its precipitation under lab conditions continue to hamper progresses. The buffer condition storing the purified linker have been modified, including adding glycerol, adjusting the pH, changing the salt type, or adding mild detergent, to help stabilize the purified linker

domain, but the improvements were not sufficient to completely prevent precipitation. This instability hampers further biophysical characterization that requires long term protein stability, including NMR and crystallography. In that case, an experimental approach proposed to identify the potential phenol binding site is to use affinity labeling.  $\alpha$ -bromoacetosyringone (ASBr), an inhibitor of the VirA/VirG TCS, has been shown to successfully form covalent linkage and labeled proteins *in vivo* (Lee, Dudley et al. 1992). Even though this *in vivo* experiment did not successfully label VirA as predicted, presumably due to the complex environment inside the cell, the structure similarity between ASBr and AS still suggests the feasibility of probing the phenol binding site with purified protein *in vitro*. Moreover, in ASBr labeling the generic AS molecule can be added as a competitive inhibitor to protect the potential binding site, and further protease treatment and MS analysis can be coupled to determine the phenol binding site.

**CHAPTER 4**

**INTEGRATING PHENOL AND SUGER PERCEPTION FOR THE BROAD-  
HOST-RANGE IN *AGROBACTERIUM TUMEFACIENS***

**Introduction**

The discovery of inter-kingdom gene transfer and the implications the process offered for agricultural biotechnology motivated great interest in the molecular mechanisms of *Agrobacterium* pathogenesis (Gelvin 2005). The extremely broad host range of this pathogen allows DNA transfer to virtually all flowering plants (DeCleene and DeLey 1976). A mechanistic understanding of how this host range can be achieved and maintained has significant evolutionary implications. Furthermore, from a biotechnological perspective, an understanding of host detection and responses opens many practical opportunities for system-level control.

**Host-range determination of *Agrobacterium***

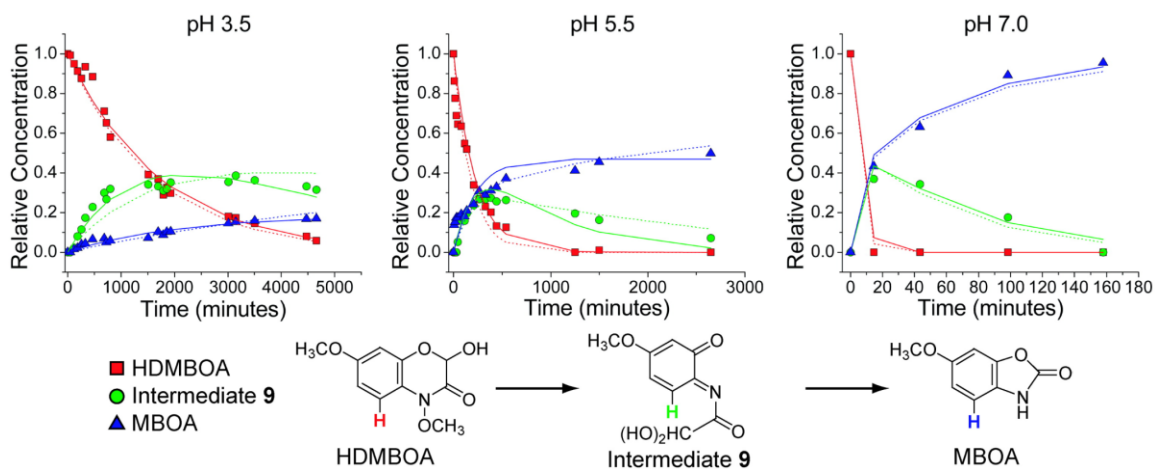
The only plant family generally not pathogenized by *Agrobacterium* is the Graminae. This group of monocotyledonous grasses includes the economically important cereal crops, most notably rice, wheat, and maize. However, laboratory procedures have now been established to genetically transform those plants with *Agrobacterium*. Such methods, called “Agroinfection,” can successfully mediate the transfer of genetic material without inducing tumor formation (Grimsley, Hohn et al. 1986; Grimsley, Hohn et al. 1987). Further, enhancement of agroinfection of Graminae with a higher efficiency was

achieved by exogenously adding acetosyringone, the *vir* gene inducer, to the inoculation medium (Hiei, Ohta et al. 1994; Ishida, Saito et al. 1996). Different strains of *Agrobacterium* seem to have different agroinfection competence. For example, the *Agrobacterium* strain C58 is agroinfection competent, while the strain A348, which has the C58 chromosomal background but the A6 Ti plasmid, is not. Apparently, the agroinfection competence is dictated by the Ti plasmid, and further gene insertion analysis determined the *virA* on the Ti plasmid to be the most critical element (Raineri, Boulton et al. 1993). Further domain swapping within *virA* assigned the linker region as the functional site dictating agroinfection competence (Heath, Boulton et al. 1997).

These results suggested that for Graminae to be resistant to the natural *Agrobacterium* pathogenesis, either some critical signal was lacking in the Graminae or alternatively some component from the grasses inhibited activation. Indeed, the organic root exudates of young maize seedlings was shown to contain 98% of a single compound benzoxazinone (Zhang, Boone et al. 2000), 2-hydroxy-4,7-dimethoxybenzoxazin-3-one (HDMBOA)(Fig. 4.1), and this compound specially inhibited the initial signal perception step in *Agrobacterium* pathogenesis. Further analysis of this labile benzoxazinone indicated the actual inhibitor to be the decomposition intermediate *o*-imidoquinone (Fig. 4.1)(Maresh, Zhang et al. 2006). The innate defensive mechanism of maize then involves the release of the benzoxazinone, structure common in the Graminae, which creates a zone of protection in the acidic rhizosphere, but once entering the neutral cytoplasm of the bacteria, decomposes to the *o*-imidoquinone to inhibit *vir* gene expression. It is unclear where in the signaling pathway HDMBOA, or the *o*-imidoquinone, targets,



**Fig. 4.1 HDMBOA decomposition.** HDMBOA (7.5 mM) was dissolved in 100 mM phosphate buffer and the  $^1\text{H}$  NMR lines 7.16 ppm (HDMBOA, red square), 6.40 ppm (*o*-imidoquinone intermediate, green circle), and 6.97 ppm (MBOA, blue triangle) were monitored (Maresh, Zhang et al. 2006). The relative intensity of these signals was used to assign relative concentrations.



and further characterization of whether the benzoxazinone is a general inhibitor to *Agrobacterium* pathogenesis is also needed.

The natural selection of dicotyledonous plant host also appears to vary among different *Agrobacterium* strains. The host range of *Agrobacterium* strain Ag162 is limited to grapevines, whereas strain A6 infects a wider range of dicotyledonous hosts. This limited vs. wide-host-range was also attributed to *virA* (Leroux, Yanofsky et al. 1987; Turk, Nester et al. 1993), and the presence of a *vir*-box upstream of the promoter region, which can be up-regulated upon signal sensing, was indicated to be in the wide-host-range strains. An alternative mechanism for host range determination, as described in Chapter 3, is to perceive structurally different signaling molecules. As mentioned, the strain KU12 appears to perceive a different group of phenolic compounds from strain A6 or C58 (Lee, Jin et al. 1995; Lee, Jin et al. 1996), which may constitute its different host range. Therefore, host range determination appears to be built upon the genetic control, the molecular detail of the initial signal perception, and the signaling landscape of the host plant.

### **The two library screening results**

The VirA receptor's central role in host commitment is suggested by the complexity of the integral membrane protein's architecture. Even though the linker domain seems to be the obvious receptor of the phenolic compounds (Fig. 1.5B)(Chang and Winans 1992; Gao and Lynn 2005), the addition of monosaccharides, a signal perceived at the periplasmic domain of VirA, enhances the phenolic response and broadens the phenol recognition profile significantly (Peng, Lee et al. 1998). In addition

to the signal sensing periplasmic and linker domains, a seemingly redundant cytoplasmic receiver (R) domain also cooperatively regulates kinase activity. This R domain, located at the C-terminus of VirA and homologous to the receiver domain of VirG as well as other general response regulators, can be a repressor of kinase activity even when expressed *in trans* (Chang, Zhu et al. 1996; Gao and Lynn 2005). The R domain's function does not seem to be tied with phosphorylation, as the same activity was detected when the conserved receiver aspartate was mutated to asparagine (D766N)(Chang, Zhu et al. 1996). Efforts to construct VirA heterodimers suggested that the R domain functions across to the opposite subunit (Brencic, Xia et al. 2004). However, a recent study observed that when VirG was expressed at lower concentrations, the R domain served as an activator, possibly acted as a guiding sequence to direct VirG association (Wise, Fang et al. 2010). Further evidence of the R domain may function in integrating with the phenolic perception was found in the R domain mutant I734N, which displays a broader phenol recognition profile (Peng, Lee et al. 1998), and implies the R domain in signaling specificity.

To address the role of the R domain in phenol perception regulation, based on the understanding that the linker domain constitutes phenol binding site and determines the specificity (Chapter 3), we hypothesized that the R domain directly interacts with the linker domain to coordinate its activity. To test this idea, we constructed a sequential library screening (Fig. 4.2A)(Fang 2007). The first library was generated by randomly mutagenizing the R domain (aa712-829) and searched for mutations abolishing phenol perception. Three null mutations, V765D, D766N, and L770P, were identified, and each of them was used as the template to generate the second library, by randomly

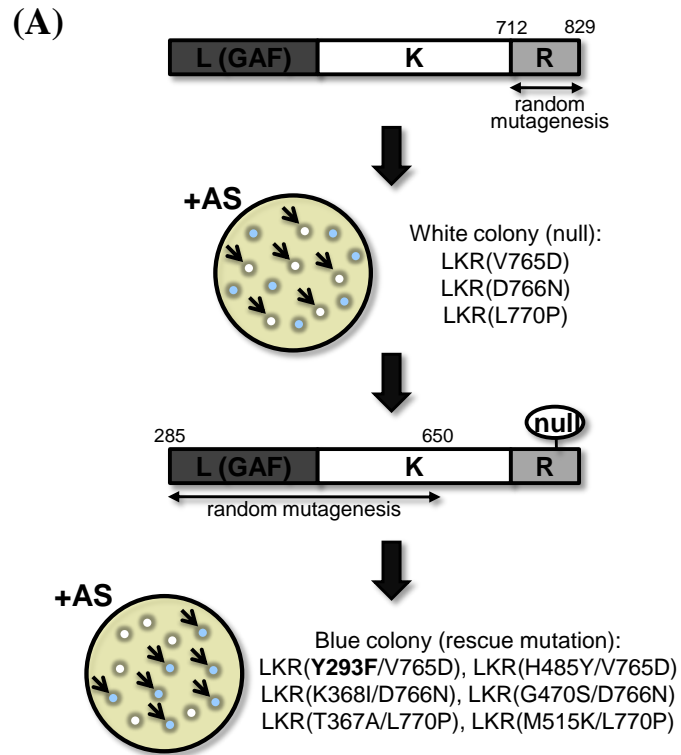
mutagenizing the linker and part of the kinase domain and searching for mutations that recover the phenol inducibility at high VirG concentrations. Among the rescued mutants identified, Y293F displayed the most significant recovering activity (Fig. 4.2B). This mutation not only recovers the null phenotype of V765D, but also significantly increased the phenol sensitivity (Fang 2007).

In Chapter 3, I have discussed the discovery of Y293T from another library, searching for phenol specificity changed mutants, and the global conformation of the helix bundle modulating *para*-substituent perception was discussed. The discovery of Y293F further indicated this position being important in L-R domain coordination. This residue directly follows the second transmembrane region (aa260-280), and the structure model of the linker domain positions it at the N-terminus of the GAF domain. Therefore, I hypothesized that this residue sits at a critical integration point where all the signal perception and domain coordination merged. To test this hypothesis, I incorporated Y293F and Y293T into full-length VirA, respectively, and analyzed their function in signal perception and integration in this chapter, and discuss the implication in host-range determination.

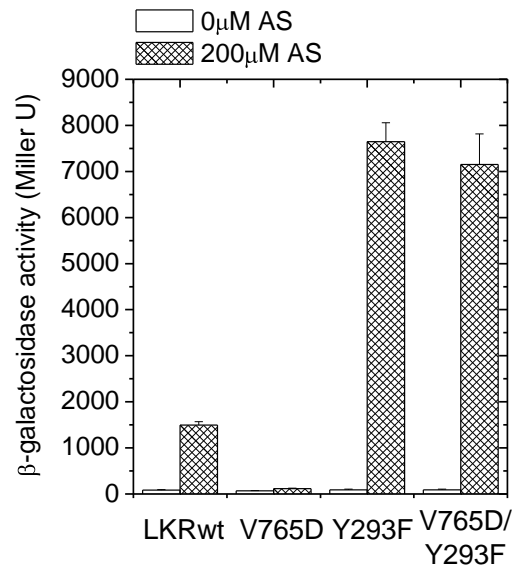
## Materials and Methods

**Bacterial strains, plasmids, and reagents.** The bacterial strains and plasmids used in this chapter are listed in Table 4.1. *E. coli* strain XL1-Blue (Stratagene) was used for routine plasmid construction. The phenolic compounds used for *vir* gene induction were listed in Table 3.3. The cloning reagents were purchased from New England Biolab and Promega.

**Fig. 4.2 The library screening probing linker-receiver interaction.** **A)** The first random mutagenesis library was generated within the R domain (aa712-829) and screened for null mutations (white colonies). The resulted three mutations, V765D, D766N, and L770P, was individually incorporated into LKR and raised for the second library, randomly mutagenized the linker and part of the kinase domain (aa285-650). Six compensating mutations were identified, with three in the linker domain and three in the kinase. **B)** *A. tumefaciens* strain A136 carrying pRG109 and the indicated LKR constructs of the null-mutation V765D, the linker mutation Y293F, and the double mutation LKR(Y293F/V765D) were assayed for the  $\beta$ -galactosidase activity in the presence or absence of 200  $\mu$ M AS.



(B)



**Table 4.1 Bacterial strains and plasmids used in Chapter 4**

Strains/plasmids	Relevant characteristics	Reference
<i>E. coli</i> strains		
XL1-Blue	<i>recA1 endA1 gyrA96 thi1 hsdR17 supE44 relA1 lac[F' proAB lacI<sup>q</sup>Z M15 Tn10 (Tc<sup>r</sup>)]</i>	Stratagene
<i>A. tumefaciens</i> strains		
A136	Strain C58 cured of pTi plasmid	(Watson, Currier et al. 1975)
Plasmids		
pYW15b	Broad-host-range expression vector, IncW, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pYW48	<i>P</i> <sub>virA</sub> - <i>virA(aa1-829)</i> in pYW15b, Ap <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pVRA8	<i>virA</i> , IncW, pBR322ori, Ap <sup>r</sup>	(Lee, Dudley et al. 1992)
pJZ4	<i>P</i> <sub>virB</sub> - <i>lacZ</i> in pMON596, IncP, Spec <sup>r</sup>	(Zhang, Boone et al. 2000)
pJZ6	IncW/ColE expression vector with <i>P</i> <sub>N25</sub> , Ap <sup>r</sup>	(Zhang, unpublished)
pSW209Ω	<i>virB::lacZ</i> , IncP, Spec <sup>r</sup>	(Wang, Mukhopadhyay et al. 2000)
pRG109	<i>P</i> <sub>N25</sub> -His <sub>6</sub> - <i>virG</i> in pJZ4, Spec <sup>r</sup>	(Gao and Lynn 2005)
pQF282	<i>P</i> <sub>N25</sub> - <i>virA(aa285-829)(Y293T)</i> in pJZ6, Ap <sup>r</sup>	(Fang 2007)
pQF431	<i>virA(Y293F)</i> , IncW, pBR322ori, Ap <sup>r</sup>	This chapter
pYL136	<i>P</i> <sub>N25</sub> - <i>virA(aa285-829)</i> in pJZ6, Ap <sup>r</sup>	Chapter 2
pYL155	<i>virA(Y293T)</i> , IncW, pBR322ori, Ap <sup>r</sup>	This chapter

**Cloning, Transformation, and growth conditions.** Same as in Chapter 2.

**Plasmid constructions.** The plasmids used in this chapter are listed in Table 4.1, and the primers are listed in Table 4.2. The full-length *virA* with Y293F or Y293T mutants were generated by the two-step PCR from pVRA8, which carries  $P_{virA}::virA$ , with *virAKpnI*, LKRA1, and the corresponding complementary primers. The ~3.9 kb amplified product was digested by *KpnI*. The ~4.6 kb *KpnI* fragment was released from pVRA8, and the vector was successively treated with shrimp alkaline phosphatase for 1 hour prior to the ligation with the *KpnI* digested PCR fragment. The resulted constructs were pQF431 and pYL155, which contain *virA*(Y293F) and *virA*(Y293T), respectively, driven by the wild-type  $P_{virA}$ .

LKR(Y293F) was generated by the two-step PCR as described in Chapter 2. The DNA was amplified from pYW48 with LKR285, LKRA1, and the complementary primers YL447 and YL448. After *BamHI* and *Acc65I* digestion, the fragment was introduced into pJZ6 to generate pYL242. The nucleotide sequence of all the plasmid constructs were confirmed by DNA sequencing facility performed by Agencourt Genomic Service (now Beckman Coulter Genomics).

**Table 4.2 Primers used in Chapter 4**

Primer	Characteristic	Sequence (5' → 3') ( <u>RE site</u> ; <i>mutation</i> )
LKR285	BamHI_285_for	CG <u>GGA TCC</u> GAT TGG TTA GCG CGG CGT
LKRA1	KpnI_stop_829_rev	GC <u>GGT ACC</u> GCA ACT CTA CGT CTT GAT
<i>virAKpnI</i>	KpnI_virA_start	GC <u>GGT ACC</u> AGC TTG GCC TAC AAG TGC
YL217	Y293T_for	CGG CGT TTA GAT <u>ACC</u> GAA GAG CTA ATC AAA GAG ATC GGA



YL218	Y293T_rev	TCC GAT CTC TTT GAT TAG CTC TTC GGT ATC TAA ACG CCG
YL447	Y293F_for	GCG CGG CGT TTA GAT TTC GAA GAG CTA ATC AAA GAG
YL448	Y293F_rev	CTC TTT GAT TAG CTC TTC GAA ATC TAA ACG CCG CGC

---

**$\beta$ -galactosidase assays for *vir* gene induction.** All of the VirA variants were transformed into *A. tumefaciens* strain A136 containing pRG109, which carries  $P_{virB}$ -*lacZ* and  $P_{N25}$ -*virG*, for *vir* gene expression. The *A. tumefaciens* strains were grown and pelleted with the same procedure as described in Chapter 2, and diluted to OD<sub>600</sub> ~ 0.1 into tubes containing a total of 1 mL induction medium with the phenolic compounds at the indicated concentration. The *A. tumefaciens* strains carrying full-length *virA* (pQF431 or pYL155) were supplemented with 14 mM glucose in addition to the 1% glycerol, as indicated, and cultured at 28 °C, 225 rpm for 15 hrs. The  $\beta$ -galactosidase activity was determined by the same method as described in Chapter 2, from the optical densities at 600 nm and 415 nm.

## Results

### VirA(Y293F)

Prior to the characterization of the activity of Y293F in full-length VirA, its phenotypic change, including phenol sensitivity and specificity, was firstly analyzed in LKR, in which the sugar effect was removed. As shown in Fig. 4.3, when compared with the wild-type LKR (green), a significant lower concentration was required for LKR(Y293F) induction (blue), as the ED<sub>50</sub> enhanced from >500  $\mu$ M to 13  $\mu$ M. While compared with the full-length VirA, which can now take the sugar effect into

consideration, the ED<sub>50</sub> of LKR(Y293F) was close to the sugar un-induced PLKR (red), while the glucose induced PLKR has an further enhanced ED<sub>50</sub> ~2 μM (black). However, in all of the activity examined, LKR(Y293F) had the highest phenol induced maximal activity.

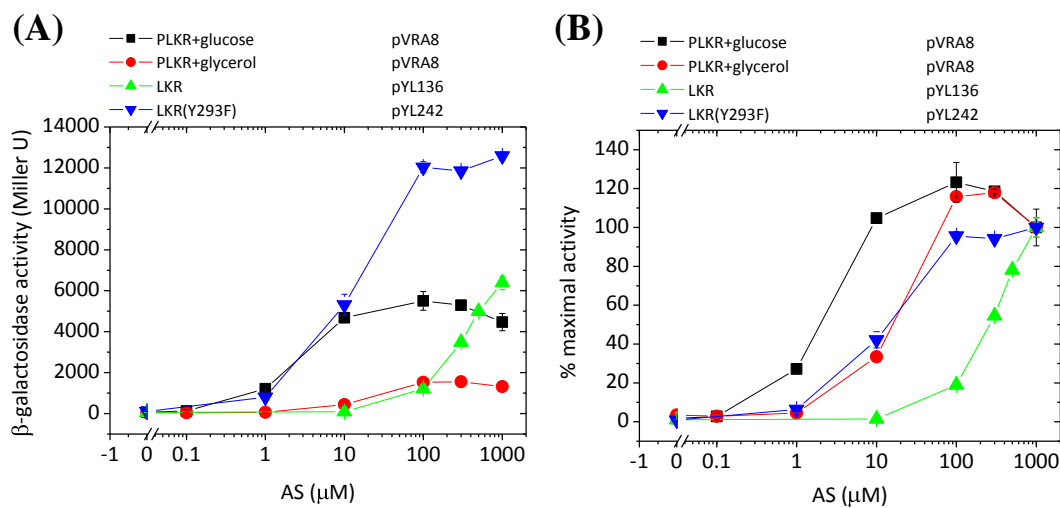
Whether changing the side chain from Tyr to Phe at aa293 alters phenol sensing specificity, as observed in Y293T in Chapter 3, was further analyzed (Fig. 4.4). The enhance of phenol response in LKR(Y293F) required only 50 μM phenolics for induction, while the wild-type LKR needed 200 μM of phenols. Moreover, LKR(Y293F) displayed a significantly broadened response to the phenols tested. AS remained the most potent inducer, but simpler phenols, including acetovanillone (AV), which lacks one *o*-methoxy group, and guaiacol (GA), which lacks both the *o*-methoxy and the *para*-acetyl group, showed significant activity.

The above phenotypes of LKR(Y293F), including an increase of phenol sensitivity and a broadened phenol perception profile, are similar to the sugar-induced VirA, which suggests that Y293F may be involved in transmitting sugar perception. To further analyze the role of Y293F in such global signal integration, the mutation was then moved into the full-length VirA to examine the impact of sugar. Initial analysis of the β-galactosidase activity of PLKR(Y293F) indicated a generally more responsive phenotype than wild-type PLKR, both in the presence or absence of glucose (Fig. 4.5), consistent with the phenomenon observed in the LKR constructs. More strikingly, however, PLKR(Y293F) showed a very high basal activity in 0 μM AS supplemented with 14 mM glucose. This phenomenon is completely different from the wild-type VirA, in which the

phenolic molecule is absolutely required for kinase activation. Further phenolic dose response was analyzed in Fig. 4.6, with increasing concentration of AS in the absence of sugar. The maximal inducibility of PLKR(Y293F) at 1000  $\mu\text{M}$  AS was about two-fold as that of wild-type VirA (Fig. 4.6), similar to the relative difference seen in LKR(Y293F) and wild-type LKR (Fig. 4.3), and the  $\text{ED}_{50}$  for AS decreased from  $\sim 37$   $\mu\text{M}$  in wild-type PLKR to 1.4  $\mu\text{M}$ . This sensitivity is comparable to the wild-type VirA in the presence of glucose,  $\sim 2$   $\mu\text{M}$  (Fig. 4.4), again consistent with Y293F impacting the sugar effect in phenol perception.

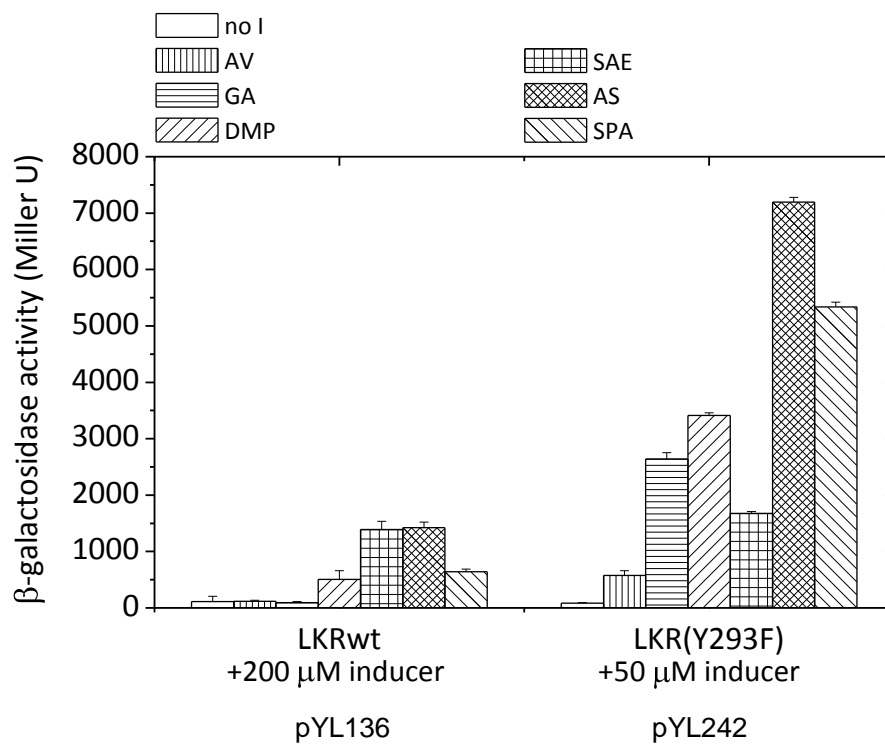
The phenol perception profile of PLKR(Y293F) was also analyzed (Fig. 4.7), and with the significant enhance in phenol response, only 5  $\mu\text{M}$  of phenolics was required for PLKR(Y293F) activity, while 50  $\mu\text{M}$  phenolics was needed for wild-type PLKR in the absence of sugar. Under such concentration, similar phenol specificity profile was observed in PLKR(Y293F) as compared with LKR(Y293F) in Fig. 4.4. To test the sugar activation directly, the glucose dose-response of PLKR(Y293F) was compared directly to wild-type VirA in the presence of AS. As shown in Fig. 4.8, the glucose sensitivity of wild-type PLKR in the presence of 20  $\mu\text{M}$  AS was identical to PLKR(Y293F) in 0  $\mu\text{M}$  AS, which means that the sugar perception is unlikely to be affected by Y293F mutation. Taken together, these results clearly connect the Y293 residue as a central point in integrating the phenol and sugar inputs. While the wild-type VirA remains responsive to both signals, the integration of the combined exposure is attenuated by removing the single oxygen atom on the 293 residue.

**Fig. 4.3 AS induction in LKR(Y293F).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression as a function of the AS concentration when supplemented with 1% glycerol. Wild-type PLKR was also assayed in the presence of 14 mM glucose. **A)** Activity represented in  $\beta$ -galactosidase activity Miller units. **B)** Activity expressed as a percent of the maximal activity with the  $ED_{50}$  value shown in the table.

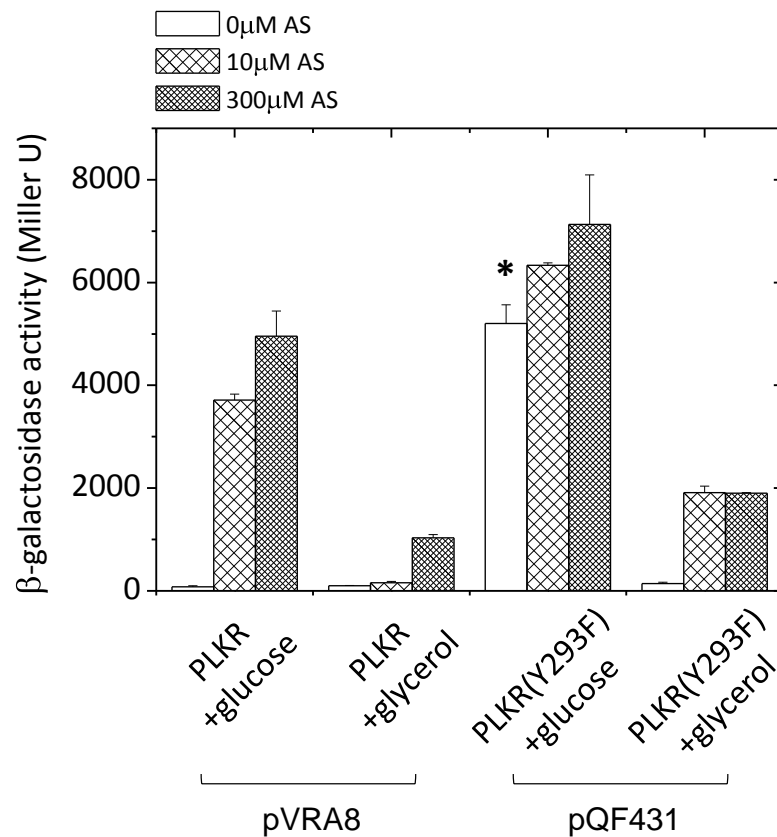


Construct	$ED_{50}$ ( $\mu$ M)
PLKR+glucose	2.2
PLKR+glycerol	18.2
LKR	569.3
LKR(Y293F)	13.3

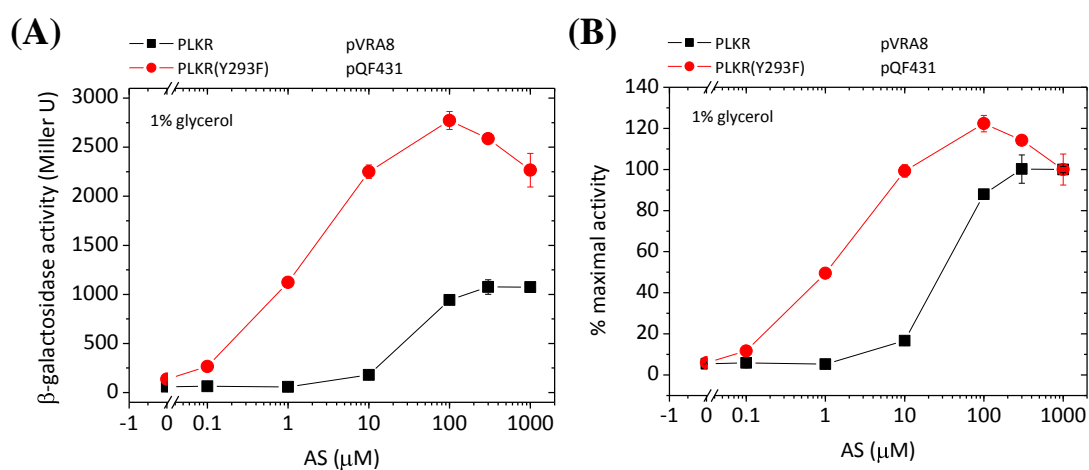
**Fig. 4.4 Phenol specificity of LKR(Y293F).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated LKR constructs were assayed for *vir* gene expression, induced by various phenols. Wild-type LKR was assayed with 200  $\mu$ M phenolic inducer, and LKR(Y293F) was assayed with 50  $\mu$ M phenolic inducer. The chemical structures of the phenolic compounds are shown in Table 3.3.



**Fig. 4.5 AS induction in PLKR(Y293F).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression.  $\beta$ -galactosidase activity of wild-type PLKR and PLKR(Y293F) with 0  $\mu$ M, 10  $\mu$ M, or 300  $\mu$ M AS were assayed with 14 mM glucose or 1% glycerol supplemented, as indicated. The high basal activity of PLKR(Y293F) with only glucose is marked with an asterisk.

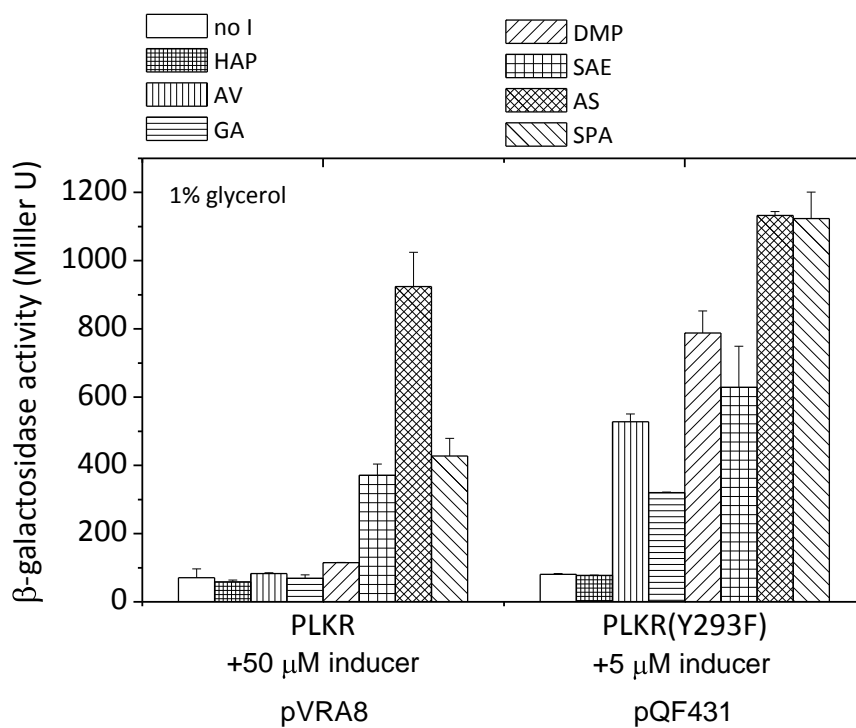


**Fig. 4.6 AS dose response of PLKR(Y293F).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression with increasing AS concentration and supplemented with 1% glycerol. **A)** Activity represented in  $\beta$ -galactosidase activity Miller units. **B)** Activity expressed as a percent of the maximal activity with the ED<sub>50</sub> value shown in the table.



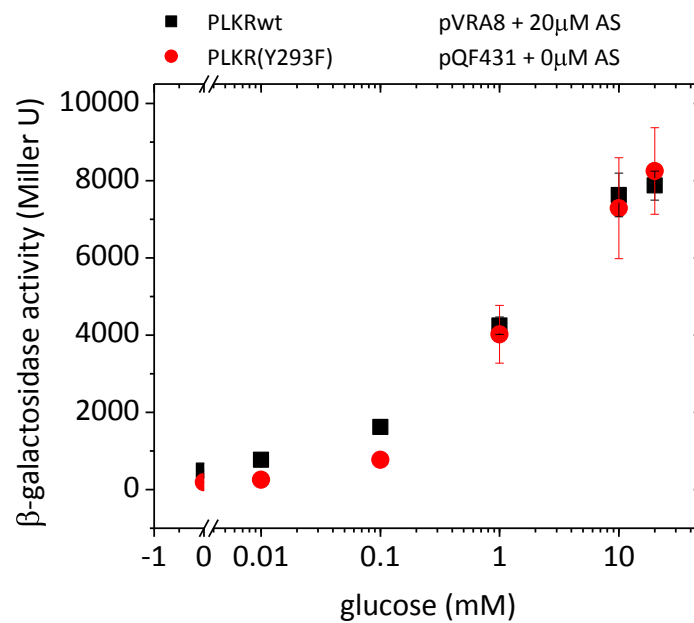
Construct	ED <sub>50</sub> ( $\mu$ M)
PLKRwt	37.2
PLKR(Y293F)	1.4

**Fig. 4.7 Phenol specificity of PLKR(Y293F).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression with indicated phenolic compounds, supplement with 1% glycerol. Wild-type PLKR was assayed with 50  $\mu$ M phenolic inducer, and PLKR(Y293F) was assayed with 5  $\mu$ M phenolic inducer.





**Fig. 4.8 Glucose dose response of PLKR(Y293F).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression as a function of the glucose concentration. Wild-type PLKR in 20  $\mu$ M AS is plotted while PLKR(Y293F) is in the absence of AS.

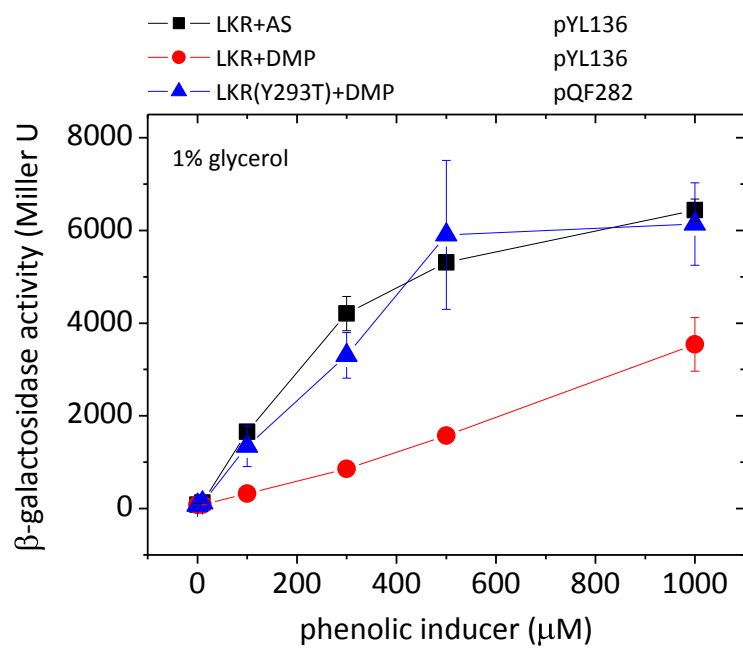


### VirA(Y293T)

Y293T, as described in Chapter 3, displays a very different change in phenol perception compared with Y293F. The generic AS molecule is no longer an inducer for LKR(Y293T), and indeed only phenol inducers without *para*-substitution, such as 2,6-dimethoxyphenol (DMP) and guaiacol (GA), or a small 1-C *para*-substitution, such as syringaldehyde (SAE), are potent inducers (Fig. 3.10B)(Fang 2007). As shown in Fig. 4.9, LKR(Y293T) responds to 2,6-dimethoxyphenol (blue) better than wild-type LKR to the same structure (red). The responsiveness was close to the level of wild-type LKR to the generic AS signal (black).

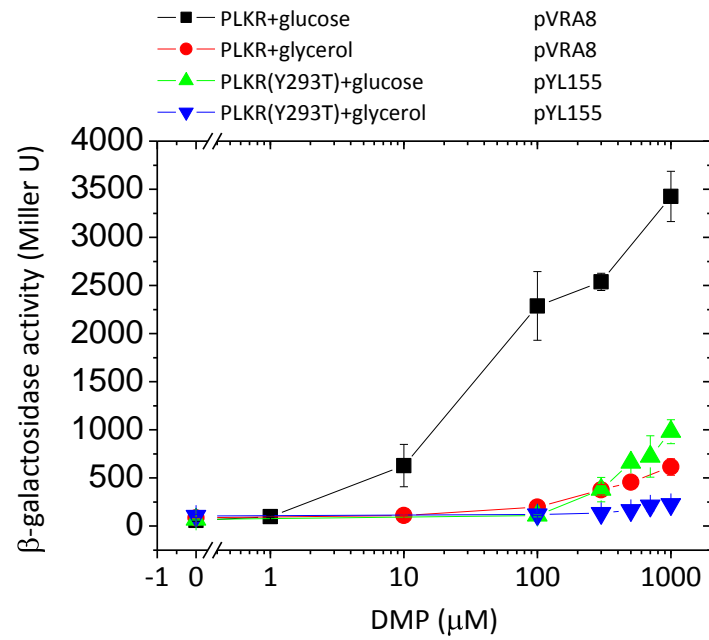
Whether removing the aromatic ring from this residue changes sugar transmission was further analyzed by moving Y293T into the full-length VirA (Fig. 4.10A). However, in full-length VirA, the maximal 2,6-dimethoxyphenol induced activity of PLKR(Y293T) (green) was not even half that of the wild-type VirA (black). The activity of PLKR(Y293T) in the absence of sugar was hardly detected (blue), and even with glucose, the activity did not plateau at 1000  $\mu$ M inducer (green). The phenol recognition profile of PLKR(Y293T) can only be observed in a significantly high phenolic concentration (500  $\mu$ M) as compared to only 10  $\mu$ M for wild-type VirA (Fig. 4.10B), in the presence of glucose, and the phenol specificity was similar to that of LKR(Y293T) (Fig. 3.10B), restricting the active inducers to those with small or no *para*-substitutions.

**Fig. 4.9 DMP dose response of LKR(Y293T).** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression with increasing phenolic inducer concentration, supplemented with 1% glycerol. Wild-type LKR was assayed with AS (black) and 2,6-DMP (red), and LKR(Y293T) was assayed with 2,6-DMP (blue).

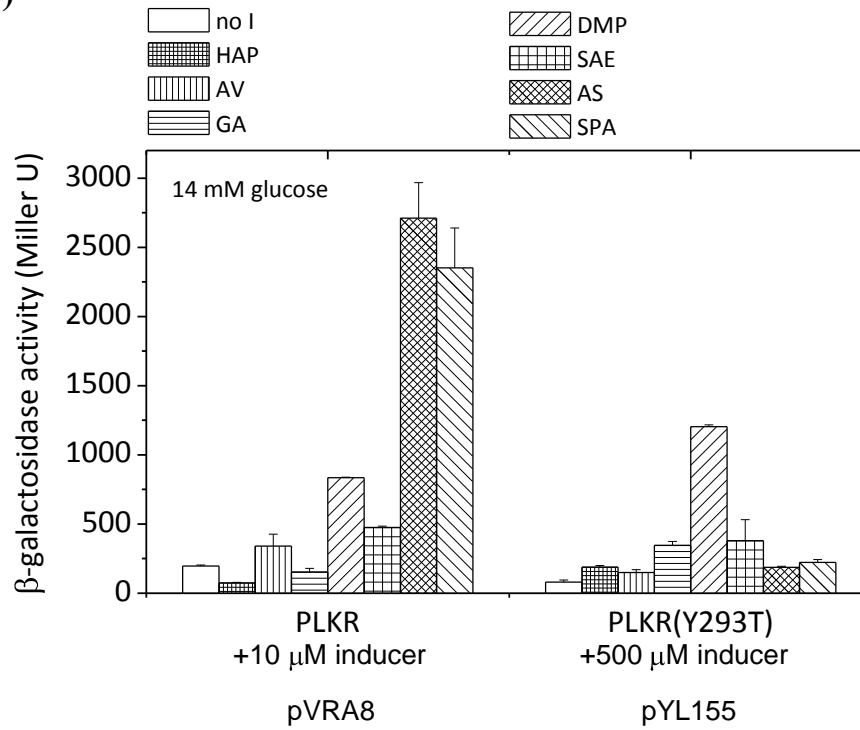


**Fig. 4.10 PLKR(Y293T).** **A)** DMP dose response of wild-type PLKR and PLKR(Y293T) with and without 14 mM glucose. *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression with increasing DMP concentration and supplemented with either 14 mM glucose or 1% glycerol as indicated. **B)** *A. tumefaciens* strain A136 carrying pRG109 and the indicated constructs were assayed for *vir* gene expression with indicated phenolic compounds, supplemented with 14 mM glucose. Wild-type PLKR was assayed with 10  $\mu$ M inducer, and PLKR(Y293T) was assayed with 500  $\mu$ M inducer.

(A)



(B)



## Discussion

The identification of Y293F revealed an unusual phenotype and provided new insights in the understanding of integrating periplasmic sugar sensing and cytoplasmic phenolic perception within VirA. In LKR, Y293F significantly enhanced phenol sensitivity and the maximal activity, and this increase was further demonstrated in PLKR(Y293F). This mutation also broadens the phenol recognition profile in both LKR and PLKR. Moreover, PLKR(Y293F) could be induced only by sugar, returning the assignment of phenol as the critical ON/OFF switch. The sugar sensitivity of wild-type PLKR and PLKR(Y293F) are similar (Fig. 4.8), which indicates that the presumable sugar binding by ChvE remains the same, and the mere sugar inducibility of PLKR(Y293F) most likely derive from a long range conformational change in VirA. Our library designed for phenol specificity, as described in Chapter 3, identified another 293 mutation with a drastically different phenotype, Y293T. When incorporated this mutation into full-length VirA, the phenol sensitivity significantly decreases without changing the phenol specificity (Fig. 4.10).

Taken together, while Y293F seems to be a “high sugar mediator”, Y293T appears to be a “low sugar mediator,” and wild-type Y293 lies in the middle. The 293 residue is located at the N-terminus of the linker HAMP region. A “piston-type” motion was proposed earlier for sugar-mediated kinase regulation, suggesting a sliding motion between the helices modifies the helix interface that lowers the energy barrier of phenol-induced helix rotation (Gao and Lynn 2007). This kind of piston-type motion has been proposed for several transmembrane signaling events in TCS, suggesting a displacement of TM2, either upwards to the extracellular space or downwards into the cytoplasm, upon

signal binding (Fig. 4.11) (Falke and Hazelbauer 2001; Sevvana, Vijayan et al. 2008; Cheung and Hendrickson 2009; Moore and Hendrickson 2009). Take the Tar domain for example, its TM2 region slides downward into the cell when the periplasmic domain binds to aspartate (Fig. 4.11B). A previous study swapping the VirA periplasmic domain with Tar showed a down-regulation upon aspartate binding (Gao and Lynn 2007), which means the VirA TM2 should use the opposite direction of sliding, moving towards the extracellular space, to mediate the sugar perception.

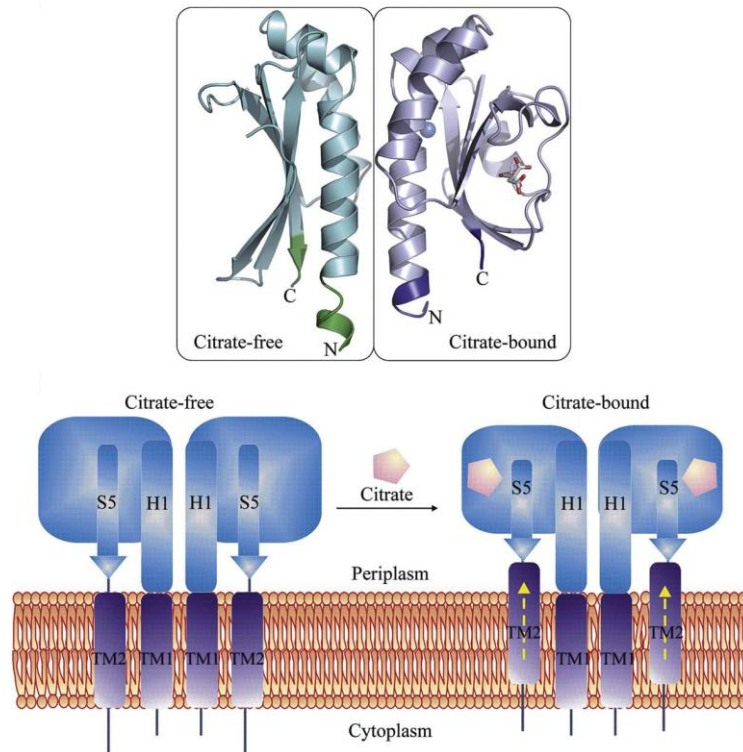
These data have now led to a model for sugar-mediated motion coupling the phenol-induced rotation within VirA (Fig. 4.12). In wild-type VirA, sugar perception mediates an upward sliding of TM2, but the hydroxyl group at aa293 could restrict the helix rotation. The presence of phenol releases this restraint. Since Y293F is missing the hydroxyl group, the helix is free of rotation upon sugar perception, therefore a “screw-type” motion can be induced that activates kinase. Finally, the “low sugar mediator” phenotype of Y293T suggests a hydrophobic environment at this position is needed to propagate sugar transmission. This non-hydrophobic environment of threonine also causes the change in phenol specificity, as described in Chapter 3.

The discovery about the critical role of aa293 in both phenol sensitivity and specificity regulation suggested a host range determinant. The alignment of VirA linker domain sequence from different *Agrobacterium* strains, as seen in Fig. 3.9A, revealed that in most of the strains, including A6 and C58, the native 293 residue is tyrosine, but in strain Ag162, AB2/73, and KU12 is phenylalanine. Strain A6, C58, and KU12 are known as the “wide-host range” pathogen, while strain Ag162 and AB2/73 show a more limited-

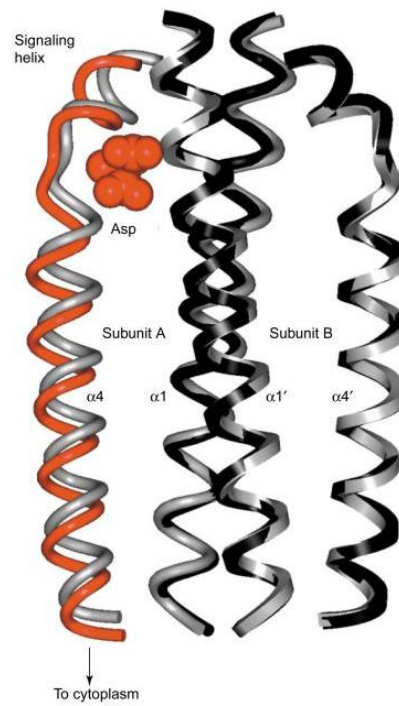
**Fig. 4.11 The “piston-type” motion.** **A)** CitA in *Klebsiella pneumoniae* regulates the transport and anaerobic metabolism of citrate in response to its extracellular concentration (Sevvana, Vijayan et al. 2008). Upon citrate binding, a contraction of the binding domain causes a displacement of TM2 moving upward toward the extracellular space, observed in the crystal structure. **B)** The Tar domain in chemotaxis of *E. coli* is responding to extracellular aspartate concentration (Falke and Hazelbauer 2001). The binding of aspartate triggers a downward displacement of  $\alpha 4$  that propagates into the cytoplasm, as shown by the red helix.



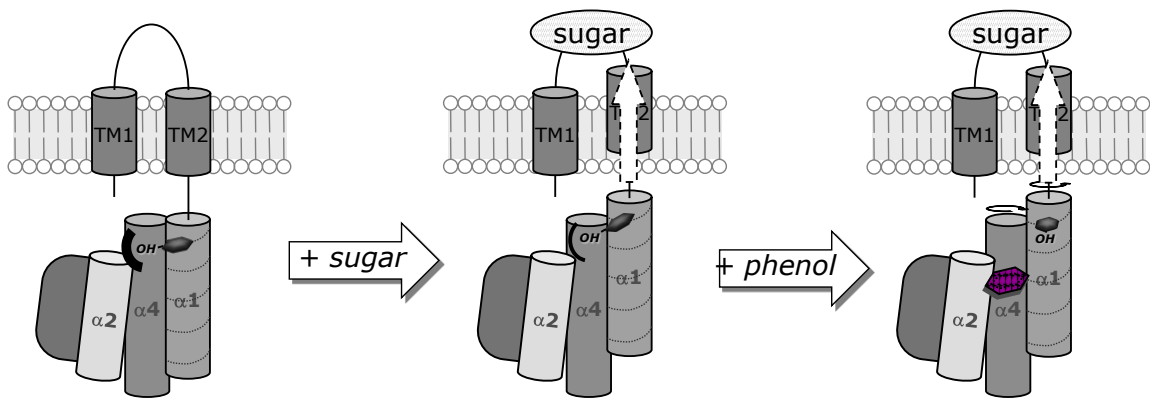
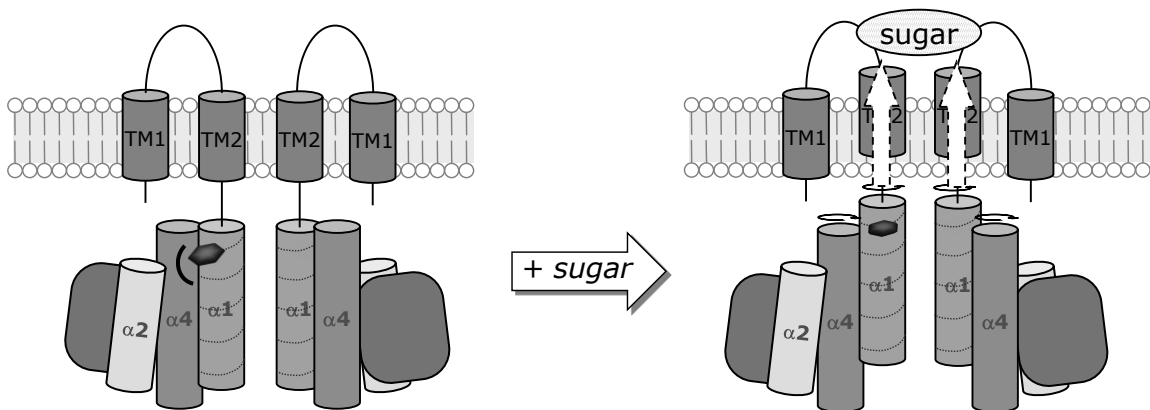
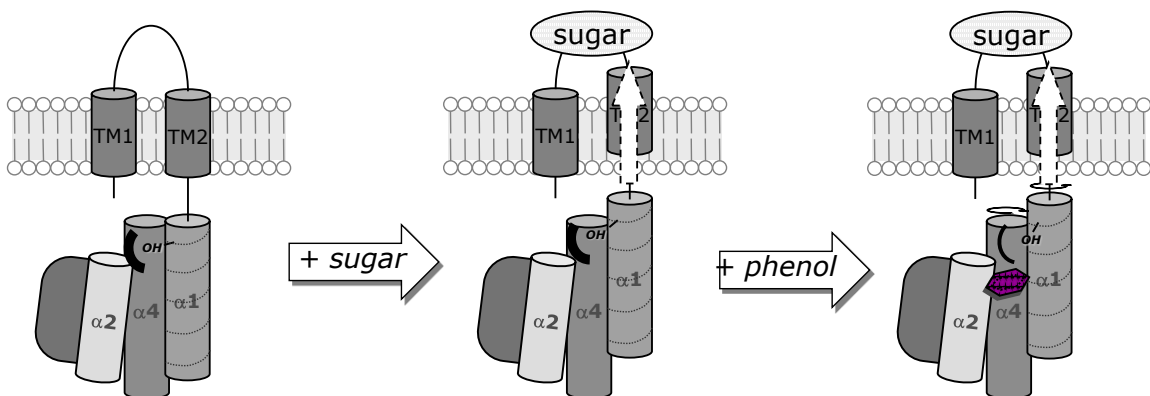
(A)



(B)



**Fig. 4.12 Proposed model of sugar-mediated kinase activation transmitted through Y293.** For clarity, the wild-type and Y293T are shown in monomer, and Y293F is shown in dimer. In wild-type VirA, sugar alone can only mediate the piston-type motion of TM2 moving toward the extracellular space. The presence of phenol is required to break the restriction from the hydroxyl of Y293, and allows the helix to rotate. In Y293F, sugar can mediate a “screw-type” movement, which allows both the piston and the rotational motion. In Y293T, sugar alone only mediates the piston motion, and the addition of phenol relieves part of the restriction of rotation, but can not completely remove it.

**wild-type Y293:****Y293F:****Y293T:**

host-range, and previous studies have suggested the host range of pathogenesis is determined by *virA*. It is indicated that in *virA* promoter region of strain A6 and C58, a *vir*-box that can be up-regulated upon signal sensing exists, but this *vir*-box region is absent in strain Ag162, AB2/73, and KU12 (Fig. 4.13) (Lee, Ha et al. 1998). Therefore, we can speculate that in order to compensate the low concentration of VirA in the limited-host-range strains Ag162 and AB2/73, *Agrobacterium* has evolved the residue at aa293 to phenylalanine to elevate the phenol sensitivity as well as broaden the phenol specificity. Moreover, sugar can be used exclusively as an inducer for plant host transformation. Strain KU12, furthermore, has an even broader phenol sensing profile, presumably constituted from a different phenol binding site as predicted in Chapter 3, which allows it to perceive an even wider range of plant species (Peng, Lee et al. 1998), and with the increase of phenol sensitivity by the phenylalanine residue at aa293, it remains successful as a wide-host-range pathogen. This hypothesis is further supported by another conserved residue at aa299. In strain A6 and C58, a negatively charged glutamate is at aa299, while strains Ag162, AB2/73 and KU12 contain a positively charged arginine or lysine residue. Previous study indicated that changing this residue from glutamate to glutamine enhances phenol sensitivity (Gao and Lynn 2007). Therefore, evolving this residue from glutamate to a positively charged residue in VirA may have a similar effect as the Y-to-F change at aa293, and both can enhance the phenol sensitivity to compensate the low VirA concentration in the cell.

*Agrobacterium* pathogenesis and the host-range determination are tightly controlled at the host-pathogen interface, and both signal sensitivity and structural specificity are critical indicators of host commitment. The need of coordinating multiple

signals with various specificities and concentration threshold represents the actual signaling landscape of the plant host in the soil environment and will define the behavior of the pathogen. Questions still remain in understanding the host-pathogen interface and the mechanism of host commitment, for example whether different strains of *Agrobacterium* triggers the pathogenesis at different levels by perceiving diverse structural characteristics at the signal receptor, or by changing the inducing threshold. Our discovery increases the opportunity to probe the signaling landscape at the host-pathogen interface, which can provide a detailed understanding of the pathogen's behavior in response to this information, and can also help understanding how this pathogenesis process can be regulated. A recently developed method to visualize such interaction using fluorescence imaging technique presents an opportunity to probe the signaling landscape.

Fig. 4.14 shows the preliminary image of *Agrobacterium*-colonized plant and induced fluorescence. The experiments were designed to visualize the roots of *Arabidopsis thaliana* and *Striga asiatica* seedlings utilizing *Agrobacterium* strain AB550, which reports the *vir*-inducing ability by the *virE::gfp* output. An *m-cherry* gene encoding RFP was contained as the internal control for the number of bacteria. *A. thaliana* is a host for *Agrobacterium*, whereas the monocot *S. asiatica* is not. The bacterium colonization and *vir* gene induction were clearly observed under fluorescence microscope in these two plants (Fig. 4.14). As shown in Fig. 4.14A, *A. tumefaciens* colonizes *A. thaliana* wounded root tissues quickly, while the unwounded *A. thaliana* root tissue can be colonized but the colonies are not induced (Fig. 4.14B). On the other hand, *A. tumefaciens* does not colonize *S. asiatica*, either wounded or unwounded (Fig.

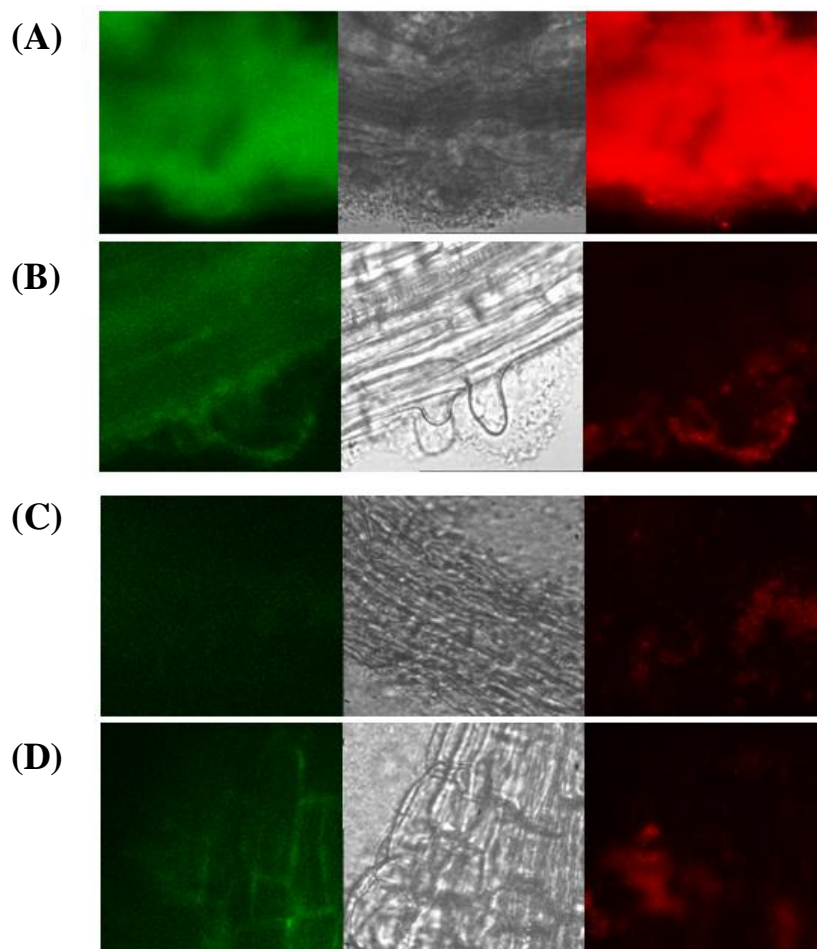
4.14C and D), and even the bacteria that do colonize the *S. asiatica* seedlings do not induce GFP expression, even after 7 days of co-culturing (Fig. 4.14C).

These early fluorescence experiments now can be used to detect the signal distribution of the plant host. The hemiparasite *S. asiatica* maintains virtually no wall bound phenols (Keyes, Palmer et al. 2007). While *A. thaliana* could be successfully colonized by *Agrobacterium* for *vir* gene induction, specifically localized to the wound site, the absence of wall phenols appears to make *S. asiatica* invisible to the bacteria. Alternatively, the root may be producing compounds that inhibit *vir* induction, similar to HDMBOA identified from maize, to resist *Agrobacterium* colonization and pathogenesis. In either case, these data suggest features of the plant wound site that are critical to pathogenesis have gone unnoticed in the *in vitro* analyses, and defining these differences is critical to plant biotechnology as well as the general understanding of the signaling landscape in pathogenesis. The ability to monitor pathogenesis of *A. tumefaciens* *in planta* sets the stage for changing the input information with the VirA mutants identified in this chapter, and further analyses of the diverse plant hosts.

**Fig. 4.13 *vir*-box in *virA*.** Nucleotide sequence alignment of the *virA* promoter region of different *Agrobacterium* strains. The ATG start codon is shown in bold with an asterisk. The conserved Shine-Dalgarno (SD) sequence, the -10 and -35 regions, and the signal-inducible *vir*-box are underlined.

	<b><i>vir</i>-box</b>			
pTi15955	<u>GCTTGTGTGTTCAATTCTTCTCTGT</u> <u>TTC</u> <u>ACTTGAAACAA</u> ACTGAATATATATTCCC----GCTTTCAAAGCC			
pTiA6	<u>GCTTGTGTGTTCAATTCTTCTCTGT</u> <u>TTC</u> <u>ACTTGAAACAA</u> ACTGAATATATATTCCC----GCTTTCAAAGCC			
pTiBo542	<u>GCTTGTGTGTTCAATTCTTCTCTGT</u> <u>TTC</u> <u>ACTTGAAACAA</u> ATTGAATAGATATTCCC----GCTTTCAAAGCC			
pTiSAKURA	ATCCATTGGAATGCCGAGGCTGT <u>GTTTCATTTGAAACAC</u> ACTGAGTCGACGTTTGT----TGCTTCAA-CCC			
pTiC58	ATCCATTGGAATGCCGAGGCTGT <u>GTTTCATTTGAAACAC</u> ACTGAGTCGACGTTTGT----TGCTTCAAACCC			
pRiA4	ATC-GTTGGAATGCCAGGGCTGT <u>TTCATTTGAAACAA</u> ACTGAGTCGACGTTTGT----GATTTCAAACCC			
pTiS4	GCC-ATATTTTTGTCTGACC--AGCTTGGCTTGTTCGT-CACCACCGAAATGCCG----AGTAAACAGAAA			
pTiKU12	GCC-ATATTTTTGTCTGACC--AGCTTGGCTTGTTCGT-CAACTATGGAAGGCCG----AGTGACGGAAAA			
pTiAB2/73	TAG-CTTGGTTTGTTCGGTCGAAATCTGACTAGCAACAAGTCGCACCGGGATGCAG-----GTGATAAGAAA			
pTiAg162	GCTGGAATTTTCTCCGTGCGGCAATTCGTATCACGGATGTAAGAATGATAAGTAATCCGGGGTGAGATCA			
	<b>-35</b>	<b>-10</b>		<b>SD</b>
pTi15955	<u>ATTTACAAATCCTCTCGTGCAGC-CTAA-CGCAGTTGGCGTG--GATCAAAGC--GTTGGCACGAGGAAGTA</u>			
pTiA6	<u>ATTTACAAATCCTCTCGTGCAGC-CTAA-CGCAGTTGGCGTG--GATCAAAGC--GTTGGCACGAGGAAGTA</u>			
pTiBo542	<u>ATTTACAAATCCTCTCGTGCAGC-CTAA-CGCCATTGGCGTG--GAACAA-GC--GTTGGCACGAGGAAGTA</u>			
pTiSAKURA	<u>ATTTACAAACCCTACTGTGCGGC-CTAAGGGCCACCGGGGTG--GGACTGAGC--GCTGGTACGAGGACGTA</u>			
pTiC58	<u>ATTTACAAACCCTACTGTGCGGC-CTAAGGGCCACCGGGGTG--GGACTGAGC--GCTGGTACGAGGACGTA</u>			
pRiA4	<u>ATTTACAAAGCCTACCGTGCAGC-CTAAGCGCCACGGGAGTG--GGACTGAGC--GTTGGCACGAGGACATA</u>			
pTiS4	CTGCATCAAATC-ACCGGGCGTCGCCGTTGCACGATGTCTG-TGAGTTGGATTGGCTGT-CAGAAGGCGAGG			
pTiKU12	CTGTGTCAGATC-ACTGGGCGTTGCCGTTGCACGACGTTGAATGAGTTGGATTGGCTGT-GAAGGATGAGG			
pTiAB2/73	CTGGGTGAGA-C-AC-GGGCGCC-TTGGTTGCAACCGCTCTGTTCAATTGGATTGGGTATTAGAAGGAG-GG			
pTiAg162	TCCGGGTGAGATC-ATCCGGTGACGCTAATCCCTCGACATCCGCATAATCGGGTCGGCTGTTGCAAAGAGAGG			
	<b>***</b>			
pTi15955	AGTGCG- <b>ATGAAC</b> GGAAGA			
pTiA6	AGTGCG- <b>ATGAAC</b> GGAAGATATTC			
pTiBo542	AGTGCG- <b>ATGAAC</b> GGAAGGTATTCA			
pTiSAKURA	AGTGCG <b>ATGAAT</b> GGAAGGTATTCACC			
pTiC58	AGTGCG <b>ATGAAT</b> GGAAGGTATTCA			
pRiA4	AGTGCG <b>ATGAAT</b> GGAAGGTATTCA			
pTiS4	GGACCGC <b>ATG</b> TTGGGATCATTTATC			
pTiKU12	TGACCAC <b>ATG</b> TTGGGATCATTT			
pTiAB2/73	TGACCGC <b>ATG</b> TTGGGATCATTCG			
pTiAg162	TAACTAA <b>ATG</b> ACTTGGGATCGTCTG			

**Fig. 4.14 Fluorescence image of the plants co-cultivating with *Agrobacterium*.** The three columns in each panel include: the bright field image (middle), the confocal image taken with a GFP emission filter (left), and the confocal image taken with an *m-cherry* emission filter (right). **A)** Wounded root tissue of a 5-day-old *A. thaliana* seedling, imaged 66 hrs after vacuum infiltration with AB550. **B)** Unwounded root tissue of a 5-day-old *A. thaliana* seedling 24 hrs after vacuum infiltration. **C)** Wounded root tissue of a *S. asiatica* seedling, imaged 7 days after co-culturing with AB550. **D)** Unwounded root tissue of a *S. asiatica* seedling, imaged 7 days after co-culturing with AB550.





## CHAPTER 5

### SUMMARY

*Agrobacterium* pathogenesis was initially identified in crown gall tumors (Smith and Townsend 1907). This general broad-host-range pathogenesis is remarkable, allowing it to infect most dicotyledonous plants. The ability to mediate inter-kingdom gene transfer has attracted much attention and has been exploited to engineer plants outside the natural host range, for example, cereal crops. Host recognition involves specifically recognizing, attaching to, and colonizing the host all prior to the initiation of gene transfer. Specific molecular signals at the host-pathogen interface, sugar, acidic pH, and phenols, are all that is required for sensing the host.

I was most interested in phenol perception by the type I TCS of VirA/VirG not only because phenols are the most critical plant-derived signals, but also because it offers the greatest structural diversity and accessibility. What enables VirA to respond to many plant-derived signals and how is the coordination with other regulatory domains structurally controlled? Combining various results allowed a simple rotational ratchet motion to be proposed and specific tests to probe such motion directly were carried out. Specifically, I fused the GCN4 leucine zipper domains throughout the coiled-coil region and was able to predict the His474 positions by counting the heptad repeats. The activities of different fusions allowed me to assign both “ON” and “OFF” conformations and to further characterize this ratcheting motion. Furthermore, the clear ratcheted activity of GCN4-426 establishes the rotational motion to be coherently mediated from

the linker domain to the kinase core. However, I did not observe continuous and rigid heptads maintained throughout aa426-474 (Fig. 2.15), possibly indicating that the flexible heptads can allow the coiled-coil to be dynamic in motion and regulation.

Successfully engineering the rotational motion together with the observation that aa426-437 are repressive to the kinase activity led to the prediction that mutations could be identified that would bias the coiled-coil interface and attenuate signal transmission. Indeed, mutations were identified that stabilized the dimerization interface by hydrophobic or electrostatic interactions at  $\alpha 4$ - $\alpha 4'$ . Further identification of a salt-bridge between  $\alpha 1$  and  $\alpha 4$  supports the four-helix bundle architecture at the linker dimerization interface and strengthens the argument that the conformation of the helix bundle core is critical in signal transmission. Analysis of the *trans*-phosphorylation mechanism was achieved by direct amino acid insertion, which identified the clockwise rotation being the active motion. Combining the predicted coiled-coil sequence, extensive genetic data, and the structure of the VirA homolog *T. maritima* HK0853, a functional model of *trans*-phosphorylation is proposed (Fig. 2.16). This proposed model, using a simple, general motion for kinase activation, could be applied to future studies on other functional proteins.

Based on the mechanistic foundation that a general motion mediates kinase activation, further understanding of the host commitment then requires identification of the signal binding site and the mechanism of perceiving different chemical specificities. The structural model of the linker domain was used to probe the possible phenol binding site computationally. In addition to the linker domain structure initially developed (Fig. 1.6B), another structural model, generated by the automated method SAM-T06, showed

better docking results. Both the docking of phenolics to this structure, and the structure of linker<sub>KU12</sub>, were mostly consistent with the GAF prediction. To the first approximation, the two linker domains from different *Agrobacterium* strains should adopt a similar fold and binding site, so further genetic analysis was carried out, focusing on the hydrophobic residues in the predicted phenol binding region. The mutagenesis study did not clearly identify the phenol binding site, and more mutations could be rationally designed. However, a clean *in vitro* experiment should better reveal the protein-ligand association, so I started a preliminary characterization of the linker domain and showed it to be consistent with the GAF fold. Even though some technical problems to stabilize the linker domain exist, these results should set the stage for the future characterization of the phenol binding site, the specificity regulation, and the proposed "proton transfer" model upon phenol binding.

One strategy for the pathogen to identify and infect different hosts is by perceiving a broad range of chemical structures released from different hosts. In *Agrobacterium*, the phenol specificity should also be determined by the linker domain. The mutations identified for the specificity change were not located at the predicted phenol binding site but at the helix bundle region. This kind of "distant mutation" is observed in many enzymes, regulating substrate binding by a long-range conformational change that propagates to the binding site. This would make sense to the specificity-switch mutations identified in the linker domain because these mutations mediate the specificity towards the *para*-substitution, whose binding region was predicted to be most flexible, therefore should be most susceptible to a long-range conformational change.

Another important regulation of the histidine kinase's activity is due to the tightly controlled domain coordination. Previous studies have indicated the importance of coordinating the periplasmic, the linker, and the receiver domains in kinase activity regulation. The regulation includes augmenting the phenolic response by the sugar/H<sup>+</sup> perception at the periplasmic domain, to change the phenol specificity by the receiver mutants, and to enhance or decrease the kinase activity according to VirA/VirG concentration, which is mediated by the receiver domain. Previously we hypothesized that the receiver domain interacts with the linker to regulate phenol perception, and a mutagenesis library was designed to probe such interaction. Surprisingly, one of the mutations was identified at the same residue as the previously constructed specificity switch library, Y293F. This Y293 residue seems to be important in integrating multiple signal inputs, both by altering the signal binding site (distantly) and coordinating domain interactions. I further moved these mutations into the full-length VirA because both sugar and phenol effects could be nicely represented. The *in vivo* activity showed that a small perturbation at the side chain of aa293 induces great differences in the phenol response. Removing a single oxygen from the side chain, resulting in Y293F, greatly enhanced the phenol response, both in sensitivity and specificity, while removing the aromatic ring, resulting in Y293T, caused the complete opposite effect, to reduce the phenol sensitivity and narrowed the specificity. Another remarkable phenotype is that Y293F no longer requires a phenol for kinase activation, and sugars can function as independent inducers.

Comparison of the VirA amino acid sequence and the *virA* promoter sequence revealed a conservation of a Y residue in most of the *Agrobacterium* strains with a *vir*-box at the promoter region, while an F is maintained in the strains without it. The *vir*-box

was indicated to up-regulate *vir* expression thereby mediating the wide-host-range vs. the limited-host-range (Lee, Ha et al. 1998), so I speculated that the amino acid residue at aa293 may have evolved from Y to F in the limited-host-range strains to augment the signal sensing. The phylogenetic tree constructed from ten *Agrobacterium* strains (Fig. 5.1) showed that the limited-host-range bacterium without *vir*-box and has F at aa293 evolved from the wide-host-range *Agrobacterium*. A particular strain, KU12, has no *vir*-box but is still a wide-host-range bacterium. This may be the result of a further evolved amino acid sequence adapting a broader phenol specificity so that it can perceive a wider range of plant hosts.

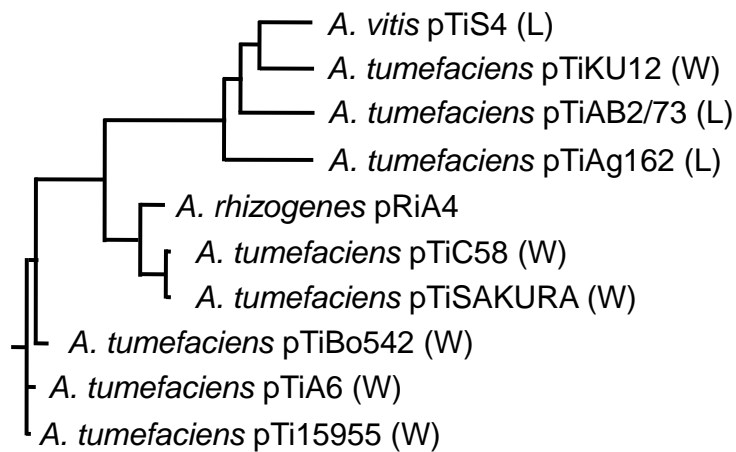
The possible linker-receiver interaction was further analyzed *in vitro* by chemical cross-linking. The result indicated that this region interacts with the R domain through aa445-471, which was previously assigned as the coiled-coil of the kinase in Chapter 2. I hypothesize that the Y293F mutation, analyzed in Chapter 4, creates a more flexible helical association, which can be propagated through the four-helix bundle and the coiled-coil, to relieve R domain repression. Further characterization of such long-range conformational changes regulating domain association is needed.

Based on the discoveries in this dissertation and previous literature, I propose an overall domain coordination model of VirA (Fig. 5.2A). In this model, the linker domain is connected to the kinase through a continuous helix core. The Y293 residue located at the N-terminus of this helix bundle integrates the periplasmic sugar/H<sup>+</sup> perception, transmitted from TM2, and the phenol perception at the linker domain for activity. The receiver domain is proposed to interact with the coiled-coil of the kinase, N-terminus of the conserved His474, and functions across the subunits. The conserved His474 is buried

at the dimerization interface of the DHP domain, and the rotation, mediated by the coiled-coil, brings it close to the CA domain of the other subunit for *trans*-phosphorylation (Chapter 2). In this model, complete regulation of kinase activity comes from the N-terminus of His474, and the only association at the C-terminus of the kinase is VirG. The association interface with VirG is proposed to be similar to the interaction between other HK/RR cognate pairs, such as Spo0B-Spo0F, which was characterized in previous studies, and the crystal structure of *T. maritima* HK853-RR468 (Fig. 5.2B)(Zapf, Sen et al. 2000; Varughese 2002; Laub and Goulian 2007; Casino, Rubio et al. 2009).

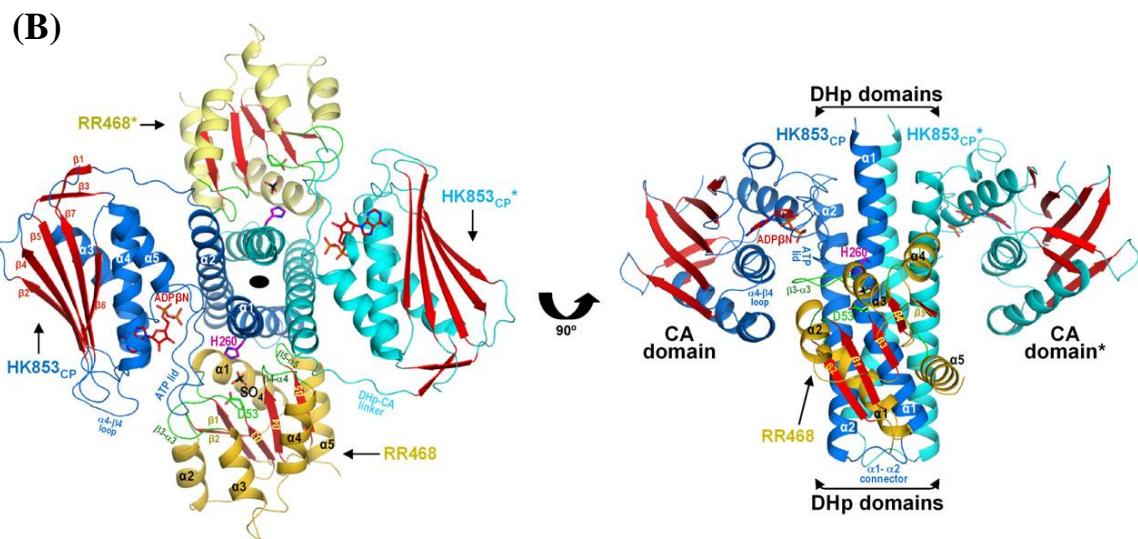
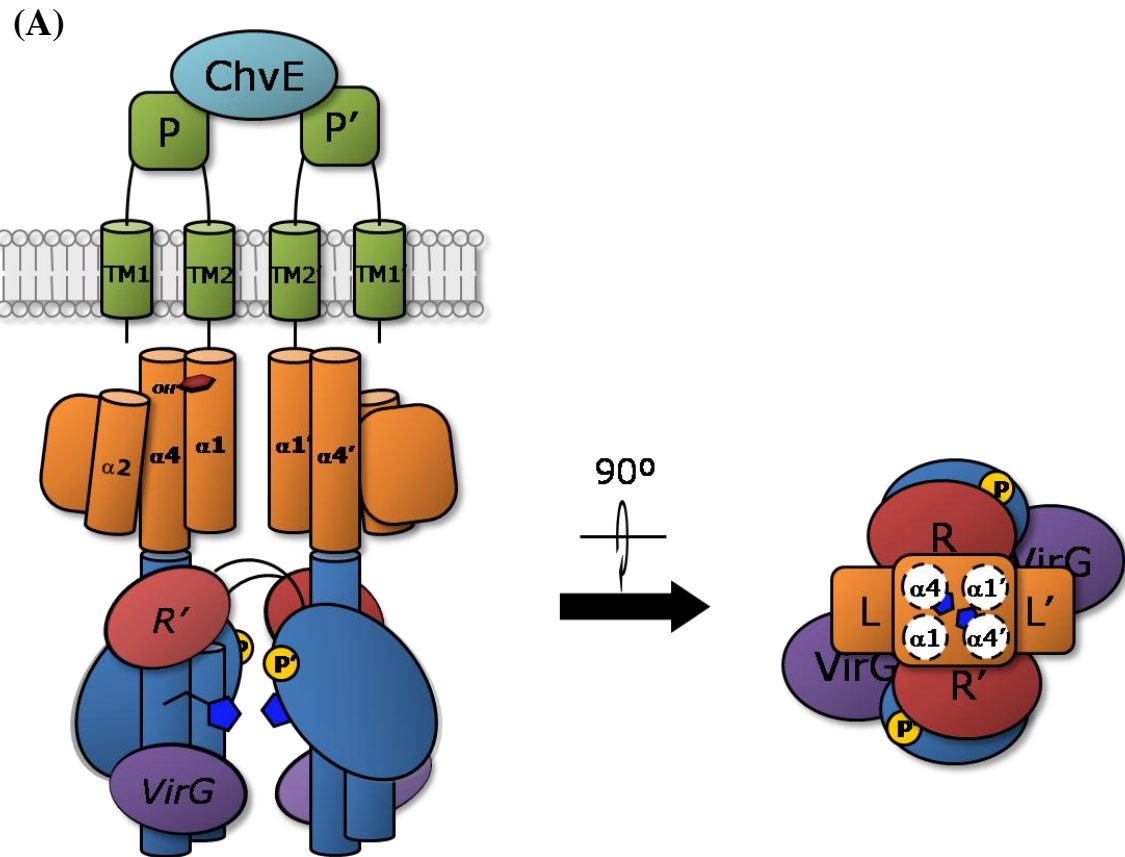
The functional model of VirA will greatly facilitate further research to understand the molecular mechanism at the host-pathogen interface. A recently developed fluorescence imaging technique can be further applied to probe the plant signaling landscape in real time. An in-depth knowledge of the signal perception and integration in *A. tumefaciens* enhances our understanding of general host perception and expands our understanding of the communication between different kingdoms of life. Further engineering of these natural mechanisms will not only allow for the design of inhibitors to interrupt the signaling pathways, but will also make possible the engineering of general signal perception modules for synthetic biology.

**Fig. 5.1 Phylogenetic tree of ten *Agrobacterium* strains.** The phylogenetic tree was constructed based on the VirA sequence of ten different *Agrobacterium* strains. The wide-host-range pathogen is marked with a (W) and the limited-host-range pathogen with an (L).



**Fig. 5.2 Proposed model of VirA/VirG domain coordination of.** **A)** The periplasmic, linker, kinase, and receiver domains of VirA are shown in green, orange, blue, and red, respectively. ChvE is shown in cyan, and VirG is purple. The two subunits are differentiated by a prime (') sign. The R domain is proposed to interact with the kinase coiled-coil region of the other subunit, while VirG is proposed to associate at the C-terminus of the conserved His474. A different view from the membrane shows the corresponding domain coordination. The periplasmic domain is omitted in this view for clarity. **B)** Crystal structure of *Thermotoga maritima* HK853-RR468, shown from two different angles (Casino, Rubio et al. 2009). The dimer of HK853 is shown (blue and cyan:  $\alpha$ -helices; red:  $\beta$ -sheets). The two RR468 are shown (gold and light yellow:  $\alpha$ -helices; red:  $\beta$ -sheets). The conserved His260 in HK853, Asp53 in RR468, and the ADP $\beta$ N molecules are shown in ball-and-stick.





## REFERENCES

- Airola, M. V., K. J. Watts, et al. (2010). "Structure of Concatenated HAMP Domains Provides a Mechanism for Signal Transduction." Structure **18**(4): 436-448.
- Albanesi, D., M. Martin, et al. (2009). "Structural plasticity and catalysis regulation of a thermosensor histidine kinase." Proceedings of the National Academy of Sciences **106**(38): 16185-16190.
- Ankenbauer, R. G. and E. W. Nester (1990). "Sugar-mediated induction of *Agrobacterium tumefaciens* virulence genes: structural specificity and activities of monosaccharides." J. Bacteriol. **172**(11): 6442-6446.
- Appleman, J. A., L.-L. Chen, et al. (2003). "Probing Conservation of HAMP Linker Structure and Signal Transduction Mechanism through Analysis of Hybrid Sensor Kinases." J. Bacteriol. **185**(16): 4872-4882.
- Aravind, L. and C. P. Ponting (1997). "The GAF domain: an evolutionary link between diverse phototransducing proteins." Trends in Biochemical Sciences **22**(12): 458-459.
- Aravind, L. and C. P. Ponting (1999). "The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins." FEMS Microbiology Letters **176**(1): 111-116.
- Arnold, K., L. Bordoli, et al. (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.
- Brencic, A., E. R. Angert, et al. (2005). "Unwounded plants elicit *Agrobacterium* vir gene induction and T-DNA transfer: transformed plant cells produce opines yet are tumour free." Molecular Microbiology **57**(6): 1522-1531.
- Brencic, A., Q. Xia, et al. (2004). "VirA of *Agrobacterium tumefaciens* is an intradimer transphosphorylase and can actively block vir gene expression in the absence of phenolic signals." Molecular Microbiology **52**(5): 1349-1362.
- Bundock, P., A. den Dulk-Ras, et al. (1995). Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*.
- Burkhard, P., J. Stetefeld, et al. (2001). "Coiled coils: a highly versatile protein folding motif." Trends in Cell Biology **11**(2): 82-88.
- Cai, S.-J. and M. Inouye (2003). "Spontaneous Subunit Exchange and Biochemical Evidence for Trans-autophosphorylation in a Dimer of *Escherichia coli* Histidine Kinase (EnvZ)." Journal of Molecular Biology **329**(3): 495-503.
- Cangelosi, G. A., R. G. Ankenbauer, et al. (1990). "Sugars Induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein." PNAS **87**(17): 6708-6712.
- Casino, P., V. Rubio, et al. (2009). "Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction." Cell **139**(2): 325-336.
- Chang, C. and S. Winans (1992). "Functional roles assigned to the periplasmic, linker, and receiver domains of the *Agrobacterium tumefaciens* VirA protein." J. Bacteriol. **174**(21): 7033-7039.

- Chang, C. H., J. Zhu, et al. (1996). "Pleiotropic phenotypes caused by genetic ablation of the receiver module of the *Agrobacterium tumefaciens* VirA protein." J. Bacteriol. **178**(15): 4710-4716.
- Charles, T. C., S. Jin, et al. (1992). "Two-Component Sensory Transduction Systems in *Phytophthora*." Annual Review of Phytopathology **30**(1): 463-484.
- Cheung, J. and W. A. Hendrickson (2009). "Structural Analysis of Ligand Stimulation of the Histidine Kinase NarX." Structure (London, England : 1993) **17**(2): 190-201.
- Cho, H. S., S.-Y. Lee, et al. (2000). "NMR structure of activated CheY." Journal of Molecular Biology **297**(3): 543-551.
- Christie, P. J. (2004). "Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems." Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1694**(1-3): 219-234.
- Cochran, A. G. and P. S. Kim (1996). "Imitation of *Escherichia coli* Aspartate Receptor Signaling in Engineered Dimers of the Cytoplasmic Domain." Science **271**(5252): 1113-1116.
- Cornilescu, G., A. T. Ulijasz, et al. (2008). "Solution Structure of a Cyanobacterial Phytochrome GAF Domain in the Red-Light-Absorbing Ground State." Journal of Molecular Biology **383**(2): 403-413.
- de Groot, M. J. A., P. Bundock, et al. (1998). "*Agrobacterium tumefaciens*-mediated transformation of filamentous fungi." Nat Biotech **16**(9): 839-842.
- DeCleene, M. and J. DeLey (1976). "The host range of crown gall." Bot. Rev. **42**: 389-466.
- Dixon, R. A., L. Achnine, et al. (2002). "The phenylpropanoid pathway and plant defence—a genomics perspective." Molecular Plant Pathology **3**(5): 371-390.
- Doty, S. L., M. C. Yu, et al. (1996). "Mutational analysis of the input domain of the VirA protein of *Agrobacterium tumefaciens*." J. Bacteriol. **178**(4): 961-970.
- Duban, M. E., K. Lee, et al. (1993). "Strategies in pathogenesis: mechanistic specificity in the detection of generic signals." Molecular Microbiology **7**(5): 637-645.
- Dye, F. and F. M. Delmotte (1997). "Purification of a protein from *Agrobacterium tumefaciens* strain A348 that binds phenolic compounds." Biochemical Journal **321**: 319-324.
- Falke, J. J. and G. L. Hazelbauer (2001). "Transmembrane signaling in bacterial chemoreceptors." Trends in Biochemical Sciences **26**(4): 257-265.
- Fang, F. (2007). Signal perception and transmission through the VirA/VirG two-component system in *Agrobacterium tumefaciens*. Department of Chemistry and Biology. Atlanta, Emory University. **Doctor of philosophy**: 1-143.
- Finn, R. D., J. Mistry, et al. (2010). "The Pfam protein families database." Nucleic Acids Research **38**(suppl 1): D211-D222.
- Gao, R. and D. G. Lynn (2005). "Environmental pH Sensing: Resolving the VirA/VirG Two-Component System Inputs for *Agrobacterium* Pathogenesis." J. Bacteriol. **187**(6): 2182-2189.
- Gao, R. and D. G. Lynn (2007). "Integration of Rotation and Piston Motions in Coiled-Coil Signal Transduction." J. Bacteriol. **189**(16): 6048-6056.
- Gao, R. and A. M. Stock (2009). "Biological Insights from Structures of Two-Component Proteins." Annual Review of Microbiology **63**(1): 133-154.

- Gelvin, S. B. (2000). "Agrobacterium and plant genes involved in T-DNA transfer and integration." Annu. Rev. Plant Physiol. Plant Mol. Biol. **51**: 223-256.
- Gelvin, S. B. (2003). "Agrobacterium-mediated plant transformation: the biology behind the "Gene-Jockeying" tool." Microbiol. Mol. Biol. Rev. **67**(1): 16-37.
- Gelvin, S. B. (2005). "Agricultural biotechnology: Gene exchange by design." Nature **433**(7026): 583-584.
- Gilles-Gonzalez, M.-A. and G. Gonzalez (2004). "Signal transduction by heme-containing PAS-domain proteins 10.1152/jappphysiol.00941.2003." J Appl Physiol **96**(2): 774-783.
- Goetz, H., K. Christine, et al. (2009). Agrobacterium-Mediated Gene Transfer to Cereal Crop Plants: Current Protocols for Barley, Wheat, Triticale, and Maize, Hindawi Publishing Corporation.
- Gong, W., B. Hao, et al. (1998). "Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction." Proceedings of the National Academy of Sciences **95**(26): 15177-15182.
- Grimsley, N., B. Hohn, et al. (1986). "'Agroinfection,' an alternative route for viral infection of plants by using the Ti plasmid." Proceedings of the National Academy of Sciences of the United States of America **83**(10): 3282-3286.
- Grimsley, N., T. Hohn, et al. (1987). "Agrobacterium-mediated delivery of infectious maize streak virus into maize plants." Nature **325**(6100): 177-179.
- He, F., G. R. Nair, et al. (2009). "Molecular basis of ChvE function in sugar binding, sugar utilization and virulence in *Agrobacterium tumefaciens*." J. Bacteriol. **191**: 5802-5813.
- Heath, J. D., M. I. Boulton, et al. (1997). "Discrete Regions of the Sensor Protein VirA Determine the Strain-Specific Ability of *Agrobacterium* to Agroinfect Maize." Molecular Plant-Microbe Interactions **10**(2): 221-227.
- Hefti, M. H., K.-J. Francoijs, et al. (2004). "The PAS fold: A redefinition of the PAS domain based upon structural prediction." Eur J Biochem **271**(6): 1198-1208.
- Hess, K. M., M. W. Dudley, et al. (1991). "Mechanism of Phenolic Activation of *Agrobacterium* Virulence Genes: Development of a Specific Inhibitor of Bacterial Sensor/Response Systems." PNAS **88**(17): 7854-7858.
- Hiei, Y., S. Ohta, et al. (1994). "Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA." The Plant Journal **6**(2): 271-282.
- Ho, Y.-S. J., L. M. Burden, et al. (2000). "Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor." EMBO J. **19**(20): 5288-5299.
- Hoch, J. A. (2000). "Two-component and phosphorelay signal transduction." Current Opinion in Microbiology **3**(2): 165-170.
- Horsman, G. P., A. M. F. Liu, et al. (2003). "Mutations in Distant Residues Moderately Increase the Enantioselectivity of *Pseudomonas fluorescens* Esterase towards Methyl 3Bromo-2-methylpropanoate and Ethyl 3Phenylbutyrate." Chemistry – A European Journal **9**(9): 1933-1939.
- Hulko, M., F. Berndt, et al. (2006). "The HAMP Domain Structure Implies Helix Rotation in Transmembrane Signaling." Cell **126**(5): 929-940.

- Ishida, Y., H. Saito, et al. (1996). "High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*." Nat Biotech **14**(6): 745-750.
- Jackson, L. K., J. Baldwin, et al. (2004). "Multiple Active Site Conformations Revealed by Distant Site Mutation in Ornithine Decarboxylase<sup>†,‡</sup>." Biochemistry **43**(41): 12990-12999.
- Jin, S., T. Roitsch, et al. (1990). "The VirA protein of *Agrobacterium tumefaciens* is autophosphorylated and is essential for vir gene regulation." J. Bacteriol. **172**(2): 525-530.
- Jin, S. G., R. K. Prusti, et al. (1990). "Phosphorylation of the VirG protein of *Agrobacterium tumefaciens* by the autophosphorylated VirA protein: essential role in biological activity of VirG." J. Bacteriol. **172**(9): 4945-4950.
- Johnson, W. C. (1999). "Analyzing protein circular dichroism spectra for accurate secondary structures." Proteins: Structure, Function, and Bioinformatics **35**(3): 307-312.
- Joubert, P., D. Beaupere, et al. (2002). "Effects of phenolic compounds on *Agrobacterium vir* genes and gene transfer induction--a plausible molecular mechanism of phenol binding protein activation." Plant Science **162**(5): 733-743.
- Karplus, K., C. Barrett, et al. (1998). "Hidden Markov models for detecting remote protein homologies." Bioinformatics **14**(10): 846-856.
- Karplus, K., R. Karchin, et al. (2003). "Combining local-structure, fold-recognition, and new fold methods for protein structure prediction." Proteins: Structure, Function, and Bioinformatics **53**(S6): 491-496.
- Keyes, W. J., A. G. Palmer, et al. (2007). "Semanogenesis and the parasitic angiosperm *Striga asiatica*." Plant J. **51**: 707-716.
- Kiefer, F., K. Arnold, et al. (2009). "The SWISS-MODEL Repository and associated resources." Nucleic Acids Research **37**(suppl 1): D387-D392.
- Kunik, T., T. Tzfira, et al. Genetic transformation of HeLa cells by *Agrobacterium*, The National Academy of Sciences.
- Laub, M. T. and M. Goulian (2007). "Specificity in Two-Component Signal Transduction Pathways." Annual Review of Genetics **41**(1): 121-145.
- Lee, J.-M., H. Y. Cho, et al. (2008). "O<sub>2</sub>- and NO-Sensing Mechanism through the DevSR Two-Component System in *Mycobacterium smegmatis*." J. Bacteriol. **190**(20): 6795-6804.
- Lee, K. (1997). "A Structure-Based Activation Model of Phenol-Receptor Protein Interactions." Bull. Korean Chem. Soc. **18**: 18-23.
- Lee, K., M. Dudley, et al. (1992). "Mechanism of Activation of *Agrobacterium* Virulence Genes: Identification of Phenol-Binding Proteins 10.1073/pnas.89.18.8666." PNAS **89**(18): 8666-8670.
- Lee, K., Y.-L. Tzeng, et al. (1993). Possible evolutionary relationships in a signal transduction system. 18th International Symposium on the Chemistry of Natural Products. Strasbourg, France. **65**: 1241-1248.
- Lee, S.-Y., H. S. Cho, et al. (2001). "Crystal structure of an activated response regulator bound to its target." Nat Struct Mol Biol **8**(1): 52-56.
- Lee, Y.-W., U.-H. Ha, et al. (1998). "Characterization of an unusual sensor gene (*virA*) of *Agrobacterium*." Gene **210**(2): 307-314.

- Lee, Y.-W., S. Jin, et al. (1996). "The sensing of plant signal molecules by Agrobacterium: genetic evidence for direct recognition of phenolic inducers by the VirA protein." *Gene* **179**(1): 83-88.
- Lee, Y., S. Jin, et al. (1995). "Genetic Evidence for Direct Sensing of Phenolic Compounds by the VirA Protein of Agrobacterium tumefaciens." *PNAS* **92**(26): 12245-12249.
- Lee, Y., S. Jin, et al. (1996). "The sensing of plant signal molecules by Agrobacterium: genetic evidence for direct recognition of phenolic inducers by the VirA protein." *Gene* **179**(1): 83-88.
- Leroux, B., M. F. Yanofsky, et al. (1987). "Characterization of the virA locus of Agrobacterium tumefaciens: a transcriptional regulator and host range determinant." *EMBO J.* **6**(4): 849-856.
- Levdikov, V. M., E. Blagova, et al. (2006). "The Structure of CodY, a GTP- and Isoleucine-responsive Regulator of Stationary Phase and Virulence in Gram-positive Bacteria." *J. Biol. Chem.* **281**(16): 11366-11373.
- Li, J., S. G. Wolf, et al. (2005). "Exploring cargo transport mechanics in the type IV secretion systems." *Trends in Microbiology* **13**(7): 295-298.
- Li, L., Y. Jia, et al. (2002). "A global pH sensor: Agrobacterium sensor protein ChvG regulates acid-inducible genes on its two chromosomes and Ti plasmid." *PNAS* **99**(19): 12369-12374.
- Liebeton, K., A. Zonta, et al. (2000). "Directed evolution of an enantioselective lipase." *Chemistry & Biology* **7**(9): 709-718.
- Lin, Y. H., R. Gao, et al. (2008). Capturing the VirA/VirG TCS of Agrobacterium tumefaciens. *Bacterial Signal Transduction: Networks and Drug Targets*. New York, Springer. **631**: 161-177.
- Little, R. and R. Dixon (2003). "The Amino-terminal GAF Domain of Azotobacter vinelandii NifA Binds 2-Oxoglutarate to Resist Inhibition by NifL under Nitrogen-limiting Conditions." *J. Biol. Chem.* **278**(31): 28711-28718.
- Lohrke, S. M., S. Nechaev, et al. (1999). "Transcriptional Activation of Agrobacterium tumefaciens Virulence Gene Promoters in Escherichia coli Requires the A. tumefaciens rpoA Gene, Encoding the Alpha Subunit of RNA Polymerase." *J. Bacteriol.* **181**(15): 4533-4539.
- Lohrke, S. M., H. Yang, et al. (2001). "Reconstitution of Acetosyringone-Mediated Agrobacterium tumefaciens Virulence Gene Expression in the Heterologous Host Escherichia coli." *J. Bacteriol.* **183**(12): 3704-3711.
- Lupas, A., M. V. Dyke, et al. (1991). "Predicting Coiled Coils from Protein Sequences." *Science* **252**(5009): 1162-1164.
- Lupas, A. N., M. Gruber, et al. (2005). The Structure of [alpha]-Helical Coiled Coils. *Advances in Protein Chemistry*, Academic Press. **Volume 70**: 37-38.
- Madhusudan, J. Zapf, et al. (1997). "A Response Regulatory Protein with the Site of Phosphorylation Blocked by an Arginine Interaction: Crystal Structure of Spo0F from Bacillus subtilis<sup>†,‡</sup>." *Biochemistry* **36**(42): 12739-12745.
- Mantis, N. J. and S. C. Winans (1992). "The Agrobacterium tumefaciens vir gene transcriptional activator virG is transcriptionally induced by acid pH and other stress stimuli." *J. Bacteriol.* **174**(4): 1189-1196.

- Maresh, J., J. Zhang, et al. (2006). "The Innate Immunity of Maize and the Dynamic Chemical Strategies Regulating Two-Component Signal Transduction in *Agrobacterium tumefaciens*." ACS Chem. Biol. **1**(3): 165-175.
- Marina, A., C. Mott, et al. (2001). "Structural and Mutational Analysis of the PhoQ Histidine Kinase Catalytic Domain." Journal of Biological Chemistry **276**(44): 41182-41190.
- Marina, A., C. D. Waldburger, et al. (2005). "Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein." The EMBO Journal **24**: 4247-4259.
- Martinez, S. E., J. A. Beavo, et al. (2002). "GAF Domains: Two-Billion-Year-Old Molecular Switches that Bind Cyclic Nucleotides." Mol. Interv. **2**(5): 317-323.
- Martinez, S. E., S. Bruder, et al. (2005). "Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: Modes of ligand binding and dimerization." PNAS **102**(8): 3082-3087.
- McCullen, C. A. and A. N. Binns (2006). "Agrobacterium tumefaciens and Plant Cell Interactions and Activities Required for Interkingdom Macromolecular Transfer." Annual Review of Cell and Developmental Biology **22**(1): 101-127.
- Melchers, L. S., A. J. G. Regensburg-Tuïnk, et al. (1989). "Specificity of signal molecules in the activation of *Agrobacterium* virulence gene expression." Molecular Microbiology **3**(7): 969-977.
- Miller, J. H. (1972). Experiments in molecular genetics Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
- Moore, J. O. and W. A. Hendrickson (2009). "Structural Analysis of Sensor Domains from the TMAO-Responsive Histidine Kinase Receptor TorS." Structure (London, England : 1993) **17**(9): 1195-1204.
- Morris, G. M., D. S. Goodsell, et al. (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of Computational Chemistry **19**(14): 1639-1662.
- Mukhopadhyay, A., R. Gao, et al. (2004). "Integrating Input from Multiple Signals: The VirA/VirG Two-Component System of *Agrobacterium tumefaciens*." ChemBioChem **5**(11): 1535-1542.
- Nechushtai, R., H. Lammert, et al. (2011). "Allostery in the ferredoxin protein motif does not involve a conformational switch." Proceedings of the National Academy of Sciences **108**(6): 2240-2245.
- Ninfa, E. G., M. R. Atkinson, et al. (1993). "Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits." J. Bacteriol. **175**(21): 7024-7032.
- O'Shea, E., J. Klemm, et al. (1991). "X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil." Science **254**(5031): 539-544.
- Ouali, M. and R. D. King (2000). "Cascaded multiple classifiers for secondary structure prediction." Protein Science **9**(6): 1162-1176.
- Pan, S. Q., T. Charles, et al. (1993). "Prefomed Dimeric State of the Sensor Protein VirA is Involved in Plant- *Agrobacterium* Signal Transduction." PNAS **90**(21): 9939-9943.
- Parkinson, J. S. and E. C. Kofoid (1992). "Communication Modules in Bacterial Signaling Proteins." Annual Review of Genetics **26**(1): 71-112.

- Peach, M. L., G. L. Hazelbauer, et al. (2002). "Modeling the transmembrane domain of bacterial chemoreceptors." Protein Sci **11**(4): 912-923.
- Peng, W.-T., Y.-W. Lee, et al. (1998). "The Phenolic Recognition Profiles of the *Agrobacterium tumefaciens* VirA Protein Are Broadened by a High Level of the Sugar Binding Protein ChvE." J. Bacteriol. **180**(21): 5632-5638.
- Perutz, M. F., M. Paoli, et al. (1999). "Fix L, a haemoglobin that acts as an oxygen sensor: signalling mechanism and structural basis of its homology with PAS domains." Chemistry & Biology **6**(11): R291-R297.
- Raineri, D. M., M. I. Boulton, et al. (1993). "VirA, the plant-signal receptor, is responsible for the Ti plasmid-specific transfer of DNA to maize by *Agrobacterium*." Proceedings of the National Academy of Sciences of the United States of America **90**(8): 3549-3553.
- Robinson, V. L., D. R. Buckler, et al. (2000). "A tale of two components: a novel kinase and a regulatory switch." Nat Struct Mol Biol **7**(8): 626-633.
- Robinson, V. L., T. Wu, et al. (2003). "Structural Analysis of the Domain Interface in DrrB, a Response Regulator of the OmpR/PhoB Subfamily." J. Bacteriol. **185**(14): 4186-4194.
- Sardiwal, S., S. L. Kendall, et al. (2005). "A GAF Domain in the Hypoxia/NO-inducible *Mycobacterium tuberculosis* DosS Protein Binds Haem." Journal of Molecular Biology **353**(5): 929-936.
- Sevvana, M., V. Vijayan, et al. (2008). "A Ligand-Induced Switch in the Periplasmic Domain of Sensor Histidine Kinase CitA." Journal of Molecular Biology **377**(2): 512-523.
- Shimoda, N., A. Toyoda-Yamamoto, et al. (1993). "Genetic evidence for an interaction between the VirA sensor protein and the ChvE sugar-binding protein of *Agrobacterium*." Journal of Biological Chemistry **268**(35): 26552-26558.
- Shimoda, N., A. Toyoda-Yamamoto, et al. (1990). "Control of Expression of *Agrobacterium vir* Genes by Synergistic Actions of Phenolic Signal Molecules and Monosaccharides." PNAS **87**(17): 6684-6688.
- Shrawat, A. K. and H. Lörz (2006). "Agrobacterium-mediated transformation of cereals: a promising approach crossing barriers." Plant Biotechnology Journal **4**(6): 575-603.
- Singh, M., B. Berger, et al. (1998). "Computational learning reveals coiled coil-like motifs in histidine kinase linker domains." Proceedings of the National Academy of Sciences of the United States of America **95**(6): 2738-2743.
- Smith, E. F. and C. O. Townsend (1907). "A PLANT-TUMOR OF BACTERIAL ORIGIN." Science **25**(643): 671-673.
- Spencer, P. A. and G. H. N. Towers (1988). "Specificity of signal compounds detected by *Agrobacterium tumefaciens*." Phytochemistry **27**(9): 2781-2785.
- Stachel, S. E., G. An, et al. (1985). "A Tn3 LacZ Transposon for the Random Generation of Beta-Galactosidase Gene Fusions - Application to the Analysis of Gene-Expression in *Agrobacterium*." Embo Journal **4**(4): 891-898.
- Stachel, S. E., E. Messens, et al. (1985). "Identification of the signal molecules produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*." Nature **318**(6047): 624-629.



- Stachel, S. E., E. W. Nester, et al. (1986). "A plant cell factor induces *Agrobacterium tumefaciens* vir gene expression." Proceedings of the National Academy of Sciences of the United States of America **83**(2): 379-383.
- Stachel, S. E. and P. C. Zambryski (1986). "VirA and VirG control the plant-induced activation of the T-DNA transfer process of *A-tumefaciens*." Cell **46**: 325-333.
- Stephenson, K. and R. J. Lewis (2005). "Molecular insights into the initiation of sporulation in Gram-positive bacteria: new technologies for an old phenomenon." FEMS Microbiology Reviews **29**(2): 281-301.
- Stewart, V. and L.-L. Chen (2010). "The S Helix Mediates Signal Transmission as a HAMP Domain Coiled-Coil Extension in the NarX Nitrate Sensor from *Escherichia coli* K-12." J. Bacteriol. **192**(3): 734-745.
- Stock, A. M., E. Martinez-Hackert, et al. (1993). "Structure of the magnesium-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis." Biochemistry **32**(49): 13375-13380.
- Stock, A. M., V. L. Robinson, et al. (2000). "TWO-COMPONENT SIGNAL TRANSDUCTION." Annu. Rev. Biochem. **69**: 183-215.
- Szurmant, H., R. A. White, et al. (2007). "Sensor complexes regulating two-component signal transduction." Current Opinion in Structural Biology **17**(6): 706-715.
- Tanaka, T., S. K. Saha, et al. (1998). "NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ." Nature **396**(6706): 88-92.
- Tao, W., C. L. Malone, et al. (2002). "A cytoplasmic coiled-coil domain is required for histidine kinase activity of the yeast osmosensor, SLN1." Molecular Microbiology **43**: 459-473.
- Taylor, B. L. and I. B. Zhulin (1999). "PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light." Microbiol. Mol. Biol. Rev. **63**(2): 479-506.
- Tetsch, L. and K. Jung (2009). "The regulatory interplay between membrane-integrated sensors and transport proteins in bacteria." Molecular Microbiology **73**(6): 982-991.
- Tomomori, C., T. Tanaka, et al. (1999). "Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ." Nat Struct Mol Biol **6**(8): 729-734.
- Toro-Roman, A., T. R. Mack, et al. (2005). "Structural Analysis and Solution Studies of the Activated Regulatory Domain of the Response Regulator ArcA: A Symmetric Dimer Mediated by the [alpha]4-[beta]5-[alpha]5 Face." Journal of Molecular Biology **349**(1): 11-26.
- Turk, S. C. H. J., E. W. Nester, et al. (1993). "The virA promoter is a host-range determinant in *Agrobacterium tumefaciens*." Molecular Microbiology **7**(5): 719-724.
- Tzfira, T. and V. Citovsky (2002). "Partners-in-infection: host proteins involved in the transformation of plant cells by *Agrobacterium*." Trends in Cell Biology **12**(3): 121-129.
- Utsumi, R., R. E. Brissette, et al. (1989). "Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate." Science **245**(4923): 1246-1249.

- Utsumi, R., S. Yamada, et al. (2008). Structural Basis of the Signal Transduction in the Two-Component System. Bacterial Signal Transduction: Networks and Drug Targets, Springer New York. **631**: 22-39.
- Varughese, K. I. (2002). "Molecular recognition of bacterial phosphorelay proteins." Current Opinion in Microbiology **5**(2): 142-148.
- Wadhams, G. H. and J. P. Armitage (2004). "MAKING SENSE OF IT ALL: BACTERIAL CHEMOTAXIS." Nature Reviews Molecular Cell Biology **5**(12): 1024-1037.
- Wang, Y., R. Gao, et al. (2002). "Ratcheting Up *vir* Gene Expression in *Agrobacterium tumefaciens*: Coiled Coils in Histidine Kinase Signal Transduction." ChemBioChem **3**(4): 311-317.
- Wang, Y., A. Mukhopadhyay, et al. (2000). "Construction of an efficient expression system for *Agrobacterium tumefaciens* based on the coliphage T5 promoter." Gene **242**(1-2): 105-114.
- Watson, B., T. C. Currier, et al. (1975). "Plasmid required for virulence of *Agrobacterium tumefaciens*." J. Bacteriol. **123**(1): 255-264.
- West, A. H. and A. M. Stock (2001). "Histidine kinases and response regulator proteins in two-component signaling systems." Trends in Biochemical Sciences **26**(6): 369-376.
- Winans, S. C. (1990). "Transcriptional induction of an *Agrobacterium* regulatory gene at tandem promoters by plant-released phenolic compounds, phosphate starvation, and acidic growth media." J. Bacteriol. **172**(5): 2433-2438.
- Wise, A. A., F. Fang, et al. (2010). "The receiver domain of hybrid histidine kinase VirA: an enhancing factor for *vir* gene expression in *Agrobacterium tumefaciens*." J. Bacteriol. **192**: 1534-1542.
- Yang, J. T., C. S. Wu, et al. (1986). "Calculation of protein conformation from circular dichroism." Methods in enzymology **130**: 208-269.
- Yang, X., J. Kuk, et al. (2008). "Crystal structure of *Pseudomonas aeruginosa* bacteriophytochrome: Photoconversion and signal transduction." Proceedings of the National Academy of Sciences **105**(38): 14715-14720.
- Yuan, Z.-C., P. Liu, et al. (2007). "Transcriptome profiling and functional analysis of *Agrobacterium tumefaciens* reveals a general conserved response to acidic conditions (pH5.5) and a complex acid-mediated signaling involved in *Agrobacterium*-plant interactions." J. Bacteriol.: JB.01387-01307.
- Zapf, J., U. Sen, et al. (2000). "A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction." Structure (London, England : 1993) **8**(8): 851-862.
- Zhang, J., L. Boone, et al. (2000). "At the maize/*Agrobacterium* interface: natural factors limiting host transformation." Chemistry & Biology **7**(8): 611-621.
- Zhu, J., P. M. Oger, et al. (2000). "The Bases of Crown Gall Tumorigenesis." J. Bacteriol. **182**(14): 3885-3895.
- Zhu, Y. and M. Inouye (2003). "Analysis of the Role of the EnvZ Linker Region in Signal Transduction Using a Chimeric Tar/EnvZ Receptor Protein, Tez1." Journal of Biological Chemistry **278**(25): 22812-22819.

