



UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES

CENTRO DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MORFOLOGÍA

TESIS

ESTUDIO DE SISTEMAS DE DOS COMPONENTES EN
ACTINOBACILLUS PLEUROPNEUMONIAE

CANDIDATO

IBq. Eduardo Mauricio Martín Álvarez Tostado

PARA LA OBTENCIÓN DEL GRADO DOCTOR EN CIENCIAS
BIOLÓGICAS

DIRECTOR

Dra. Alma Lilian Guerrero Barrera

MIEMBROS DEL COMITÉ TUTORAL

Mario Jacques, PhD

Dr. Ricardo Oropeza Navarro

Aguascalientes, Ags.; 9 de Junio de 2026

MTRO. GUILLERMO DOMÍNGUEZ AGUILAR
DECANO (A) DEL CENTRO DE CIENCIAS BÁSICAS

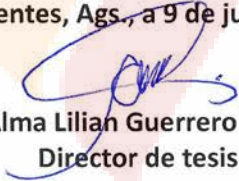
P R E S E N T E

Por medio del presente como **DIRECTOR** designado del estudiante **EDUARDO MAURICIO MARTIN ÁLVAREZ TOSTADO** con ID 23653 quien realizó la tesis titulada: **ESTUDIO DE SISTEMAS DE DOS COMPONENTES EN *Actinobacillus pleuropneumoniae***, un trabajo propio, innovador, relevante e inédito y con fundamento en la fracción IX del Artículo 43 del Reglamento General de Posgrados, doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE
"Se Lumen Proferre"

Aguascalientes, Ags., a 9 de junio de 2026.


Dr. Alma Lillian Guerrero Barrera
Director de tesis

c.c.p.- Interesado
c.c.p.- Coordinación del Programa de Posgrado

CARTA DE VOTO APROBATORIO

MTRO. GUILLERMO DOMÍNGUEZ AGUILAR
DECANO (A) DEL CENTRO DE CIENCIAS BÁSICAS

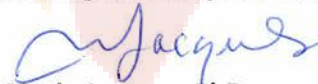
P R E S E N T E

Por medio del presente como **ASESOR** designado del estudiante **EDUARDO MAURICIO MARTIN ÁLVAREZ TOSTADO** con ID 23653 quien realizó la tesis titulada: **ESTUDIO DE SISTEMAS DE DOS COMPONENTES EN *Actinobacillus pleuropneumoniae***, un trabajo propio, innovador, relevante e inédito y con fundamento en la fracción IX del Artículo 43 del Reglamento General de Posgrados, doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE
"Se Lumen Proferre"

Aguascalientes, Ags., a 9 de junio de 2026.


Mario Jacques, PhD
Asesor de tesis

c.c.p.- Interesado
c.c.p.- Coordinación del Programa de Posgrado

CARTA DE VOTO APROBATORIO

MTRO. GUILLERMO DOMÍNGUEZ AGUILAR
DECANO (A) DEL CENTRO DE CIENCIAS BÁSICAS

P R E S E N T E

Por medio del presente como **ASESOR** designado del estudiante **EDUARDO MAURICIO MARTIN ÁLVAREZ TOSTADO** con ID 23653 quien realizó la tesis titulada: **ESTUDIO DE SISTEMAS DE DOS COMPONENTES EN *Actinobacillus pleuropneumoniae***, un trabajo propio, innovador, relevante e inédito y con fundamento en la fracción IX del Artículo 43 del Reglamento General de Posgrados, doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE
"Se Lumen Proferre"

Aguascalientes, Ags., a 9 de junio de 2026.

Dr. Ricardo Oropeza Navarro
Asesor de tesis

c.c.p.- Interesado
c.c.p.- Coordinación del Programa de Posgrado

Fecha de dictaminación (dd/mm/aaaa): 27/05/2026

NOMBRE: Eduardo Mauricio Martín Álvarez Tostado ID 23653

PROGRAMA: Doctorado en Ciencias Biológicas LGAC (del posgrado): Toxicología y Bioingeniería Ambiental

MODALIDAD DEL PROYECTO DE GRADO: Tesis (X) *Tesis por artículos científicos () **Tesis por Patente () Trabajo Práctico ()

TÍTULO: ESTUDIO DE SISTEMAS DE DOS COMPONENTES EN *Actinobacillus pleuropneumoniae*

IMPACTO SOCIAL (señalar el impacto logrado): El análisis de los sistemas de dos componentes en bacterias es de suma importancia para comprender y combatir la virulencia bacteriana. Este análisis sirve como base para el diseño de nuevos antibióticos.

INDICAR SEGÚN CORRESPONDA: SI, NO, NA (No Aplica)

Elementos para la revisión académica del trabajo de tesis o trabajo práctico:	
SI	El trabajo es congruente con las LGAC del programa de posgrado
SI	La problemática fue abordada desde un enfoque multidisciplinario
SI	Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI	Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI	Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
SI	El trabajo demuestra más de una aportación original al conocimiento de su área
SI	Las aportaciones responden a los problemas prioritarios del país
NO	Generó transferencia del conocimiento o tecnológica
SI	Cumple con la ética para la investigación (reporte de la herramienta antiplagio)
El egresado cumple con lo siguiente:	
SI	Cumple con lo señalado por el Reglamento General de Posgrados
SI	Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc.)
SI	Cuenta con los votos aprobatorios del comité tutorial
N.A.	Cuenta con la carta de satisfacción del Usuario (En caso de que corresponda)
SI	Coincide con el título y objetivo registrado
SI	Tiene congruencia con cuerpos académicos
SI	Tiene el CVU de la SECHTI actualizado
SI	Tiene el o los artículos aceptados o publicados y cumple con los requisitos institucionales (en caso de que proceda)
*En caso de Tesis por artículos científicos publicados (completar solo si la tesis fue por artículos)	
	Aceptación o Publicación de los artículos en revistas indexadas de alto impacto según el nivel del programa
	El (la) estudiante es el primer autor(a)
	El (la) autor(a) de correspondencia es el Director (a) del Núcleo Académico
	En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
	Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
**En caso de Tesis por Patente	
	Cuenta con la evidencia de solicitud de patente en el Departamento de Investigación (anexarla al presente formato)

Con base en estos criterios, se autoriza continuar con los trámites de titulación y programación del examen de grado:

SI X
No

Elaboró:

FIRMAS

*NOMBRE Y FIRMA DEL(LA) CONSEJERO(A) SEGÚN LA LGAC DE ADSCRIPCIÓN:

Dr. Daniel Cervantes García

* En caso de conflicto de intereses, firmará un revisor miembro del NA de la LGAC correspondiente distinto al director o miembro del comité tutorial, asignado por el Decano.

NOMBRE Y FIRMA DEL COORDINADOR DE POSGRADO:

Dra. Lucía Isabel Chávez Ortiz

Revisó:

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:

Dra. Iliana Ernestina Medina Ramírez

Autorizó:

NOMBRE Y FIRMA DEL DECANO:

Mtro. Guillermo Domínguez Boule

Nota: procede el trámite para el Depto. de Apoyo al Posgrado

En cumplimiento con el Art. 24 fracción V del Reglamento General de Posgrado, que a la letra señala entre las funciones del Consejo Académico: Proponer criterios y mecanismos de selección, permanencia, egreso y titulación de estudiantes para asegurar la eficiencia terminal y la titulación y el Art. 28 fracción IX, atender, asesorar y dar el seguimiento del estudiantado desde su ingreso hasta su titulación.



Article

From Interfaces to Networks: Energetic Control of Specificity in Bacterial Two-Component Systems

Eduardo M. Martin ¹, Alma L. Guerrero-Barrera ^{1,*} , F. Javier Avelar-Gonzalez ², Rogelio Salinas-Gutierrez ³  and Mario Jacques ⁴

¹ Laboratorio de Biología Celular y Tisular, Departamento de Morfología, Universidad Autónoma de Aguascalientes, Aguascalientes 20100, Mexico; mauricio.martin@edu.uaa.mx

² Laboratorio de Estudios Ambientales, Departamento de Fisiología y Farmacología, Universidad Autónoma de Aguascalientes, Aguascalientes 20100, Mexico; javier.avelar@edu.uaa.mx

³ Departamento de Estadística, Universidad Autónoma de Aguascalientes, Aguascalientes 20100, Mexico; rogelio.salinas@edu.uaa.mx

⁴ Groupe de Recherche sur les Maladies Infectieuses du Porc, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC J2S 2M2, Canada; mario.jacques@umontreal.ca

* Correspondence: lilian.guerrero@edu.uaa.mx

Abstract

Bacterial two-component systems (TCSs) mediate environmental sensing and adaptive responses through signal transduction between histidine kinases (HKs) and response regulators (RRs), thereby regulating biochemical processes essential for survival and, in pathogenic species, infection. How signaling specificity and insulation are maintained in organisms encoding multiple paralogous two-component systems remains an open question. Here, we investigate specificity in the *Actinobacillus pleuropneumoniae* TCS signaling network using an integrated computational framework that combines coevolutionary analysis, structural modeling, molecular dynamics simulations, and free-energy calculations. We show that cognate HK-RR recognition is established locally through clusters of coevolving interface residues, termed the orthologue interface specificity core (OISC), which mediate symmetric molecular recognition at individual interaction interfaces. However, interface-level recognition alone is insufficient to explain signaling fidelity across the network. Instead, system-wide specificity and pathway insulation emerge in this network from asymmetric energetic discrimination among cognate and non-cognate interactions across the ensemble of paralogous interfaces. Graded free-energy profiles reveal that broadly compatible interfaces can coexist with robust signaling insulation, reconciling interface promiscuity with stable network organization. Together, these findings support a two-tiered model for the TCS network analyzed here, in which symmetric interface constraints enable cognate recognition, while asymmetric network-level energetics govern signaling specificity. This framework may extend to other paralogous TCS networks.

Keywords: two-component systems; signaling specificity; molecular recognition; protein–protein interaction; binding free energy; molecular dynamics; signaling network architecture; *Actinobacillus pleuropneumoniae*; virulence regulation



Academic Editor: Rainer Breitling

Received: 17 April 2026

Revised: 18 May 2026

Accepted: 18 May 2026

Published: 25 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Actinobacillus pleuropneumoniae is a Gram-negative bacterium. This encapsulated coccobacillus, which is hemolytic, facultative anaerobic, and fermentative, causes porcine pleuropneumonia, a highly contagious disease [1]. Porcine pleuropneumonia is characterized by fibrinohemorrhagic necrotizing bronchopneumonia and fibrinous pleuritis [2].



Review

Two-Component Systems in *Pasteurellaceae* and Their Roles in Virulence

Eduardo M. Martin ¹, Alma L. Guerrero-Barrera ^{1,*} and F. Javier Avelar-Gonzalez ²

¹ Laboratorio de Biología Celular y Tisular, Departamento de Morfología, Universidad Autónoma de Aguascalientes, Aguascalientes 20131, Mexico; mauricio.martin@edu.uaa.mx

² Laboratorio de Estudios Ambientales, Departamento de Fisiología y Farmacología, Universidad Autónoma de Aguascalientes, Aguascalientes 20131, Mexico; javier.avelar@edu.uaa.mx

* Correspondence: lilian.guerrero@edu.uaa.mx

Simple Summary

Bacteria survive by detecting changes around them and adjusting their behavior accordingly. To do this, they use protein pairs that act like on/off switches, known as two-component systems. When activated by a specific signal, these systems trigger a response that helps bacteria tolerate stress or cause infection. This review examines how these systems function in members of the *Pasteurellaceae* family and how even closely related species can respond very differently to similar signals. The review summarizes those systems that are best understood, where knowledge gaps remain, and how differences in gene regulation may explain variations in virulence. New computational approaches—including artificial intelligence, molecular modeling, molecular dynamics, and network analysis—that can predict how these protein switches interact and control bacterial behavior are also highlighted. Understanding these signaling systems may help identify new drug targets and improve strategies to prevent or treat infections caused by these and other pathogens.

Abstract

Two-component systems (TCSs) are widespread in bacteria and archaea, with only limited presence in eukaryotes. These signaling mechanisms detect environmental changes and adjust gene expression to survive and adapt. In this review, TCSs were examined within the *Pasteurellaceae* family, focusing on how closely related organisms employ similar systems to regulate infections and stress responses. Comparative analysis revealed that homologous TCSs can differ markedly in the signals they detect and in the genes or virulence factors they control, underscoring the functional diversity that has evolved even within this family. Inconsistencies in nomenclature across studies are also identified, which complicate data integration and cross-species comparisons. Given these challenges, the need for unified naming conventions and broader, system-level analyses is highlighted. It is further proposed that emerging computational tools—including molecular modeling, molecular dynamics, and neural network-based analyses—offer powerful strategies to predict signaling interactions, identify conserved regulatory models, and clarify how these systems contribute to bacterial adaptation and pathogenicity.

Keywords: *Pasteurellaceae*; two-component systems; signal transduction; transcriptional regulation; virulence regulation; comparative genomics; regulatory networks; computational modeling



Academic Editor: Yafeng Qiu

Received: 17 October 2025

Revised: 21 November 2025

Accepted: 26 November 2025

Published: 29 November 2025

Citation: Martin, E.M.; Guerrero-Barrera, A.L.; Avelar-Gonzalez, F.J. Two-Component Systems in *Pasteurellaceae* and Their Roles in Virulence. *Vet. Sci.* **2025**, *12*, 1140. <https://doi.org/10.3390/vetsci12121140>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

AGRADECIMIENTOS

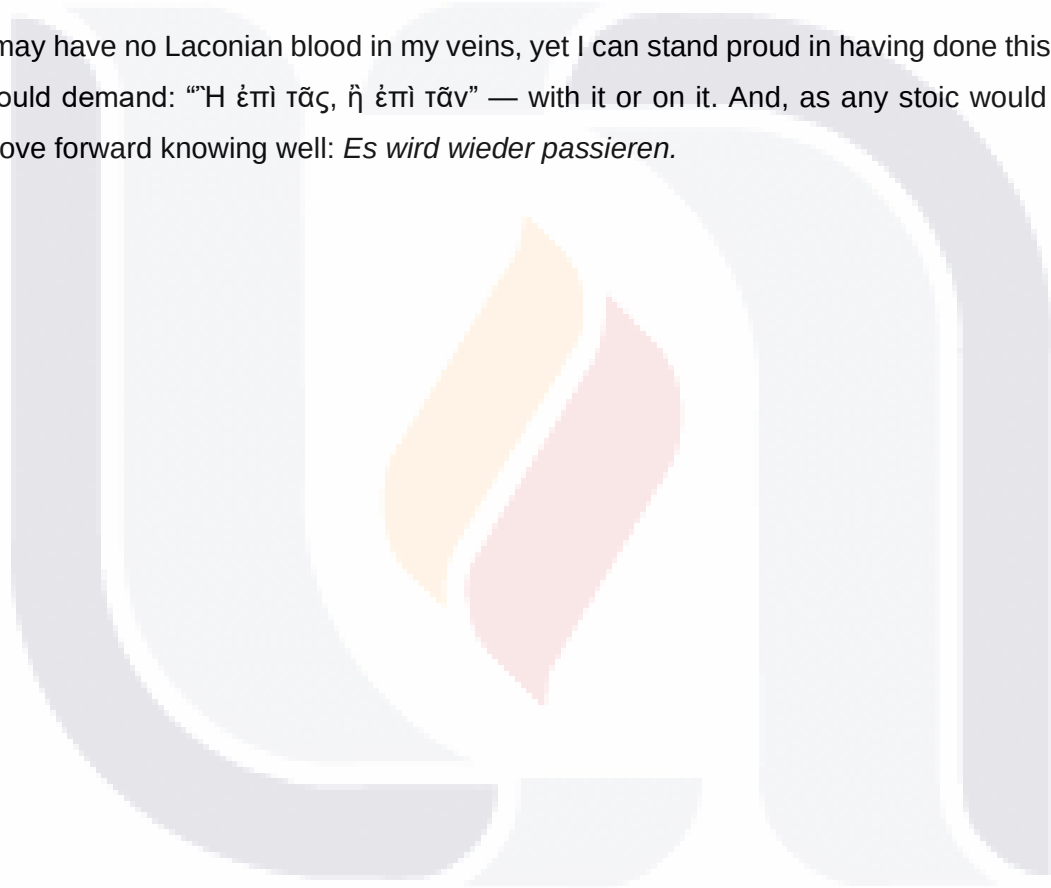
A la Universidad Autónoma de Aguascalientes y al SECIHTI por la beca No. 297145 ya que sin ninguna de estas anteriores esto no sería posible.

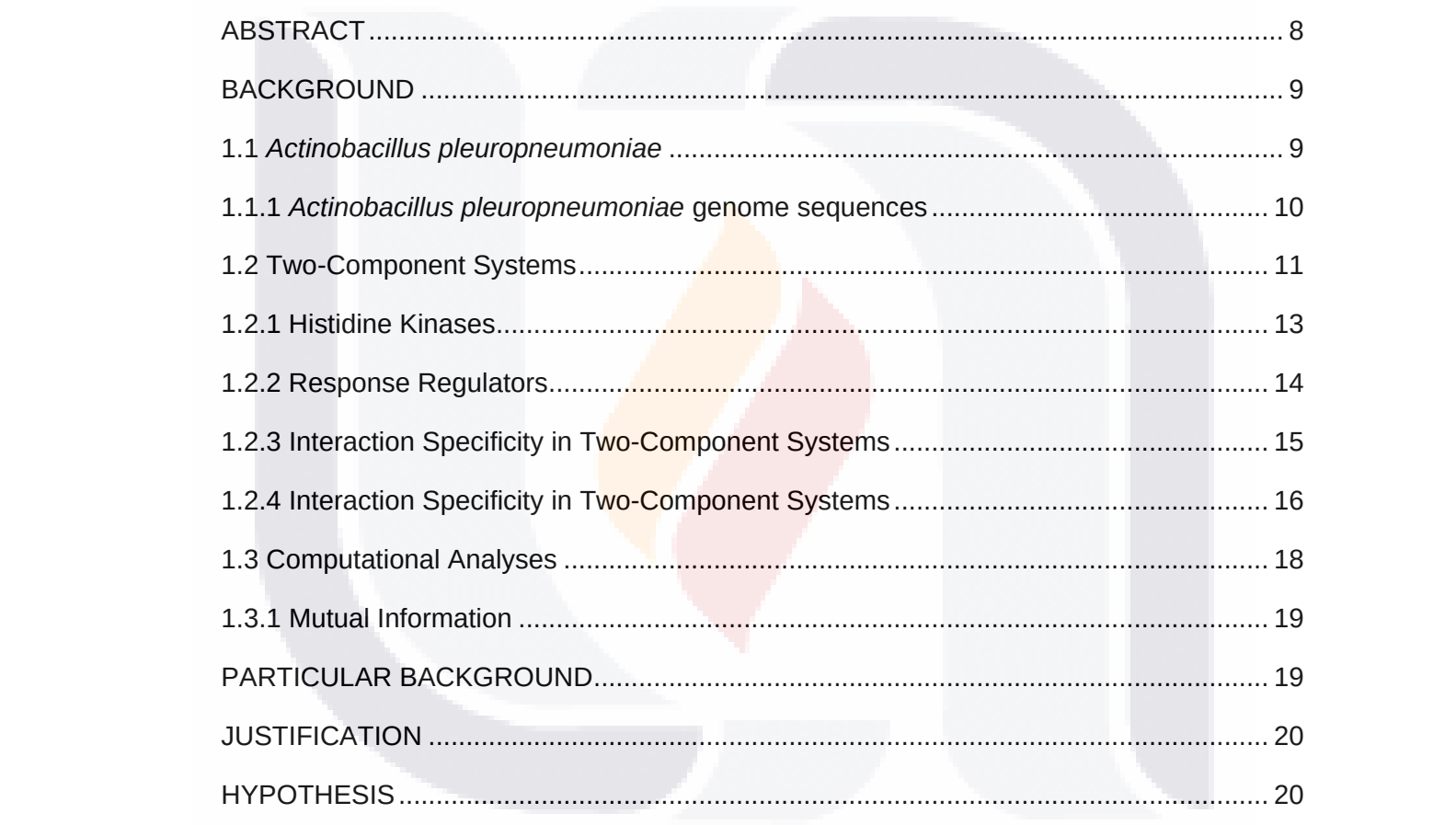


DEDICATORIA

To everyone who accompanied me on this journey, who watched me wander through the Möbius labyrinth of my own making, carrying the boulder endlessly. If you cannot picture me as a happy Sisyphus, you can at least be certain that every moment was borne with pride.

I may have no Laconian blood in my veins, yet I can stand proud in having done this as they would demand: “Ἡ ἐπὶ τῷ, ἢ ἐπὶ τῷ” — with it or on it. And, as any stoic would know, I move forward knowing well: *Es wird wieder passieren.*





INDEX

INDEX 1

TABLES INDEX 3

GRAPHICS INDEX 4

LIST OF ABBREVIATIONS..... 6

RESUMEN 7

ABSTRACT 8

BACKGROUND 9

1.1 *Actinobacillus pleuropneumoniae* 9

1.1.1 *Actinobacillus pleuropneumoniae* genome sequences 10

1.2 Two-Component Systems 11

1.2.1 Histidine Kinases..... 13

1.2.2 Response Regulators..... 14

1.2.3 Interaction Specificity in Two-Component Systems 15

1.2.4 Interaction Specificity in Two-Component Systems 16

1.3 Computational Analyses 18

1.3.1 Mutual Information 19

PARTICULAR BACKGROUND..... 19

JUSTIFICATION 20

HYPOTHESIS 20

OBJECTIVE 20

General Objective 20

Particular Objectives 20

MATERIALS AND METHODS 21

2.1 Identification of putative two-component systems proteins in *Actinobacillus pleuropneumoniae* 21

2.2 Identification of genomic context of two-component system putative proteins of <i>Actinobacillus pleuropneumoniae</i>	22
2.3 Computational analyses.....	22
2.3.1 Two-Component Systems paralogous families' datasets	22
2.3.2 Mutual information with average product correction (<i>Mip</i>)	23
2.3.3 Joint column occupancy and weighted <i>Mip</i>	31
2.3.4 Tridimensional mapping of specificity-determining residue positions	32
2.3.5 Protein-protein interface amino acids.....	33
2.3.6 Prototypical interaction interface's distance among amino acids.....	33
2.3.7 Amd PseudoCode	34
2.3.8 Specificity determining co-evolving interface matrices.....	37
RESULTS	38
3.1 Putative two-component system proteins	38
3.2 Prototypical interaction interface	44
3.3 Amino acids determining specificity in non-Hybrid Two-Component Systems interface	47
3.3.1 Specificity amino acids SDMCEII	47
3.3.2 Interface coupling index	56
DISCUSSION.....	59
4.1 Pair wise testing	60
4.2 Interaction interfaces' discriminating components	62
4.3 System-wide discrimination.....	63
4.4 TCSs' discrimination in <i>A. pleuropneumoniae</i>	67
CONCLUSIONS.....	70
BIBLIOGRAPHY	72

TABLES INDEX

Table 3.1.1 Two-Component Systems identified in *Actinobacillus pleuropneumoniae* 42

Table 3.2.1 Prototypical interaction interface’s amino acid positions 46

Table 3.3.1.1 Mlp mean values of orthologue interface specificity positions 54

Table 3.3.1.2 Orthologue interface specificity core clusters in *A. pleuropneumoniae* 55

Table 3.3.2.1 OISCC amino acid pairs 58



GRAPHICS INDEX

Figure 1.2.1 Two-component system model	12
Equation 2.3.2.1 Mutual information with average product correction	24
Equation 2.3.2.2 Mutual information	24
Equation 2.3.2.3 Shannon's Entropy	24
Equation 2.3.2.4 Joint Shannon's Entropy	24
Equation 2.3.2.5 Average product correction	24
Equation 2.3.2.6 Mean mutual information	24
Equation 2.3.2.7 Overall mean mutual information	24
Pseudo-Code 2.3.2.1 Mlp pseudo-code	24
Equation 2.3.3.1 Weighted mutual information with average product correction	32
Pseudo-Code 2.3.3.1 wMlp pseudo-code	32
Equation 2.3.6.1 Adjusted minimal distance	33
Equation 2.3.6.2 Interface coupling index	34
Pseudo-Code 2.3.7 Amd pseudo-code	34
Equation 2.3.8.1 Matrices absolute distance value	37
Equation 2.3.8.2 Matrices mean value	37
Figure 3.1.1 <i>arcA</i> Gene Context	38
Figure 3.1.2 <i>arcB</i> Gene Context	38
Figure 3.1.3 <i>cpxRcpxA</i> Gene Context	38
Figure 3.1.4 <i>narP</i> Gene Context	39
Figure 3.1.5 <i>narQ</i> Gene Context	39
Figure 3.1.6 <i>phoBphoR</i> Gene Context	39
Figure 3.1.7 <i>qseBqseC</i> Gene Context	39
Figure 3.1.8 Two-Component Systems Architecture in <i>Actinobacillus pleuropneumoniae</i>	40
Figure 3.1.9 CpxR gapped proteins from <i>Actinobacillus pleuropneumoniae</i>	41

Figure 3.2.1 Histidine kinase-Response regulator interaction complex	44
Figure 3.2.2 Histidine kinase-Response regulator secondary interaction interface	46
Figure 3.3.1 Mlp and wMlp dot blots.....	47
Figure 3.2.1.1 Histidine kinase-Response regulator Interaction Interface Mlp heatmaps .	48
Figure 3.3.1.2 Sequence logos from parental TCSs MSAs Mlp matrices	49
Figure 3.3.1.3 Sequence logos from non-hybrid TCSs MSAs of <i>A. pleuropneumoniae</i>	49
Figure 3.3.1.4 PhoRB::CpxA Mlp matrix mean values distribution .	50
Figure 3.3.1.5 PhoRB::NarQP Mlp matrix mean values distribution	50
Figure 3.3.1.6 PhoRB::QseCB Mlp matrix mean values distribution	50
Figure 3.3.1.7 CpxAR::NarQP Mlp matrix mean values distribution	51
Figure 3.3.1.8 CpxAR::QseCB Mlp matrix mean values distribution	51
Figure 3.3.1.9 NarQP::QseCB Mlp matrix mean values distribution	51
Figure 3.3.1.10 Pairwise undirected graphs of interaction interfaces' residue network.....	52
Figure 3.3.1.11 Pairwise maximum spanning tree interaction interfaces' residue network	53
Figure 3.3.1.12 OISCC spatial localization	56
Figure 3.3.2.1 ICI vs amd dot blot	57
Figure 3.3.2.2 OISCC spatial localization with ICI	57
Figure 3.3.2.3 Sequence logo from parental TCSs' MSAs	59
Figure 3.3.2.4 Sequence logo from non-hybrid <i>A. pleuropneumoniae</i> 's TCSs MSAs	59

LIST OF ABBREVIATIONS

NAD: Nicotinamide adenine dinucleotide

CDS: Coding sequence

Mbp: Mega base pairs

TCS: Two-component system

HK: Histidine Kinase

RR: Response Regulator

His: Histidine

ATP: Adenosine triphosphate

ADP: Adenosine diphosphate

Asp: Aspartate

REC: Receiver Domain

DHp: Dimerization and histidine phosphotransfer domain

CA: Catalytic and Phosphorylation Domain

HPT: Histidine Phosphotransfer Domain

App: Actinobacillus pleuropneumoniae

ORF: Open Reading Frame

Pi: Inorganic Phosphate

EPI: Epinephrine

NE: Norepinephrine

MSA: Multiple Sequence Alignment

MI: Mutual information

MIp: Mutual Information with Average Product Correction

wMIp: Weighted mutual Information with Average Product Correction

SDMCEII: Specificity Determining Matrices of Co Evolving Interaction Interface

ICI: Interface compatibility index

PII: Prototypical interaction interface

amd: Adjusted minimal distance

OISCC: Orthologue interface specificity core cluster

ISAC: Interface specificity accessory cluster

ISAR: Interface specificity accessory residues

RESUMEN

La transducción de señales de los sistemas de dos componentes depende del reconocimiento exacto entre proteínas cognadas, lo cual se realiza de manera selectiva por dichos sustratos cognados, mientras que la comunicación cruzada con proteínas no cognadas es prácticamente nula. Sin embargo, las propiedades esenciales que explican dicha especificidad no han sido establecidas de manera exhaustiva. Para abordar este problema, el presente trabajo evalúa la participación de los residuos de la interfase proteína-proteína en la discriminación. A través del análisis de las secuencias de aminoácidos de los cuatro sistemas de dos componentes prototípicos codificados en *Actinobacillus pleuropneumoniae*, se emplearon alineamientos múltiples de secuencias, información mutua, datos estructurales y el índice de acoplamiento de la interfase para evaluar el reconocimiento molecular. Se utilizaron más de dos mil secuencias homólogas individuales con el fin de incluir la mayor diversidad posible de secuencias pertenecientes al dominio Bacteria.

Se identificaron cuatro grupos distintos de residuos determinantes de especificidad para todos los sistemas evaluados, denominados “núcleo de especificidad de la interfase ortóloga”. La capacidad global de discriminación de los sistemas de dos componentes depende de la composición de estos grupos, compuestos respectivamente por 4, 2, 5 y 3 pares de residuos en los sistemas CpxAR, NarQP, PhoRB y QseCB. Los pares de residuos se encuentran espacialmente próximos; la forma y composición de los grupos son específicas de cada sistema, y sus interfaces de interacción presentan un traslape mínimo entre sí. De esta manera, se definió la composición de las interfaces de interacción de la red de sistemas de dos componentes en *A. pleuropneumoniae* y se describieron sus elementos discriminatorios.

En conjunto, se propone que el reconocimiento molecular depende de componentes concretos de la interfase de interacción, y que la especificidad emergente y el consecuente aislamiento entre componentes cognados y no cognados se derivan de ello. Esto refleja la capacidad general de estos sistemas para discriminar componentes altamente similares mediante reconocimiento molecular.

ABSTRACT

Two-component systems' signal transduction depends on the exact recognition between cognate proteins, which is selectively done by cognate substrates and cross-talk to non-cognate proteins is negligible. However, the fine underlying properties explaining specificity have not been thoroughly established. To solve this, the present work evaluates the involvement of protein-protein interface residues in discrimination. Through analyses of the amino acid sequences from the four prototypical two-component systems coded in *Actinobacillus pleuropneumoniae*. Multiple sequence alignments, mutual information, structural data, and the interface coupling index, were used to evaluate molecular recognition. Over two thousand individual homologue sequences were used, to include the broadest diversity of sequences belonging to the bacteria domain. Four different clusters of specificity determining residues were found for all evaluated systems, named Orthologue interface specificity core. The gross system-wide discrimination capability of two component systems relies on the composition of these cluster and are composed of 4, 2, 5, and 3 residue pairs, from systems CpxAR, NarQP, PhoRB, and QseCB, respectively. The residue pairs in these are spatially nearby, the shape and composition of the clusters are system-specific and its interaction interfaces have minimal overlap among them. The interaction interface composition of the two-component systems network in *A. pleuropneumoniae* was defined and their discriminatory components described. Overall, here is suggested how molecular recognition relies on concrete components from the interaction interface, and the emergent specificity and resultant insulation of cognate and non-cognate components, respectively. Which encompasses the system's overall capability to discriminate highly similar components, through molecular recognition.

BACKGROUND

1.1 *Actinobacillus pleuropneumoniae*

Actinobacillus pleuropneumoniae is a Gram-negative, fermentative, hemolytic, biofilm forming, facultative anaerobic encapsulated coccobacillus, which is the etiological agent of porcine pleuropneumonia. This disease is highly contagious and characterized by fibrinohemorrhagic necrotizing bronchopneumonia and fibrinous pleuritis, ranging from peracute to chronic. The chronic condition has a high prevalence resulting in growth impairment and increased susceptibility to secondary infections. Chronic infected individuals serve as carriers and some animals can develop subclinical infections, both of them responsible for transmission of the microorganism. The disease has been described worldwide, having serious impact on economy, ecology and animal welfare in domestic pig industrial farming (Bossé et al., 2002; Chiers, Waele, Pasmans, Ducatelle, & Haesebrouck, 2010; Dubreuil, Jacques, Mittal, & Gottschalk, 2000).

Actinobacillus pleuropneumoniae is differentiated in two biotypes, Biotype 1 strains require nicotinamide adenine dinucleotide (NAD), whereas Biotype 2 strains synthesize NAD. Furthermore, Biotype 1 is divided in 12 serotypes as well as Biotype 2 has 6 serotypes, based on surface polysaccharide antigens. Serotypes 1 and 5 are further differentiated into 1a, 1b, 5a and 5b, based on minor differences in polysaccharide structures. Thirteen serotypes have been sequenced so far, from those serotypes 1, 3, 5b, 7, and 8 are completely sequenced. Draft genome sequences for the remaining serotypes are available with the exception of serotypes 14 and 15, from a total of 16 recognized up to this date (Bossé et al., 2016; Foote et al., 2007; G. Li, Xie, Zhang, & Wang, 2012; Pereira et al., 2015; Sárközi, Makrai, & Fodor, 2015; Xu et al., 2010, 2008; Zhan, Angen, Hedegaard, Bendixen, & Panitz, 2010).

All serotypes are capable of developing porcine pleuropneumonia, although some serotypes are more virulent than others. The virulence of this pathogen is accomplished with the help of many factors, including exotoxins, endotoxin, capsule polysaccharides, adhesins, and outer membrane proteins, as well as iron acquisition systems. These variations appear to correlate with the production of diverse combinations of Apx toxins, but not solely; with most virulent serotypes producing both ApxI and ApxII. The four pore-forming exotoxins of *A. pleuropneumoniae*, called Apx, are cytolytic and/or hemolytic. Serotypes 1, 5, 9, and 11 are known to be especially virulent, and all express both ApxI and ApxII (Auger et al., 2009; Bossé et al., 2002; Chiers et al., 2010).

Besides *A. pleuropneumoniae* pathogenesis is complex, involving different virulence factors, among which are related to host colonization and establishment of infection, essential nutrients acquisition, metabolic adaptation and persistence within the host. Some of these aforementioned virulence factors, have been described to be regulated in *A. pleuropneumoniae* by two component systems; such as, the histidine kinases CpxA, PhoR and QseC, and the response regulators ArcA and QseB (Auger et al., 2009; Baltes et al., 2005; Bossé et al., 2002; Buettner, Bendallah, et al., 2008; Buettner, Maas, & Gerlach, 2008; Chiers et al., 2010; Deslandes et al., 2010; Liu et al., 2015).

1.1.1 *Actinobacillus pleuropneumoniae* genome sequences

The first complete genome sequence for *A. pleuropneumoniae* was sequence from serovar 5b L20 with 2,274,482bp in length and containing 2,012 putative open reading frames. No plasmid DNA was found in L20, although plasmids have been described for other strains of this species. The chromosome of L20 does not contain duplicate regions. The average GC content is 41.3%, similar to those of other members of the *Pasteurellaceae*. The genes encoding the serotype 5b-specific capsule, i.e., *cps5C*, *cps5B*, and *APL_1581* (a putative glycosyltransferase gene), have significantly lower GC contents indicating recent acquisition of these capsule genes. Five two-component regulatory systems were predicted and annotated, encoded by *arcA/arcB*, *cpxA/cpxR*, *narQ/narP*, *phoB/phoR*, and *qseB/qseC* (Foote et al., 2007).

The complete genome sequence of *A. pleuropneumoniae* serovar 3 strain JL03 is composed of 2,242,062bp with a single circular chromosome. The JL03 genome is approximately 1.4% smaller than that of strain L20. Notably, strain L20 possesses a strain specific genomic island of 37.7kbp encoding a number of phage-related proteins, which is absent in strain JL03. The JL03 genome contained 2,097 potential CDSs with an average size of 941bp, which in sum account for 88.1% of the whole chromosome. As shown in other completed microbial genomes, 18.3% of the CDSs were found to be similar to hypothetical proteins of unknown functions (Xu et al., 2008).

Nine draft genomes from different *A. pleuropneumoniae* serovar reference strains, 4074 of serovar 1, S1536 of serovar 2, M62 of serovar 4, Fem ϕ of serovar 6, CVJ13261 of serovar 9, D13039 of serovar 10, 56153 of serovar 11, 1096 of serovar 12, and N273 of serovar 13, were sequenced by using 454 GS FLX (Xu et al., 2010). All of the genomes were comprised of a circular chromosome with ~2.19 to ~2.33 Mbp in length. The average GC content of

each genome was 41%, which was consistent with that of an entire *A. pleuropneumoniae* chromosome. The median number of CDSs per strain was 2,174; the largest number was 2,223 for serovar 4 strain M62, and the least was 2,096 for serovar 12 strain 1096. Although these draft genomes were extracted from the reference strains of different *A. pleuropneumoniae* serovars that can be assigned to three levels of virulence, the overall genomic characteristics were quite similar (Xu et al., 2010).

1.2 Two-Component Systems

Two-component systems (TCS) sense and respond to changes in the bacterial environment through signal transduction mechanisms. These systems participate in a wide range of signal-processing mechanisms including sporulation, chemotaxis, flagellar motility, gliding, intermediary metabolism, internal nitrogen availability, pathogenicity and virulence constituting an efficient way of transmitting a specific signal from the extracellular environment to the transcriptional machinery as an adequate response to environmental fluctuations (Cases & Lorenzo, 2005; J. B. Stock, Stock, & Mottonen, 1990).

The prototypical two-component system is composed of a sensor and a response protein (Figure 1.2.1). A typical Histidine kinase (HK) sensor protein, consists of a variable N-terminal input or sensory domain connected to conserved C-terminal domains that catalyze phosphorylation. Response proteins, named Response regulators (RRs), consists of a conserved N-terminal receiver domain and a variable C-terminal output or regulatory domain. The HK serves as a receptor for stimuli and as a regulator that controls the activity of downstream signaling components via phosphorylation of the RR. Biochemical regulation of these systems is achieved through phosphorylation, transphosphorylation and dephosphorylation reactions between the proteins involved in a given signal transduction system, the whole of these known as phosphotransfer. HKs autophosphorylate upon recognition of its specific stimulus and transfers the phosphoryl group to its cognate RR turning them in most cases, to an active transcriptional state. The specific recognition between cognate histidine kinases and response regulators is then determinant for the coupling of external stimuli to particular and adequate cellular responses. As soon as, a stimulus has been recognized; the autophosphorylated HK must discriminate its cognate RR from other non-cognate substrates, before transferring the phosphoryl group.

There are three primary mechanisms for ensuring TCSs pathway specificity: molecular recognition, phosphatase activity and substrate competition; additionally, spatial temporal

regulation could also be involved but it has not been thoroughly assessed. The predominant mechanism enforcing specificity is molecular recognition, the intrinsic ability of an autophosphorylated HK to recognize its cognate partner through interface surface interactions. These specificity-determining residues were identified initially through computational analyses of amino acid co-variation in large sets of cognate co-operonic TCS proteins. Mutations in these positions can occur because they are either accompanied or preceded by compensatory changes in other variable positions. Such compensation would result in a coupling between changes in the two positions, or co-evolution between interacting proteins (Goulian, 2009; Jung, Fried, Behr, & Heermann, 2012; Podgornaia & Laub, 2013; Salazar & Laub, 2015; Skerker et al., 2008; A. M. Stock, Robinson, & Goudreau, 2000).

A typical sensor protein, Histidine Kinase (HK), consists of a variable N-terminal input or sensory domain connected to conserved C-terminal domains that catalyze phosphorylation. A typical response protein, response regulator (RR) from now on, consists of a conserved N-terminal receiver domain and a variable C-terminal output or regulatory domain (Bourret & Silversmith, 2010; Gao & Stock, 2010a). The Histidine Kinase serves as a receptor for stimuli and as a regulator that controls the activity of downstream signaling components via phosphorylation of the Response Regulator (Stewart, 2010). Biochemical regulation of these systems is achieved through phosphorylation, transphosphorylation and dephosphorylation reactions between the proteins involved in a given signal transduction system, the whole of these known as phosphotransfer.

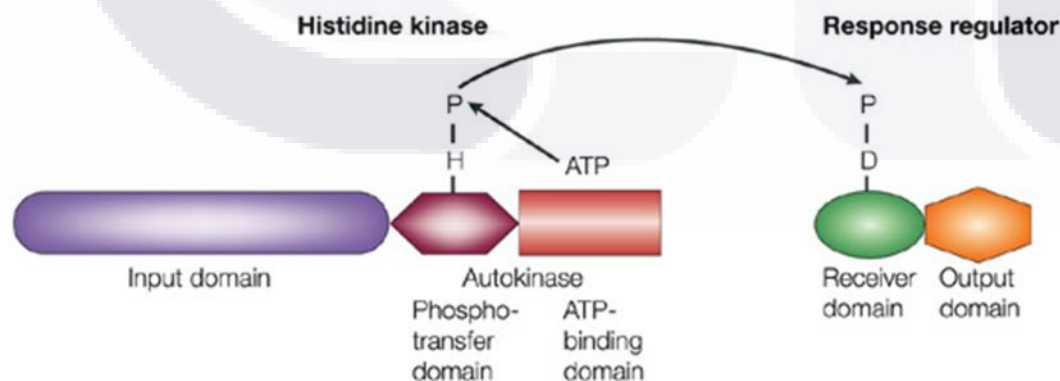


Figure 1.2.1 Two component system model (Jensen, Wang, & Shapiro, 2002a).

The chemistry of the basic two-component phosphoryl transfer signal transduction pathway involves three phosphotransfer reactions and two phosphoprotein intermediates:

1. Autophosphorylation: $\text{HK-His} + \text{ATP} \rightarrow \text{HK-His-P} + \text{ADP}$

2. Phosphotransfer: $\text{HK-His-P} + \text{RR-Asp} \rightarrow \text{HK-His} + \text{RR-Asp-P}$

3. Dephosphorylation: $\text{RR-Asp-P} + \text{H}_2\text{O} \rightarrow \text{RR-Asp} + \text{P}_i$

The γ -phosphoryl group in ATP is first transferred to a conserved Histidine side chain of the HK. The RR then catalyzes the transfer of this phosphoryl group from the phospho-Histidine residue to a conserved Aspartate side chain within its own regulatory domain. Finally, the phosphoryl group is transferred from the phospho-Aspartate residue to water in a hydrolysis reaction (A. M. Stock et al., 2000).

1.2.1 Histidine Kinases

These proteins have a modular architecture with distinct structural domains each with its specific functional role (Stewart, 2010). The amino-terminal sensor contains a variable input domain that is adapted for the detection of a specific stimulus. These sensory domains in most of cases can be accompanied by two transmembrane segments that can either flank or be at the N-terminal or C-terminal portion of the sensor domain, thus anchoring the protein to the cell membrane of bacteria. Histidine kinases can have different domain organization thus permitting extracellular, membrane-embedded or intracellular input domain, as in the later configuring a fully cytoplasmic protein (Cheung & Hendrickson, 2010).

The sensor domain is then followed by a conserved transmitter domain, which transfers the signal to its cognate response regulator through phosphotransfer reactions. This domain can be dissected into two parts: the H-box containing dimerization and histidine phosphotransfer domain (DHp), and the histidine kinase-type ATPase catalytic domain (HATPase_c), better known as catalytic and ATP-binding domain (CA) (Jung et al., 2012). The phospho-accepting conserved histidine resides in the DHp domain (Stewart, 2010).

Interaction between DHp domains elicits the formation of a homodimer where DHp and CA domains associate, at least transiently, to allow transfer of the phosphoryl group from ATP to the phospho-accepting conserved histidine. Autophosphorylation occurs *in trans* within the homodimer; where the phospho-accepting conserved histidine from one protomer receives its phosphoryl group from the ATP molecule bound to the CA domain of the other protomer (Stewart, 2010); although it has been demonstrated that at least some histidine kinase homodimers exhibit an *in cis* mechanism for autophosphorylation (Casino, Rubio, & Marina, 2009; Peña-Sandoval & Georgellis, 2010).

Once a HK has accomplished autophosphorylation, it needs to allow a cognate response regulator to grab its high-energy phosphoryl group. Through protein-protein interaction between HK and the cognate RR, the later transfer the phosphoryl group from the conserved histidine to a conserved aspartate in its receiver domain. The response regulator interacts with the DHp domain, the CA domain and the DHp–CA linker (Stewart, 2010).

1.2.2 Response Regulators

These effector proteins are configured essentially of two domains, a conserved regulatory or receiver domain and a variable output domain; the receiver domain it's commonly located in the N-terminal portion and the variable output domain in C-terminal portion of the protein (A. M. Stock et al., 2000). The primary known function of most receiver domains is to act as phosphorylation-mediated switches. This phosphorylation takes place at the highly conserved active site of the receiver domain in a specific aspartate residue. Typical response regulators receive phosphoryl groups from histidine kinases *in vivo*, receiver domains can catalyze their own phosphorylation *in vitro* using suitable small molecule phosphodonors. Thus, receiver domains are active participants in the phosphorylation reaction rather than passive substrates (Bourret, 2010).

Phosphorylation of the receiver domains induces intramolecular or intermolecular, and inhibitory or activating interactions with themselves and effector domains. The receiver domain exists in equilibrium between inactive and active conformations, with phosphorylation stabilizing the active conformation. Phosphorylation mediates dimerization of the receiver domains, which promotes the output response. Upon phosphorylation, the interaction partner can be the receiver domain itself or a different domain. Heterodomain interactions with the unphosphorylated receiver domain typically prevent effector domain activity by restricting it in an unfavorable conformation (Gao & Stock, 2010a). Receiver domains catalyze its own autodephosphorylation by a mechanism that involves a water molecule in a nucleophilic attack on to the phosphoryl group, leading to the formation of inorganic phosphorus in solution. The autodephosphorylation reaction rates correlates to the different timescales of the specific biological processes. Although the primary route of dephosphorylation is through the phosphatase activity of other proteins (Bourret, 2010).

Output domains are highly variable depending on the diversity of functions that they perform. One of the first identified functions of these output domains was as DNA-binding, but they can also perform other activities such as RNA-binding, ligand-binding, protein-binding,

enzymes or membrane transporters. A great majority of bacterial DNA-binding output domains serve as transcriptional regulators. The most common types of DNA-binding output domains contain H-T-H, Winged Helix and LytTR structural motifs. LytTR-type output domains control production of virulence factors in several important bacterial pathogens. Some other output domains have enzymatic activity and themselves participate in signal transduction: methylsterases, adenylate cyclases, diguanylate cyclases, c-di-GMP-specific phosphodiesterases, histidine kinases, serine/threonine protein kinases and protein phosphatases. Functions through protein–protein interactions have been also reported, in which an output domain shares a sequence similarity to a sigma factor and upon its activation it sequesters the anti-sigma factor then eliciting the sigma factor to function proper. In alkane-degrading sulfate-reducing delta-proteobacteria output domains are membrane transporters for nitrate, sulfate, and dicarboxylates. Several other bacteria encode fusions of the REC domain with membrane-bound and/or ATPase components of the ABC-type transporters. Output domains control bacterial response at a transcriptional and post-transcriptional level, apparently locating two-component systems at a high signaling hierarchy (Galperin, 2010).

1.2.3 Interaction Specificity in Two-Component Systems

Most bacteria encode as few as a single pair and up to a hundred of two-component systems proteins in their genomes and thus its regulation pathways responding to a diverse range of signals. The specific recognition between cognate histidine kinases and response regulators is then determinant for the coupling of an external signal to an adequate cellular response. As soon as a signal has been recognized, the autophosphorylated histidine kinase must discriminate its cognate response regulator from other noncognate substrates, before transferring the phosphoryl group. There are three key mechanisms for ensuring the two-component systems pathway specificity: molecular recognition, phosphatase activity and substrate competition. In addition, spatial-temporal regulation could also be involved but it has not been thoroughly assessed (Podgornaia & Laub, 2013).

In vitro assays have shown that the DHp domain from histidine kinases presents a strong kinetic preference to its cognate response regulator receiver domain over non-cognate substrates (Laub, Biondi, & Skerker, 2007; Skerker, Prasol, Perchuk, Biondi, & Laub, 2005). Also, the recognition of the cognate partner is driven by a small set of amino acid residues that covariate between the histidine kinase and response regulator sequences (Skerker

et al., 2008) and that are in close proximity at the surface of interaction from a solved crystal structure of a cognate pair in complex (HK853-RR468) from *Thermotoga maritima* (Casino et al., 2009).

Most histidine kinases have been described as bifunctional as they can drive the phosphorylation and dephosphorylation of their cognate response regulator, thus acting as a phosphatase modulating the output response and inhibiting it after the signal has subsided, and further minimizing the cross-talk when a cognate response regulator has been phosphorylated by a non-cognate histidine kinase or by a phosphodonor small molecule (Huynh, Noriega, & Stewart, 2010; Huynh & Stewart, 2011b; Siryaporn & Goulian, 2008).

In *Escherichia coli* the relative cellular concentration of the two-component system EnvZ-OmpR cognate pair was found to be in a ratio of about ~1:35, other pathways cognate pairs have been reported with similar ratios (Cai & Inouye, 2002; Miyashiro & Goulian, 2007; Yoshida, Cai, & Inouye, 2002). The higher abundance of response regulators elicits a substrate competition in which a given response regulator effectively outcompetes non-cognate substrates in binding to a cognate histidine kinase (Podgornaia & Laub, 2013).

1.2.4 Interaction Specificity in Two-Component Systems

Ten different two-component system proteins were identified in each *Actinobacillus pleuropneumoniae* strain of all serovars sequenced genomes (Table 1). These were grouped into five two component system pairs with substantial homology to ArcB-ArcA, CpxA-CpxR, QseC-QseB, PhoR-PhoB, and NarQ-NarP. Four of these systems belong to the OmpR family, while the latter is from the NarL family. ArcB is the only hybrid protein coded in *A. pleuropneumoniae* (*App*), hence the ArcB/ArcA system is a multicomponent phosphorelay (Jung et al., 2012; Laub et al., 2007). The relatively small size of the two-component systems signaling network and the divergence among the orthodox systems present in *App*, makes it a suitable model to assess the interaction among prototypical two-components systems' families.

ArcB-ArcA is involved in the metabolic adaptation from aerobic to anaerobic conditions and sensing redox state variations from the quinone pool. Besides, ArcB-ArcA might have a concomitant function with the anaerobic sensor HlyX, serving as an accessory protein within the anaerobic response regulon (Alvarez, Rodriguez, & Georgellis, 2013; Buettner et al., 2009; Buettner, Bendallah, et al., 2008; Buettner, Maas, et al., 2008; Georgellis, Kwon, Lin, Wong, & Akerley, 2001; Green & Baldwin, 1997). An ArcA partial deletion mutant in *App* 7

AP76 showed attenuated infection persistence, deficient auto-aggregation in anaerobic condition and biofilm formation differences only 18 hours after inoculation under anaerobic growth conditions, but no significant difference in survival or growth in both aerobic or anaerobic conditions (Buettner, Bendallah, et al., 2008; Buettner, Maas, et al., 2008). Microarrays comparing an ArcA deletion versus a wild-type isogenic strain under anaerobic conditions, demonstrated that it upregulates itself *arcA* but also *narQ* and *cpxR*, and downregulates *aceE*, *aceF*, and *lpdA*. AceE, AceF, and LpdA are protein subunits from the pyruvate dehydrogenase complex and were downregulated approximately threefold, this suggest pyruvate concentration might be building up because it is not being converted to acetyl-CoA (Buettner, Bendallah, et al., 2008; Buettner, Maas, et al., 2008).

The CpxA-CpxR system has been reported to be involved in the sensing of diverse stimuli, such as: cell envelope integrity, adhesion, central metabolism and misfolded protein either from the inner membrane or the periplasmic space. Misfolded proteins could be sensed directly by CpxA or through accessory proteins for instance NlpE, given the fact that the inhibitory protein CpxP homologue is not coded in *App* (Auger et al., 2009; Hathroubi, Fontaine-Gosselin, Tremblay, Labrie, & Jacques, 2015; Hunke, Keller, & Müller, 2012; Tremblay, Deslandes, & Jacques, 2013; Wolfe, 2010; Wolfe, Parikh, Lima, & Zemaitaitis, 2008). In *App* this system has been reported to be contributing to biofilm formation and virulence (Hathroubi et al., 2015; H. Li et al., 2018; Yan et al., 2020). In *A. pleuropneumoniae* field isolates form biofilms in response to sub-inhibitory concentrations of Penicillin G and upregulate the *cpxRA* operon during biofilm formation, albeit, both of these genes were differentially regulated between static or drip-flow biofilm formation conditions. *cpxA* is upregulated in a 4h static biofilm, while, *cpxR* was downregulated in drip-flow conditions (Hathroubi et al., 2015; Tremblay et al., 2013). Seems that the CpxA-CpxR system in *App* is involved in central metabolism and stress envelope sensing. Nitrogen availability as either nitrite or nitrate; whether within or outside the host, might be the specific stimuli of the NarQ-NarP system (Buettner et al., 2009; Huynh, Chen, & Stewart, 2015a; L. Li et al., 2012, 2014; Noriega, Lin, Chen, Williams, & Stewart, 2010). In *App* this system has been reported to be involved in the response to catecholamines, the transcriptional regulatory response harnessed NarQ-NarP genes. The system was differentially regulated, EPI induced *narQ* and inhibited *narP*, while NE induced *narP* and inhibited *narQ* (L. Li et al., 2012).

Phosphate starvation response is mediated by the PhoR-PhoB system, with inorganic phosphate (Pi) as its specific stimulus (Buettner et al., 2009; Crépin et al., 2011; Lamarche, Wanner, Crépin, & Harel, 2008; Santos-Beneit, 2015). Up-to-date information regarding the system PhoR-PhoB has not been published, nor directly or indirectly.

In *Escherichia coli* HK QseC was described to sense specifically mammalian hormones EPI and NE, but also a bacterial hormone like compound AI 3; furthermore, QseC dependent gene regulation was described to be comprised by non-cognate RRs KdpE and QseF and cognate RR QseB, activated upon phosphorylation (Clarke, Hughes, Zhu, Boedeker, & Sperandio, 2006; Hughes, Clarke, Yamamoto, Rasko, & Sperandio, 2009; L. Li et al., 2012). In *App* the quorum sensing system QseC-QseB possibly recognizes bacterial and animal hormones alike, Norepinephrine (NE), epinephrine (EPI) and Gram-negative bacterial hormones acylhomoserine lactones (AHLs), autoinducer 2 (AI 2) or autoinducer 3 (AI 3) (L. Li et al., 2011, 2012; Liu et al., 2015; Ng & Bassler, 2009; Sperandio, Torres, Jarvis, Nataro, & Kaper, 2003). When *App* was treated with mammalian hormones EPI and NE, the transcriptional regulatory response harnessed TCS genes. The TCS QseC-QseB was positively regulated by EPI and negatively regulated by NE (L. Li et al., 2012).

1.3 Computational Analyses

Comprehensive mathematical approaches to identify specificity conferring residue positions have been published relying on sequence rather than structural information. Nevertheless, the solved structures are crucial for validating the results of these computational approaches. Co-variance based approach, has been applied extensively to individual proteins to gain insights into tertiary structure, these intramolecular contacts residue positions may be linked either functionally, energetically or being in physical proximity making them necessary for structural integrity or for promoting certain protein conformations (Fodor & Aldrich, 2004; Podgornaia & Laub, 2013; Szurmant & Hoch, 2010).

Multiple sequence alignments can be used to predict residue contacts in proteins. If one or both columns in a protein alignment are not at all conserved, that is, are completely random, then there can be only spurious correlations between the two columns and application of any correlated mutation algorithm should yield a low score. A perfectly conserved column, on the other hand, presents a conceptual challenge to co-variance; a correlated mutation algorithm is supposed to detect how the changes in a column affect another given column.

TESIS TESIS TESIS TESIS TESIS

If there are no changes in a column, a correlated mutation algorithm must choose to report no score, a perfectly high score, or a perfectly low score (Fodor & Aldrich, 2004).

1.3.1 Mutual Information

Correlation of residue positions can be measured as mutual information. Applied to interacting proteins, mutual information compares the frequency of occurrence for individual amino acids at two residue positions in the multiple sequence alignment to the frequency of the co-occurrence of these amino acids at the particular residue positions. The basic premise of such an approach is that specificity determining residues positions between interacting proteins need to co-evolve and hence are correlated. Therefore, mutations in these positions can occur because they are either accompanied or preceded by compensatory changes in other variable positions. Such compensation would result in a coupling between changes in the two positions, or co-evolution. Since compensation often involves residues that are proximal in the folded structure, identification of co-evolving positions is expected to be useful in the *ab initio* prediction of protein structure from multiple sequence alignments (Dunn, Wahl, & Gloor, 2008; Szurmant & Hoch, 2010).

Mutual information (MI) calculation compares individual frequencies to the co-frequency of amino acid choice for two residue positions. If two positions are unlinked, the co-frequency is equal to the product of the individual frequencies this results in a value close to 0. If two positions are linked, co-frequency is larger than the product of the individual frequencies and therefore a value larger than 0 (Szurmant & Hoch, 2010).

PARTICULAR BACKGROUND

Bacterial pathogens of porcine respiratory disease complex such as *A. pleuropneumoniae*, *Streptococcus suis*, *Pasteurella multocida*, *Bordetella bronchiseptica*, *Haemophilus parasuis* and *Mycoplasma hyopneumoniae*, were present in the upper respiratory tract of pigs in all swine farms sampled, in the State of Aguascalientes, Mexico. These results show the great respiratory health problems present in the swine farms, which is a reflection of the current situation of the Mexican field (V. M. Loera-Muro et al., 2013). Furthermore *A. pleuropneumoniae* is present and viable in the drinking water of pig farms. The prevalence of *A. pleuropneumoniae* in the drinking water samples from pig farms was 20 %. Biofilms were observed by scanning electron microscopy. Suggesting *A.*

pleuropneumoniae uses biofilms as a strategy to survive in the environment (A. Loera-Muro, Avelar-González, Loera-Muro, Jacques, & Guerrero-Barrera, 2013).

JUSTIFICATION

Two-component systems are the most important signal transduction pathway in bacteria, enabling bacteria to sense its surroundings and to respond adequately in order to adapt an ever-changing environment. Members of this family of signal transduction have been identified in archaeal, bacterial and in some lower eukaryotic genomes. Determining the *in vivo* interaction between the five predicted two-component systems of *Actinobacillus pleuropneumoniae*, in a cognate pair wise fashion or with cross-talk between them as a salvaging strategy for damaged signal processing mechanisms, will allow us to infer how *A. pleuropneumoniae* interacts with and responds to the environmental signals. This is an important question to be addressed, given the small number of systems present for signal processing, such as host-pathogen interaction, virulence, metabolic activities and survival. Given the fact that these signaling pathways are not known to be coded in higher eukaryotes, this information can be used to design potentially new drug targets or therapies that combine current antibiotics with two-component systems inhibitors, for more successful approaches for treating pathogenic bacterial infections.

HYPOTHESIS

Ho: Two-component systems in *A. pleuropneumoniae* interact with cognate pairs with a set of amino acids that confer specificity to the molecular recognition.

OBJECTIVE

General Objective

Detect two-component systems protein interaction amino acids in *A. pleuropneumoniae*.

Particular Objectives

Determine the Two-components systems coded in *Actinobacillus pleuropneumoniae*.

Evaluate the genomic context of the Two-component systems.

Generate datasets of homologue protein from Two-components systems in the bacteria domain.

Write a code that computes the covariance between amino acids from interacting cognate proteins.

Analyze the tridimensional structure of the Two-component system complex to infer a prototypical interaction interface.

Determine the distance among amino acid from the prototypical interaction interface.

Define a function that represents the coupling of amino acids that take into account structural and homology information.

Suggest a set of amino acids that help discriminate through molecular recognition between cognate and non-cognate proteins.

MATERIALS AND METHODS

2.1 Identification of putative two-component systems proteins in *Actinobacillus pleuropneumoniae*

The complete genome sequences from *A. pleuropneumoniae* serovar 5b strain L20, serovar 3 strain JL03 and serovar 7 strain AP76, as well as the draft genome sequences from serovar reference strains, serovar 1 strain 4074, serovar 2 strain S1536, serovar 4 strain M62, serovar 6 strain Fem ϕ , serovar 9 strain CVJ13261, serovar 10 strain D13039, serovar 11 strain 56153, serovar 12 strain 1096 and serovar 13 strain N273; were used to infer the coding sequences of two-component system proteins. The putative coding sequences of histidine kinase were identified, using as a query the conserved domains DHp and CA, while the model used for response regulators was the conserved domain REC. The query sequence was protein blasted against *A. pleuropneumoniae* taxid 715.

Putative TCS sequences were queried with NCBI's Blast-CD Search Standard to identify multi domain homology and possible functional roles were assigned (Marchler-Bauer et al., 2011). Multiple sequence alignments by possible functional role were assembled for each TCS putative protein coded in the *A. pleuropneumoniae* serovar strain; these multiple sequence alignments were used to determine differences between putative protein sequences. The genomic context from these TCS putative proteins was determined with NCBI Graphical Sequence Viewer 3.3.1 (Rangwala et al., 2021).

2.2 Identification of genomic context of two-component system putative proteins of *Actinobacillus pleuropneumoniae*

The identified putative protein coding sequences from *A. pleuropneumoniae* were used as input in the Gene Context Tool III to visualize the genomic context of these putative two-component system proteins (Martinez-Guerrero, Ciria, Abreu-Goodger, Moreno-Hagelsieb, & Merino, 2008). The putative sequence was queried against the complete genomes of *A. pleuropneumoniae* serovar 5b strain L20, serovar 3 strain JL03 and serovar 7 strain AP76. Images were processed so that the three different genomic contexts appear in just one figure.

2.3 Computational analyses

2.3.1 Two-Component Systems paralogous families' datasets

Sequence homologues to the TCS proteins coded in *A. pleuropneumoniae* were retrieved from NCBI's protein database, by restricted searches. The filters used were reference sequence (RefSeq) databases within the bacteria domain. Ten different databases of putatively homologue proteins were assembled by annotation for each of the TCS proteins from *A. pleuropneumoniae*. Further processing removed WP_ RefSeq annotations from these databases and only kept YP_ and NP_ annotations, which came from completely sequenced bacterial genomes.

The genomic context for each protein in the datasets were examined with NCBI Graphical sequence viewer 3.44.0 (Rangwala et al., 2021), if two proteins belong to the same putative operon they were deemed as cognates. Orphan proteins were removed from datasets to avoid possible false positives of non-cognate interactions. These cognate TCS datasets were processed with NCBI's Blast CD-Search standard to identify multi-domain homology to each protein from the following cognate protein pairs: PhoR-PhoB, CpxA-CpxR, QseC-QseB, and NarQ-NarP (M. Yang, Derbyshire, Yamashita, & Marchler-Bauer, 2020). Cognate sequences that did not had homology to both proteins were deprecated from the datasets. As a special case the NarQ-NarP dataset allowed cognate sequence pairs that had homology to both NarQ-NarP proteins, or at least to one of the former and to NarX or NarL, but not to both. Cognate sequences that did not comply to criteria were deprecated. Datasets of putatively interacting cognate proteins were merged, each cognate sequence pair from the same genome were concatenated into a single sequence arranged as Histidine kinase-Response regulatorwise (HK::RR). Dataset arrangement used the HKs' ID as the

sole identifier for each concatenated sequence. Finally, five cognate TCS proteins datasets were assembled: PhoR::PhoB (n=499), CpxA::CpxR (n=205), QseC::QseB (n=145), and NarQ::NarP (n=165).

Each dataset was processed with Clustal Omega 1.2.1 to build MSAs and Jalview 2.11.1.4 was used to allow a maximum sequence redundancy of 99 percent, further minor processing was made manually (Sievers & Higgins, 2020; Waterhouse, Procter, Martin, Clamp, & Barton, 2009). Cognate TCS concatenated MSAs were used to calculate mutual information with average product correction (Mlp), calculations were made with an algorithm written with Python 3.6 based on Dunn 2008 but that does take in to account positions which are not completely represented (Dunn et al., 2008).

2.3.2 Mutual information with average product correction (Mlp)

The mutual information with average product correction effectively separates structural and functional constraints from background, either due to shared ancestry or random noise. It is calculated as mutual information between two positions minus the average product correction (Equation 2.3.2.1). The computation of the Mlp for Specificity determining matrices of co-evolving interaction interface (SDMCEll) was performed using published Perl scripts (Dunn et al., 2008), whereas the interface coupling index (ICI) uses a completely new script written in Python 3.6. The Mlp algorithm written in Perl requires reference sequence positions for calculations. *Escherichia coli* K-12 MG1655 concatenated cognate TCS homologue sequences were used for PhoR::PhoB, CpxA::CpxR, and QseC::QseB. While *Vibrio cholerae* MJ-1266 NarQ::NarP concatenated sequence was used, *E. coli* K-12 MG1655 has no cognate homologue for the NarQ-NarP system.

MI measures the uncertainty of a position given information about the other, and can be used as a degree of correlation between to position of an alignment. The value varies from 0 to 1, the greater the value the more interdependency between position, hence amino acids. MI is calculated as the sum of Shannon's entropy of independent positions minus the joint Shannon's entropy (Equation 2.3.2.2). Shannon's entropy is the first value that needs to be calculated and it is done so by using each residue frequency, multiplied by the $\log_{20}$ of that same frequency. The result is the sum of all values for each of the 20 residues, a complete conservation will give a 0 value and 1 when all 20 residues are equally distributed (Equation 2.3.2.3). The joint Shannon's entropy is calculated similarly, except that the frequencies are

calculated for a pair of residues and the sum extends over the 400 possible combinations (Equation 2.3.2.4). The value for the joint entropy ranges from 0 to 2.

$$Mlp_{(a,b)} = MI_{(a,b)} - APC_{(a,b)} \quad (2.3.2.1)$$

$$MI_{(a,b)} = H_a + H_b - H_{(a,b)} \quad (2.3.2.2)$$

$$H_a = - \sum_x p(a) \log_{20} p(a) \quad (2.3.2.3)$$

$$H_{(a,b)} = - \sum_x \sum_y p(a,b) \log_{20} p(a,b) \quad (2.3.2.4)$$

The average product correction (Equation 2.3.2.5) gives an excellent approximation to the background MI shared by position a and b. It is calculated as the product of the mean mutual information of columns a and b, divided by the overall mean mutual information (Equation 7). The mean mutual information is the sum of the MI value of each position, divided by $m = n - 1$ for convenience, n is the total number of columns in the alignment (Equation 2.3.2.6). Except when $x = a$ or $x = b$, that means that when calculating for a given position the MI of the own column will not be taken into account. The overall mean mutation is 2 times the sum of all MI values for each position divided by the product of mn (Equation 2.3.2.7).

$$APC_{(a,b)} = \frac{MI_{(a,\bar{x})} MI_{(b,\bar{x})}}{\overline{MI}} \quad (2.3.2.5)$$

$$MI_{(a,\bar{x})} = \frac{1}{m} \sum MI_{(a,x)} \quad (2.3.2.6)$$

$$\overline{MI} = \frac{2}{mn} \sum MI_{(x,y)} \quad (2.3.2.7)$$

2.3.2.1 Mlp Pseudo-Code

In this part a detailed description of the script used to calculate the mutual information with average product correction will be described. The real code has more internal operation and will be described in the proper parts of this methodology, since the code is quite large. The program requires as input a multiple sequence alignment of protein sequences concatenated and that are in the same operon.

This program will allow a user to calculate the MIp of a multiple sequence alignment.

Counter = Initializes dictionary of all amino acids with its one letter naming convention

CountResults = Initializes the sum of all amino acids in the processed column

Probabilities = Initializes the probabilities of each amino acid in a column

ShannonEntropies = Initializes the entropies for each amino acid in a column

TotalEntropies = Initializes the sum of all entropies of a column

Aminoacids = A dictionary of all amino acids in its three-letter naming convention

File = User defines the multiple sequence alignment in fasta format to be processed

Alignments = Uses AlignIO parsing for the MSA to be processed

Print “the length of the alignment and an overview of the alignment”

i = Starts in cero

for each position in the alignment

 Position = Assigns the position number to the column

 Counter [[]] = Counts the total number of times an amino acid or a gap appears in a column

 CurrentResults = Returns the column number and frequency of each amino acid and gap and appends the values into a dictionary

 i = Changes column in the alignment

print “Frequencies” when the task is finished

Probabilities = Now has the information of each column and the frequencies of each amino acid and the gap

for each position in the alignment

 for each amino acid and the gap

 Probabilities = Returns the probabilities for each amino acid or gap. Frequencies are divided by the number of occupied positions in the column, that is the total number of positions minus the total number of gaps in that column

Print “Probabilities” when the task is finished

ShannonEntropies = Now has the information of each column and the probability of each amino acid and the gap

m = A counter that starts in cero
 n = A counter that starts in cero
 for each column in alignment
 for each amino acid probability in a column
 if probability is cero with decimal cero
 ShannonEntropies[n][m] = deposits a cero
 elif probability is cero
 ShannonEntropies[n][m] = deposits a cero
 Else
 ShannonEntropies[n][m] = returns the Shannon entropy
 for each amino acid and the gap
 m = Changes amino acid or gap in each column
 m = restarts index to one
 n = Changes column in the Shannon entropy dictionary
 print "Entropies" when the task is finished
 k = A counter that starts in cero
 TotalEntropy = Starts variable in cero
 for each column in the Shannon's entropy dictionary
 for each amino acid and gap in column
 TotalEntropy = Returns the sum all Shannon's entropies
 k = Changes the amino acid or gap in the column
 Appends the total Shannon's entropy of the column in TotalEntropies
 TotalEntropy = Restarts value to cero
 k = Resets value to one
 print "Total entropies" when the task is finished
 AAindex = Creates a dictionary for each amino acid and the gap in a matrix
 def jointProb (posicion1, posicion2) Performs a task that will return the joint entropy of two column positions
 JFM = Creates a matrix of size 21 by 21 to deposit the columns joint frequencies

JFSM = Creates a matrix of size 20 by 20 to deposit the values of joint frequencies without the values of pair formed by gap or are both gaps

JPM = Creates a matrix of size 21 by 21 to deposit the joint probability of the columns

JEM = Creates a matrix of size 20 by 20 to deposit the joint Shannon's entropy

sequence1 = Extracts the identities of the first column as a sequence

sequence2 = Extracts the identities of the second column as a sequence

AAn = Starts a variable as zero

for each position in the column

 for each identity in the amino acid index of the first column

 for each identity in the amino acid index of the second column

 if the position in the first column is the same to an identity in the amino acid index returns that identity position in the matrix

 if the position in the second column is the same to an identity in the amino acid index returns that identity position in the matrix

 AAn = Returns the sum of the present value of the position in the matrix of a pair of identities plus one

 Deposits the new value on the correct matrix position

for each amino acid in a row in the matrix JFM except those with gaps

 for each amino acid in a column in the matrix JFM except those with gaps

 JFSM = Deposits the values in a 20 by 20 matrix keeping the same order

JFS = Returns the sum of all values of the matrix JFSM of amino acid pairs frequencies

if the sum of all amino acids frequencies is not cero

 Returns the probabilities of each amino acid pair in the matrix JPM

else

 JFSM = Returns a matrix of amino acid pairs probability equal to the matrix of amino acid pairs frequencies

for each amino acid in a row in the matrix JPM except those with gaps

 for each amino acid in a column in the matrix JPM except those with gaps

 if the value position is not cero

 e = Returns the joint Shannon's entropy for a given pair of amino acids from two columns

 Deposits the joint Shannon's entropy in the correct position in the joint entropy matrix

JEMS = Returns the sum of all entropies

Values = Returns both the total joint Shannon's entropy and the number of occupied positions on the column that are amino acid pairs

 Returns Values from the task jointProb and finishes the task

MImatrix1 = Creates a matrix of the same size as the number of columns in the alignment to deposit mutual information

Weight_matrix = Creates a matrix of the same size as the number of columns in the alignment to deposit column occupancy

Joint_entropy = Creates a matrix of the same size as the number of columns in the alignment to deposit joint entropy

for each column in the alignment except the last one

 for each column in the alignment except the ones already visited in the previous cycle

 X = Performs the task in jointProb

mutualinformation = Returns the mutual information between two -
columns
Deposits the value of mutual information in its two respective
symmetric positions in MImatrix1
weights = Returns the occupancy ratio from amino acid pairs in the
two columns
Deposits the value of weight in its two respective symmetric
positions in weight_matrix
Deposits the value of the joint Shannon's entropy in its two
respective symmetric positions in joint_entropy
save = Determines the name of the output file with the first 6 characters
from the input file and the ending "_MI_matrix.csv"
Creates a .csv output file with the name as determined in the previous
variable that contains all mutual information values
print "Mutual information" when the task is finished
MI = Starts variable at cero
MIb = Starts variable at cero
n = Returns the total number of columns in the alignment
m = Returns the total number of columns in the alignment minus one
counter_position = Initializes counter at cero
for each row in the mutual information matrix
 for each column in the mutual information matrix
 if the row and column are not from the same alignment
 position
 Sums all values of one column with the rest of column
 in the alignment
 if the value to be summed is different from cero
 A counter for all position that are not cero
 if counter is cero
 Returns the mean mutual information of a column as cero
 else
 MIb = Returns the mean mutual information of a column
 Appends the mean MI of that column in an array

Resets the MI value to cero

Resets counter to cero

for each row in the MI matrix

 for a column that stars after the value of row

 MIsun = Returns the upper triangular sum of the MI matrix

 if the value to be summed is different from cero

 A counter for all position that are not cero

if counter is not cero

 Returns mean MI as cero

else

 Returns the mean MI of all the alignment

APCmatrix = Creates a matrix the size of the alignment.

APC = Starts variable at cero

for each column in the alignment

 for each column in the alignment

 if the columns are not the same and the mean value of the first column is not cero and the value of the second column is not cero

 APC = Returns the value of the average product correction for those two columns

 Deposits the APC value for those columns in its symmetric positions

save = Determines the name of the output file with the first 6 characters from the input file and the ending “_APC_matrix.csv”

Creates a .csv output file with the name as determined in the previous variable that contains all average product correction values

print “Average product correction” when the task is finished

MIp = Starts variable at cero

MIpmatrix = Creates a matrix the size of the columns in the alignment

for each row in the matrix

 for each column in the matrix

 MIp = Returns the value of the mutual information with the average product correction for two columns

Deposits the MIp value for those columns in its symmetric positions

save = Determines the name of the output file with the first 6 characters from the input file and the ending “_Mip_matrix.csv”

Creates a .csv output file with the name as determined in the previous variable that contains all MIp values

print “Mip” when the task is finished

...continues at 2.3.3.1

...after 2.3.3.1

print “Processing output file”

save = Determines the name of the output file with the first 6 characters from the input file and the ending “_Mip_detailed.csv”

writes the headers on the detailed output file

for each row in the matrix

 for each column in the matrix

 if the columns are not the same

 Writes the values of Shanno’s entropy for the first column, Shannon’s entropy for the second column, Joint Shannon’s entropy for the two columns, MI for the two columns, APC for the two columns, MIp for the two columns, weight of the occupancy of the two columns, and wMip for the two columns

Output file with all values for the processed alignment

Program ends

2.3.3 Joint column occupancy and weighted Mip

Since the algorithm takes into account pairs of columns that are not completely occupied, there are cases in which a group of highly related sequence with low occupancy can make artifacts. This artifact will consist of a pair of columns that have only a handful of amino acid pair positions but its joint frequency is overrepresented i.e., they have only 35 pairs of amino acids from a MSA with 100 sequences and all those pairs have the same pair of amino acids. Without the correction for occupancy that the weighted ratio gives, these artifacts will show high mutual information between columns. Although this would only be true for those

highly related sequence, it does not represent a real co-evolution for the orthologue proteins. Therefore, these MIp values require the adjustment for joint column occupancy. Hence, the value obtained for each pair of columns is then transformed to weighted MIp (wMIp) (Equation 3), that adjust the values according to its joint occupation. The datasets generated and analyzed can be found in GitHub (<https://github.com/PhDEMMartin/Interface-coupling-index>).

$$wMIp = MIp \cdot \frac{j}{n} \quad (2.3.3.1)$$

j = joint pair occupation

n = total sequences in MSA

2.3.3.1 wMIp PseudoCode

... starts after 2.3.3.1

wMIp = Starts the variable at zero

wMipmatrix = Creates a matrix the size of the columns in the alignment
for each row in the matrix

for each column in the matrix

wMip = Returns the weighted MIp value for two columns

Deposits the MIp value for those columns in its symmetric
positions

save = Determines the name of the output file with the first 6 characters
from the input file and the ending “_wMip_matrix.csv”

Creates a .csv output file with the name as determined in the previous
variable that contains all wMip values

print “wMip” when the task is finished

... continues at 2.3.3.1

2.3.4 Tridimensional mapping of specificity-determining residue positions

Specificity-determining residues with high intermolecular MIp values were mapped onto the HK853-RR468 crystal structure from *Thermotoga maritima* (PDB: 3DGE) by using DeepView/Swiss-PdbViewer 4.1.0 (Guex & Peitsch, 1997; Johansson, Zoete, Michielin, &

Guex, 2012). Images were obtained by visualizing only one histidine kinase structure in complex with one response regulator, and high intermolecular Mlp values residues were identified with as green, either in histidine kinase or response regulator, while the histidine kinase and response regulator chains are in contracting colors. Conserved phosphotransfer histidine and aspartate residues are depicted in red.

2.3.5 Protein-protein interface amino acids

Reference protein sequences HK853 and RR468 from *Thermotoga maritima* MSB8 were used for protein homology modelling with SWISS-MODEL. The tridimensional template structure used was the solved HK-RR complex HK843-RR468 (PDB: 3DGE) TCS from *T. maritima* MSB8 (Figure 3.2.1 A), using DeepView/Swiss-PdbViewer 4.1.0 (Guex & Peitsch, 1997; Johansson et al., 2012). Tridimensional structures of HK853 and RR468 were modelled separately and protein surface from the independent models was computed. Wherefrom, both protein surfaces from the HK843-RR468 heterodimer were used to calculate the contact surface between them and used thenceforth as the prototypical interaction interface (PII) during signal transduction (Figure 3.2.1 B and C). Amino acid positions in the interaction interface were used to evaluate molecular recognition of TCSs' proteins. Further processing added the concatenated sequence from *T. maritima* MSB8's HK853::RR468 to each of the orthologous MSAs and once again were aligned. Henceforth, amino acid positions were correlated with the PII and were organized as symmetrical matrices (Figure 3.2.1 B and C, and Table 3.2.1).

2.3.6 Prototypical interaction interface's distance among amino acids

Using a script written in Python 3.6 the minimal distance between residue pair's last backbone atoms among all interfacing residues was calculated. The minimal distance obtained for a pair of interfacing residues was used to make a parameter called adjusted minimal distance (amd) (Equation 2.3.6.1), this parameter was used to divide the wMlp of all residue pairs in the interface (Equation 2.3.3.1), giving place to the new parameter interface coupling index (ICI) (Equation 2.3.6.3). The datasets generated and analyzed, as well as the algorithm for this study can be found in GitHub.

$$amd = \frac{AApd}{md} \quad (2.3.6.1)$$

AApd = minimal distance between a pair of amino acids

md = minimum observed distance in the interface

$$ICI = \frac{wMIp}{amd} \quad (2.3.6.2)$$

2.3.7 Amd PseudoCode

This algorithm can be used to compute the minimal distance between any pair of amino acids in the interface of any pair of proteins in complex. It is required that the interface is first determined by an appropriate algorithm or software, two software that can be used for this task are DeepView/Swiss-PdbViewer 4.1.0 or ChimeraX 1.3 (Guex & Peitsch, 1997; Johansson et al., 2012; Pettersen et al., 2021). Once that the interface has been computed, this program requires as input the corresponding number for each amino acid in the interface as defined in the PDB complex archive.

This program will allow a user to compute the minimal distance between amino acids in the interface of a protein complex.

`aa_index` = Creates a dictionary for amino acid position in a matrix

`def read_pdb` Performs a task that will retrieve the global structure of a PDB data

`parser` = Calls an inbuilt task from biopython API to parse structural data from a PDB

`Assigns structure` as a global variable

`structure` = Uses parser to get the structure of a given structure it requires a name for the complex and the name of the PDB file

`print PDB's structure data`

`def aa_min_distance` Performs a task that will

`residues` = Retrieves individual amino acids from structure

`res_list` = Generates a list of the amino acids from structure with the one-letter naming convention

`interface_residues` = A user made list of the interface amino acids

aa_atoms = A dictionary of side chain atoms for each of the twenty standard amino acids

residues = for each amino acid in the PDB data get its residue three-letter naming convention code

for each amino acid position in the first protein

for each amino acid position in the second protein

x = Adjusts the amino acid position from the first protein sequence to that of the PDB's numeration

y = Adjusts the amino acid position from the second protein sequence to that of the PDB's numeration

aa_name1 = Obtains the three-letter amino acid code for the corresponding amino acid of the first protein

aa_name2 = Obtains the three-letter amino acid code for the corresponding amino acid of the second protein

min_distance = Starts minimum distance arbitrarily at 40nm

for each residue in the side chain atom index

if residue's name matches a residue in the side chain atom index

for each residue in the side chain atom index

if residue's name matches a residue in the side chain atom index

for each side chain atom for the corresponding residue from the first protein

for each side chain atom for the
corresponding residue from the
second protein

```

alpha_carbon_1 = Obtains
the      tridimensional
position   for   the
corresponding residue from
the first protein
alpha_carbon_2 = Obtains
the      tridimensional
position   for   the
corresponding residue from
the second protein
distance = computes the
Euclidean distance between
side   chain   atom
coordinates
    if min_distance is
    greater      than
    distance
        min_distance =
        the value from
        distance
        obtains first
        residue
        identifier and
        position
        obtains second
        residue
        identifier and
        position
    
```

Creates a csv print with the first residue position, the second residue, and the minimal distance between them

2.3.8 Specificity determining co-evolving interface matrices

Identified amino acid positions from the PII were arranged as matrices to plot MIP values in heatmap charts for each of the parental MSAs databases aforementioned. Pairwise compared MIP heatmap charts were performed using absolute value distance between parental matrices (Equation 2.3.8.1); but also, filtered by mean value between parental matrices (Equation 2.3.8.2), using the 80-percentile value as threshold. In the equations capital letters refer to any given matrix; while subscripts are the corresponding position of each value within the matrix. We denominate this novel method as Specificity Determining Matrices of Co-Evolving Interaction Interface (SDMCEII). This method allows us to pinpoint the exact amino acids that confer specificity between TCS; its sequence position and tridimensional location in the interaction interface. All statistical processing was made with software R 3.2.3 and gplots package (Sawitzki, 2009).

$$|A - B| = \begin{bmatrix} A_{1,1} & \dots & A_{1,40} \\ \dots & \dots & \dots \\ A_{40,1} & \dots & A_{40,40} \end{bmatrix} - \begin{bmatrix} B_{1,1} & \dots & B_{1,40} \\ \dots & \dots & \dots \\ B_{40,1} & \dots & B_{40,40} \end{bmatrix} = \begin{bmatrix} |A_{1,1} - B_{1,1}| & |\dots| & |A_{1,40} - B_{1,40}| \\ \dots & \dots & \dots \\ |A_{40,1} - B_{40,1}| & |\dots| & |A_{40,40} - B_{40,40}| \end{bmatrix} \quad (2.3.8.1)$$

$$\overline{AB} = \begin{bmatrix} A_{1,1} & \dots & A_{1,40} \\ \dots & \dots & \dots \\ A_{40,1} & \dots & A_{40,40} \end{bmatrix} - \begin{bmatrix} B_{1,1} & \dots & B_{1,40} \\ \dots & \dots & \dots \\ B_{40,1} & \dots & B_{40,40} \end{bmatrix} = \begin{bmatrix} \frac{A_{1,1} + B_{1,1}}{2} & \dots & \frac{A_{1,40} + B_{1,40}}{2} \\ \dots & \dots & \dots \\ \frac{A_{40,1} + B_{40,1}}{2} & \dots & \frac{A_{40,40} + B_{40,40}}{2} \end{bmatrix} \quad (2.3.8.2)$$

Sequence logos of the HK RR interaction core positions were acquired for each of the parental MSAs databases, but also from MSAs concatenated sequences of the non-hybrid TCSs coded in *A. pleuropneumoniae* (Crooks, Hon, Chandonia, & Brenner, 2004).

RESULTS

3.1 Putative two-component system proteins

Based on protein blast analysis as described in methodology, ten different two component system proteins were identified in each and every one of the *Actinobacillus pleuropneumoniae* reference strains of all serovars sequenced genomes (Table 3.1.1); excluding serovar 8 which was not sequenced at the time. All of these were grouped into five two component system pairs with substantial homology to ArcB ArcA, CpxA CpxR, NarQ NarP, PhoR PhoB and QseC QseB (Table 3.1.1). From these identified systems, only CpxA CpxR, PhoR PhoB and QseC QseB were identified to be coded in the same operon and thereby defined as cognates. Whereas, the remaining proteins ArcB, ArcA, NarQ and NarP were coded in separate operons and, thereby, defined as orphans. This was only determined for *A. pleuropneumoniae* serovar 3 strain JL03, serovar 5b strain L20 and serovar 7 strain AP76, which were completely sequenced thus, the genomic context could be inferred.



Figure 3.1.1 arcA Gene Context Gene neighbors of orphan *arcA* TCS gene, and *A. pleuropneumoniae* complete genome sequences for comparison.

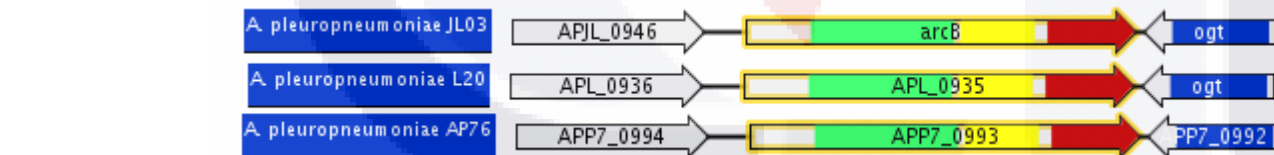


Figure 3.1.2 arcB Gene Context Gene neighbors of orphan *arcB* TCS genes, and *A. pleuropneumoniae* complete genome sequences for comparison.

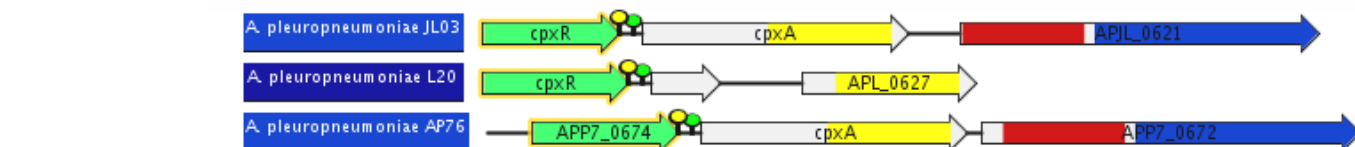


Figure 3.1.3 cpxRcpxA Gene Context Gene neighbors of co-operonic *cpxRcpxA* TCS genes, and *A. pleuropneumoniae* complete genome sequences for comparison.

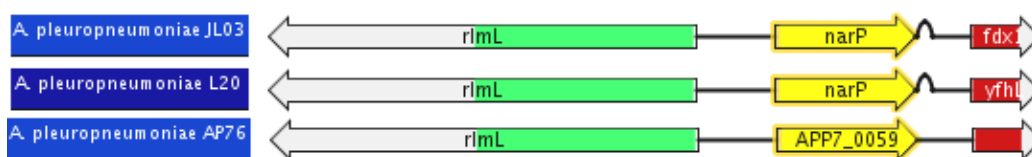


Figure 3.1.4 *narP* Gene Context Gene neighbors of orphan *narP* TCS gene, and *A. pleuropneumoniae* complete genome sequences for comparison.

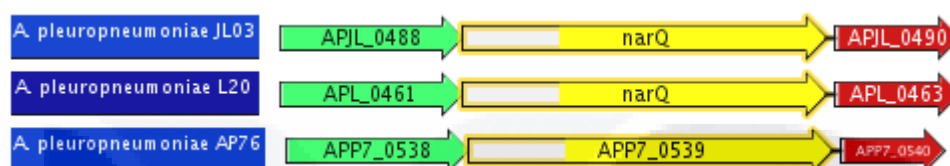


Figure 3.1.5 *narQ* Gene Context Gene neighbors of orphan *narQ* TCS gene, and *A. pleuropneumoniae* complete genome sequences for comparison.

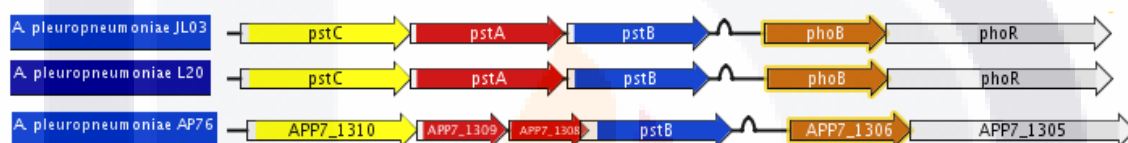


Figure 3.1.6 *phoBphoR* Gene Context Gene neighbors of co-operonic *phoBphoR* TCS genes, and *A. pleuropneumoniae* complete genome sequences for comparison.

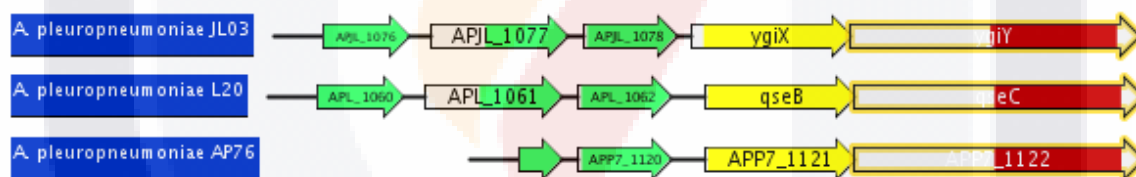


Figure 3.1.7 *qseBqseC* Gene Context Gene neighbors of co-operonic *qseBqseC* TCS genes, and *A. pleuropneumoniae* complete genome sequences for comparison.

Two-component systems' architecture in *App* suggests that these are phosphotransfer systems, with the exception of the ArcB-ArcA system, which exhibits an unorthodox architecture. ArcB harbors in its C-terminal portion two extra domains with homology to a receiver domain (REC)[CDD:00156] and a histidine phosphotransfer domain (Hpt)[CDD:cd00088] just after the catalytic and phosphorylation domain (CA)[CDD:cd00075] thus, a multi-domain unorthodox histidine kinase that constitutes a multicomponent phosphorelay system (Figure 3.1.8) (Goulian, 2009; Laub et al., 2007; Seshasayee, Bertone, Fraser, & Luscombe, 2006; A. M. Stock et al., 2000).

Multiple sequence alignment (MSA) showed that CpxA-CpxR had interesting features, such as CpxA being truncated in strains 5b L20, 6 Fem ϕ , NCTC11383, and NCTC11407. In

strains 5b L20 and NCTC11383 open reading frames (ORFs) APL_0627 and appser6_6990, respectively, code for a HK lacking the sensory portion and ad ORF APL_0628 for both codes for a partial 95 amino acid sensor domain from CpxA (Figure 3.1.8).

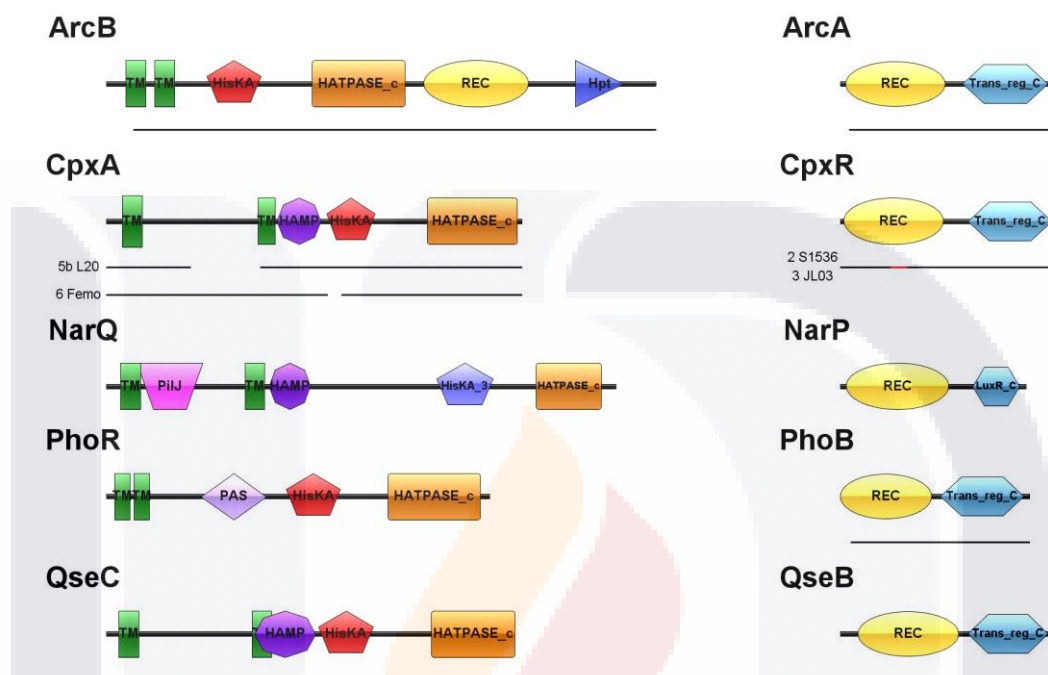


Figure 3.1.8 Two-Component Systems Architecture in *Actinobacillus pleuropneumoniae* Domain architecture from all five Two-component System putative pairs coded in *A. pleuropneumoniae* are depicted; CpxA CpxR, PhoR PhoB and QseC QseB are co-operonic whilst the remaining TCSs proteins are coded as orphans. Fine lines down wild type architectures depict serovar specific mutations that render structural or functional constraints (Table 1). TM: Transmembrane domain. HSKA: HK A dimerization/phosphoacceptor. HATPASE_c: Histidine kinase-like ATPases. HPT: Histidine containing phosphotransfer. HAMP: HK, Adenyl cyclase, Methyl accepting protein, and Phosphatase. PilJ: Type IV pili methyl accepting chemotaxis transducer N term. HSKA_3: Dimerization and phosphoacceptor of a sub family of HK. PAS: Bind ligands; sense light and oxygen in signal transduction. REC: Signal phosphate receiver. Trans_reg_C: DNA binding effector. LuxR_c: DNA binding of LuxR like proteins.

Whereas, in 6 Fem ϕ , NCTC11383, and NCTC11407 CpxA's histidine phosphorylation and phosphotransfer core is disrupted, hence, constituting a non-functional sensor protein. Presumably hindering the CpxA-CpxR signal transduction path in these strains.

Response regulator protein CpxR in *A. pleuropneumoniae* 2 4226, 2 S1536, 3 JL03, and NCTC10976 has an 18 amino acid gap comprising $\alpha 3$, $\alpha 3$ - $\beta 4$, and $\beta 4$, which might impede the phosphotransfer activity due to structural constraints (Figure 3.1.8, Table 3.1.1, Figure 3.1.9 and Table 3.4.1).

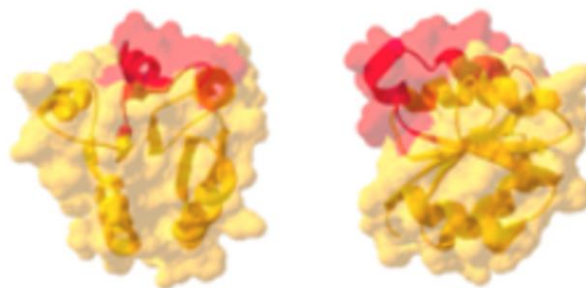


Figure 3.1.9 CpxR gapped proteins from *Actinobacillus pleuropneumoniae* Gap amino acids identified in the MSA from *A. pleuropneumoniae*, lost residues are depicted as red in the tridimensional structure from the RR468 from *Thermotoga maritima* MSB8[PDB:3DGE]. This gap comprises $\alpha 3$, $\alpha 3$ - $\beta 4$ loop and the initial part of $\beta 4$ from the RR structure.

PhoR-PhoB's MSA highlighted the fact that all PhoB RRs from sequenced *A. pleuropneumoniae* strains excluding 1 4074, 18-1342, G1-9626, S8, and S4074, lacked the homologue residues D12 and D13 to CheY's REC domain. These constitute the acidic pocket that coordinates Mg^{2+} or other divalent cations that catalyze the phosphorylation of the conserved aspartate from the phosphorylation site (Lukat, Stock, & Stock, 1990; Needham, Chen, & Falke, 1993).

NarQ protein in NCTC11383 is cleaved in two disrupting the histidine kinase dimerization and phosphoacceptor domain leading to a loss of this system's upstream pathway in this strain. While ArcB-ArcA and QseC-QseB systems have no noteworthy features in its protein sequences that could potentially disrupt the signal transduction pathways to which they belong.

Serovar	1	2	2	3	4	5b	6	7	8	9	10	11	12	13	
Strain	4047	4226	S1536	JL03	M62	L20	Fem φ	AP76		CVJ13261	D13039	56153	1096	N273	
ArcB	AWG95381	EFL78370	EFM87742 *	ABY69503	EFM89876 *	ABN74031	EFM91984 *	EFL79927	ACE61645	CUU52249	EFM94159 *	EFM96380 *	EFM98488 *	EFN00684 *	EFN02794 *
ArcA	AWG94468	EFL81615	EFM88625	ABY68655	EFM90830	ABN73156	EFM93039	EFL80674	ACE60700	CUU51367	EFM95227	EFM97386	EFM99564	EFN01595	EFN03767
CpxA	AWG95045	EFL77662	EFM88032	ABY69192	EFM90213	ABN73729 *	EFM92341 *	EFL81616 *	ACE61325	CUU51917	EFM94540	EFM96605	EFM98782	EFN01018	EFN03058
CpxR	AWG95046	EFL77661 18	EFM88033 18	ABY69193 18	EFM90214	ABN73731	EFM92343	EFL81615	ACE61326	CUU51918	EFM94541	EFM96606	EFM98783	EFN01019	EFN03059
NarQ	AWG94880	EFL79103	EFM88129	ABY69070	EFM90338	ABN73566	EFM92501	EFL81026	ACE61191	CUU51787	EFM94750	EFM96829	EFM98919	EFN01152	EFN03213
NarP	AWG94478	EFL79496	EFM88636	ABY68665	EFM90840	ABN73167	EFM93050	EFL80684	ACE60711	CUU51377	EFM95238	EFM97396	EFM99575	EFN01606	EFN03778
PhoR	AWG95686	EFL79284	EFM87389	ABY69823	EFM89558	ABN74342	EFM91701	EFL80810	ACE61957	CUU52575	EFM94002	EFM96071	EFM98389	EFN00361	EFN02524
PhoB	AWG95687	EFL79283 *	EFM87390 *	ABY69824 *	EFM89559 *	ABN74343 *	EFM91702 *	EFL80809 *	ACE61958 *	CUU52576 *	EFM94003 *	EFM96072 *	EFM98390 *	EFN00362 *	EFN02525 *
QseC	AWG95492	EFL78249	EFM86534	ABY69636	EFM89750	ABN74156	EFM91755	EFL81405	ACE61774	CUU52385	EFM94046	EFM96290	EFM98190	EFN00478	EFN02594
QseB	AWG95491	EFL78250	EFM86535	ABY69635	EFM89749	ABN74155	EFM91754	EFL81406	ACE61773	CUU52384	EFM94045	EFM96289	EFM98189	EFN00477	EFN02593

Table 3.1.1 Two Component Systems identified in *Actinobacillus pleuropneumoniae* Two-component system proteins coded in all sequenced strains of *A. pleuropneumoniae* accession number from all proteins are shown symbols used depict: † Hybrid histidine kinase, * Truncated protein, and n Gapped protein n equal the number of amino acids lost.

Serovar Strain	App6	18-1342	G1-9626	KL 16	NCTC10976	NCTC11383	NCTC11384	NCTC11407	S8	S4074
ArcB†	QSZ38965	MBT9319901	MBT9344736	ASU16956	VEJ16892	VTR36512	SQF64714	SUU64013	WP_009874855	AXA21452
ArcA	QSZ38013	MBT9318078	MBT9343212	ASU15085	VEJ15981	VTR32051	SQF63774	SUU58869	WP_005595668	AXA21118
CpxA	QSZ38636	MBT9318077	MBT9343213	ASU16600	VEJ16586	VTR62961	SQF64518	SUU62112	WP_005617129	AXA21117
CpxR	QSZ38637	MBT9318078	MBT9343212	ASU16599	VEJ16587 ₁₈	VTR62964	SQF64517	SUU62118	WP_058230464	AXA21118
NarQ	QSZ38451	MBT9319004	MBT9343837	ASU16411	VEJ16459	VTR45725	SQF64196	SUU61169	WP_005616989	AXA20953
NarP	QSZ38024	MBT9319707	MBT9344543	ASU15979	VEJ15992	VTR54715	SQF63784	SUU58930	WP_005616543	AXA20551
PhoR	QSZ39301	MBT9318669	MBT9343639	ASU15084	VEJ17254	VTR62995	SQF65027	SUU65933	WP_005617736	AXA21756
PhoB	QSZ39302	MBT9318668	MBT9343640	ASU15085	VEJ17255	VTR62996	SQF65028	SUU65971	WP_080515074	AXA21757
QseC	QSZ39102	MBT9320018	MBT9344878	ASU14884	VEJ17068	VTR34036	SQF64842	SUU65040	WP_005597899	AXA21563
QseB	QSZ39101	MBT9320019	MBT9344877	ASU14883	VEJ17067	VTR34042	SQF64841	SUU65036	WP_005619495	AXA21562

Table 3.1.1 Two Component Systems identified in *Actinobacillus pleuropneumoniae* Two-component system proteins coded in all sequenced strains of *A. pleuropneumoniae* accession number from all proteins are shown symbols used depict: † Hybrid histidine kinase, * Truncated protein, and n Gapped protein n equal the number of amino acids lost.

3.2 Prototypical interaction interface

The resolved co-crystal HK853-RR468[PDB:3DGE] from *Thermotoga maritima* MSB8 was used to infer the HK-RR interaction interface (Figure 3.2.1) (Casino et al., 2009; Guex & Peitsch, 1997; Johansson et al., 2012). The prototypical interaction interface (PII) was the framework for pairwise comparison between paralogue two-component systems, forty residues from both the histidine kinase and response regulator were predicted to be in the interface (Figure 3.2.1). These interfacing residues are used as proxy for the interaction interface between proteins upon formation of the transient complex during signal transduction. With these residues several symmetric matrices were built to assess specificity determining residues, these matrices were used in both methods ICI and SDMCEll with slightly different applications. Matrices were arranged according to the residues' relative positions assigned in the PII (Table 3.2.1).

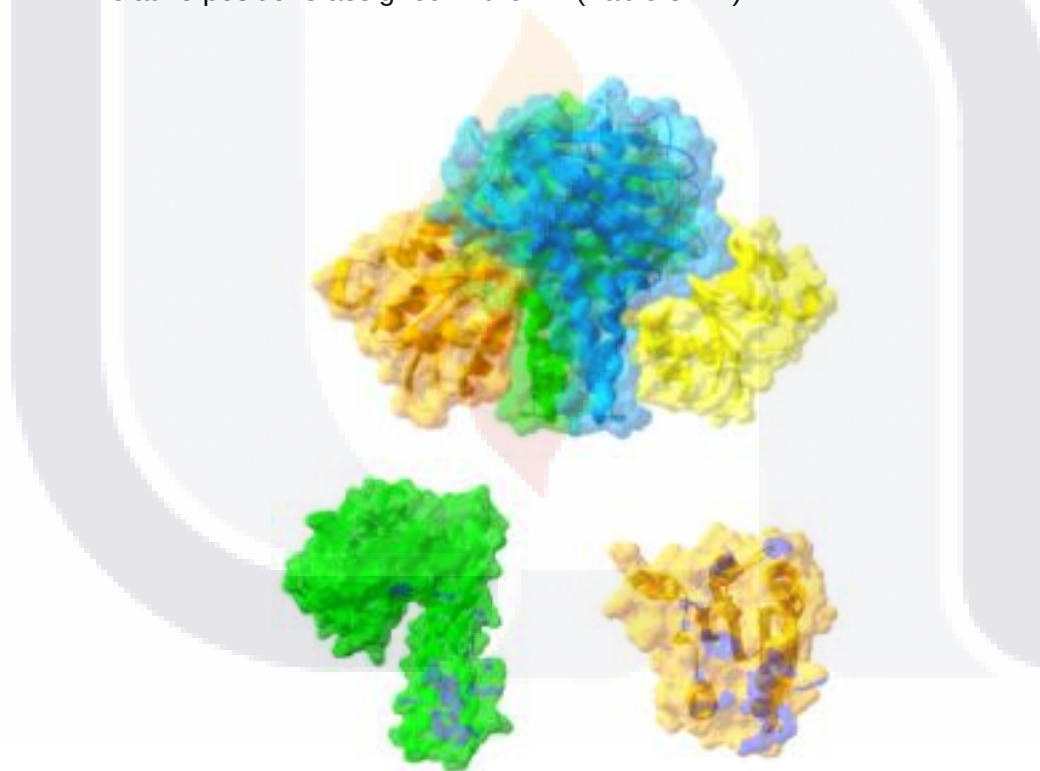


Figure 3.2.1 Histidine Kinase-Response Regulator Interaction Complex A) Tridimensional structure from HK853 RR468 complex from *Thermotoga maritima* MSB8[PDB:3DGE] [54]. Interaction interface is depicted in blue, non-interface in yellow. B) HK853 and C) RR468. The amino acid in blue were used to build matrices to calculate the co-evolution in the interaction interface and infer the specificity between TCS.

Although, the datasets analyzed were from the front facing heterodimer interface between the HK and RR (Figure 3.2.1), the tridimensional structure from HK853-RR468 shows that the HKs form a homodimer. Hence, one of the HKs is also facing backwards the RR in a

HK1-HK2-RR heterotrimer 2AB stoichiometry, therefore forming a secondary interaction interface between one of the RRs and the backward facing HK (Figure 3.2.1 A and 3.2.2).

	HK853::RR468		CpxA::CpxR			NarQ::NarP			PhoR::PhoB			QseC::QseB		
1	HIS	260	248	377	435	380	725	793	213	528	554	246	330	360
2	ARG	263	251	380	438	383	728	796	216	531	557	249	333	363
3	LEU	266	254	383	441	386	731	799	219	534	560	252	336	366
4	THR	267	255	384	442	391	732	800	220	535	561	253	337	367
5	ALA	268	256	385	443	392	733	801	221	536	562	254	338	368
6	LYS	270	258	387	445	394	735	803	223	538	564	256	340	370
7	ALA	271	259	388	446	395	736	804	224	539	565	257	341	371
8	TYR	272	260	389	447	396	737	805	225	540	566	258	342	372
9	GLU	274	262	391	449	398	739	807	227	542	568	260	344	374
10	THR	275	263	392	450	399	740	808	228	543	569	261	345	375
11	ASN	278	266	395	453	402	-	811	231	546*	572	264	348	378
12	SER	279	267	396	454	403	-	812	232^	547*	573	265^	349	379
13	GLU	282	270^	399	457	406^	-	815	233^	549*	577	-	356	386
14	LEU	283	-	400	458	407^	-	816	234^	552*	578	267^	357	387
15	THR	287	-	408*	466	412^	744	820	238	557	582	271^	361	391
16	GLU	290	272	411	469	421	747	823	241	560	585	274^	364	394
17	PHE	291	273	412	470	422^	748	824	242	561	586	275^	365	395
18	VAL	294	276	415	473	425^	751	827	245	564	589	278	368	398
19	GLN	298	280	419	477	429	767	846	249	568	593	282	372	402
20	LYS	387	367	511*	569	501	859	945	335^	681	697	370	466	496
21	TYR	437	412^	576	637	-	904*	1001	382^	777	790	410^	526	559
22	GLU	438	413^	577	638	534^	905*	1002	383	778	791	411	527	560
23	VAL	439	414^	578	639	535	906	1003	384	779	792	412	528	561
24	VAL	13	12	684	745	13^	1044	1141	13^	936	948	11	629	660
25	LEU	14	13	685	746	14^	1045	1142	14^	937	949	12	630	661
26	LYS	16	15	687	748	16^	1047	1144	16^	939	951	14	632	663
27	ILE	17	16	688	749	17^	1048	1145	17^	940	952	15	633	664
28	PHE	20	19	691	752	20	1051	1148	20^	943	955	18	636	667
29	ASN	21	20	692	753	21	1052	1149	21	944	956	19	637	668
30	LYS	24	23	695	756	24	1055	1152	24	947	959	22	640	671
31	MET	55	53	727	788	57	1088	1185	55	980	992	53	674	705
32	MET	56	54	728	789	58	1089	1186	56	981	993	54	675	706
33	PRO	57	55	729	790	59	1090	1187	57	982	994	55	676	707
34	ALA	84	79	755	818	84	1115	1214	84	1011	1023	80^	703	734
35	LYS	105	100	776	838	105^	1136	1235	105	1032	1044	101	724	755

36	PRO	106	101	777	840	106	1137	1236	106	1033	1045	102	725	756
37	PHE	107	102	778	841	107	1138	1237	107	1034	1046	103	726	757
38	SER	108	103	779	842	108	1139	1238	108	1035	1047	104	727	758
39	PRO	109	104	780	843	109	1140	1239	109	1036	1048	105	728	759
40	SER	110	105	781	844	110	1141	1240	110	1037	1049	106	729	760

Table 3.2.1 Prototypical interaction interface amino acid positions *Thermotoga maritima*'s HK853-RR468 residue positions that are in the predicted interface between the HK and RR proteins and its correlation to multiple sequence alignments' position of *App* TCS proteins. These relative positions 1 to 40 were used to calculate ICI and SDMCEII, the first 23 belong to the HK, the remaining to the RR. Grey columns contain residue position according to the reference sequenced used for the Perl script. The second and third columns contain residue overall positions according to the MSA either in the systems standalone alignments or the ones that contain also the *T. maritima*'s concatenated sequence, respectively. Chemical identities appear color coded: polar green, neutral purple, basic blue, acidic red, and hydrophobic black. Symbols to the right of the amino acid position represent * difference in residue frequencies between MSAs, ^ unaligned positions and - gapped positions.

We evaluated by SDMCEII if this secondary interaction interface had co-evolving residues, though, none of these positions Mlp values made it through the final 80th percentile cut-off. It was not assessed by the new ICI if this secondary interaction interface had co-evolving residues, hypothetically having additional specificity determining residues between cognate TCS interfaces. Even though, this secondary interface may be important for the HK-RR complex formation and molecular recognition, further research will be required to asses this possibility.

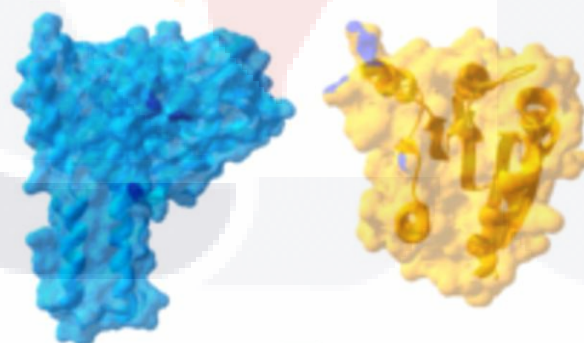


Figure 3.2.2 Histidine Kinase-Response Regulator Secondary Interaction Interface Tridimensional structure from HK853-RR468 complex from *T. maritima* MSB8 [PDB:3DGE]. Interaction interface is shown in dark blue A) HK853 backward facing second protomere in cyan and B) RR468 in orange, colors are in accordance to the whole complex (Figure 3.2.1). Interface amino acids from this secondary interaction interface may be important for the HK-RR complex formation and molecular recognition, though it wasn't assessed in the present work.

3.3 Amino acids determining specificity in non-Hybrid Two-Component Systems interface

Mutual information (MI) theory is often applied to predict positional correlations in a multiple sequence alignment, to make possible the analysis of those positions structurally or functionally important in a given fold or protein family. In order to identify co-evolving positions between putative cognate prototypical TCSs proteins, orthologue protein MSA datasets of prototypical TCSs coded in *App* were used to compute MIp. Which accurately estimates and removes background information from phylogeny and entropy, correctly identifying co-evolving positions (Dunn et al., 2008). The wMlp parameter introduced in the methodology allows to adjust for positions which are not completely represented (Figure 3.3.1).

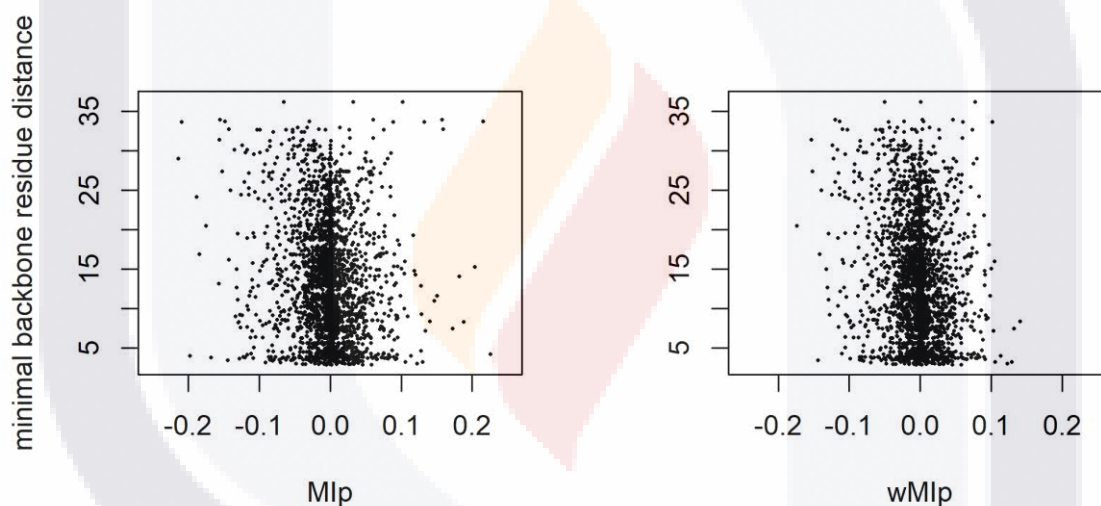


Figure 3.3.1 Dot blots of all computed values versus the minimal distance between residue pair's last side chain atoms from all four datasets Mutual information with average product correction (MIp) and weighted MIp.

3.3.1 Specificity amino acids SDMCEII

Computed MIp values and amino acids positions from the histidine kinase response regulator interaction interface were used to generate symmetric heatmap charts, reflecting MIp distance values of pairwise TCS comparing co-evolution between amino acids (Figure 3.4.2). The plotted values in the charts are the MIp distance between a pair of amino acids with corresponding positions (Table 3.2.1). Distance values were arranged in 10 different categories from zero to the highest value obtained from all pairwise calculated distances and filtered as described in methods. In the charts the relationship between a given pair of

systems can be seen, as the distance between corresponding amino acid pairs, as well as, distinct patterns emerging from each comparison (Figure 3.3.1.1).

Blank spaces in heatmaps or disconnected nodes are either due to gapped, completely conserved, below threshold, or missing positions (Figures 3.3.1.1, 3.3.1.11, and 3.3.1.12). The Mlp algorithm is unable to calculate completely conserved or missing positions, besides gapped positions are not considered (Crooks et al., 2004; Dunn et al., 2008). Positions that are missing or completely conserved are shown in the sequence logos (Figures 3.3.1.2 and 3.3.1.3).

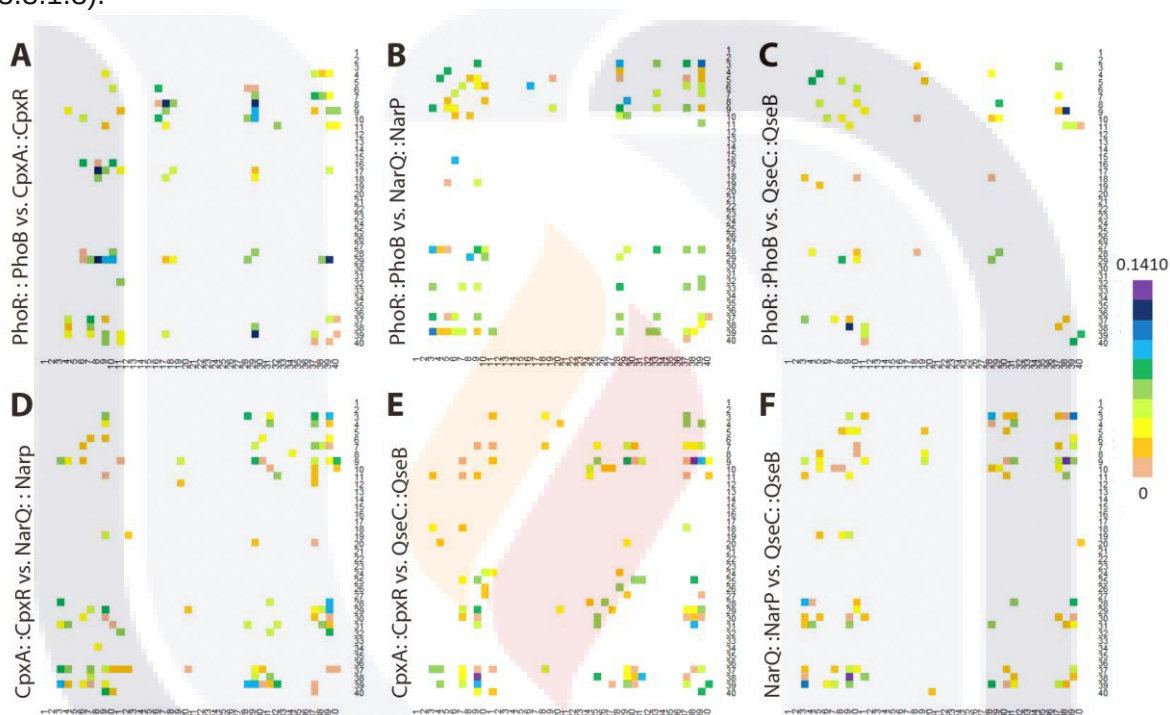


Figure 3.3.1.1 Histidine Kinase-Response Regulator Interaction Interface Mlp heatmaps Percentile 80 filtered heatmap charts of Mlp distance matrices between interaction interfaces from non-hybrid TCSs coded in *A. pleuropneumoniae*. Mlp distance values were classified in 10 categories from low to high Mlp distance values. The color key goes from blue to red, from high to low distance values; minimum and maximum Mlp values plotted are indicated. Symmetric matrices contain pairwise comparison between Mlp values from paralogous TCSs interaction interfaces; histidine kinases residues span from position 1 to 23 and response regulators from 24 to 40 in both axes. High distance values represent amino acid residues that confer specificity to cognate TCS between evaluated interfaces. Specificity between interfaces is correlated to the chemical properties from interface residues (Figure 3.4.3). A) PhoR::PhoB vs. CpxR::CpxA. B) PhoR::PhoB vs. NarQ::NarP. C) PhoR::PhoB vs. QseC::QseB. D) CpxA::CpxR vs. NarQ::NarP. E) CpxA::CpxR vs. QseC::QseB and F) NarQ::NarP vs. QseC::QseB.

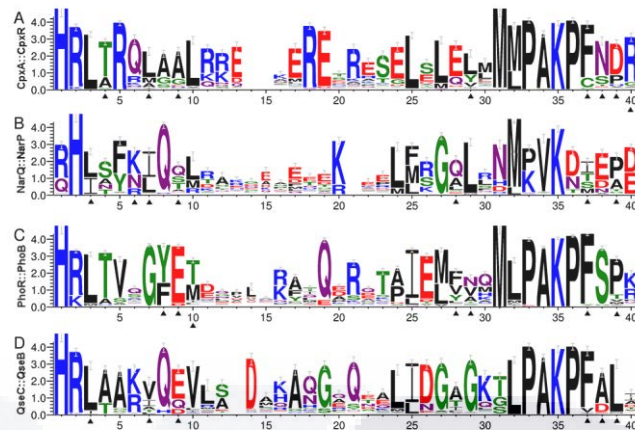


Figure 3.3.1.2 Sequence logos from parental TCSs MSAs Mlp matrices Sequence logo from Mlp matrices position corresponding to MSA from TCSs databases (Table 3.4.1), amino acid frequencies for each position sorted by chemical properties: polar green, neutral purple, basic blue, acidic red and hydrophobic black. A) CpxA::CpxR, B) NarQ::NarP, C) PhoR::PhoB and D) QseC::QseB.



Figure 3.3.1.3 Sequence logos from non-hybrid TCSs MSAs of *A. pleuropneumoniae* Sequence logo from MSA of TCSs coded in *A. pleuropneumoniae* position corresponding to Mlp matrices (Table 3.4.1), amino acid frequencies for each position sorted by chemical properties: polar green, neutral purple, basic blue, acidic red and hydrophobic black. A) CpxA::CpxR, B) NarQ::NarP, C) PhoR::PhoB and D) QseC::QseB.

The filtering threshold was chosen because Mlp distance values from the 80th percentile and forth did not follow a linear correlation after matrix pairwise comparison (Figures 3.3.1.4, 3.3.1.5, 3.3.1.6, 3.3.1.7, 3.3.1.8 and 3.3.1.9).

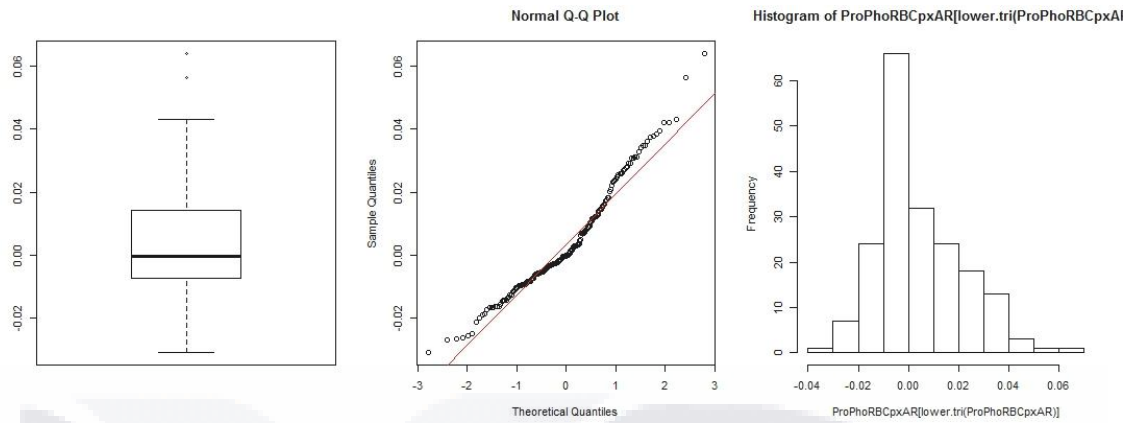


Figure 3.3.1.4 ProRB::CpxA Mlp matrix mean values distribution A) Box plot from lower triangle pair-wise Mlp mean values. B) Q-Q plot from lower triangle pair-wise Mlp mean values, the red line suggests normally distributed data. C) Histogram from lower triangle pair-wise Mlp mean values frequency.

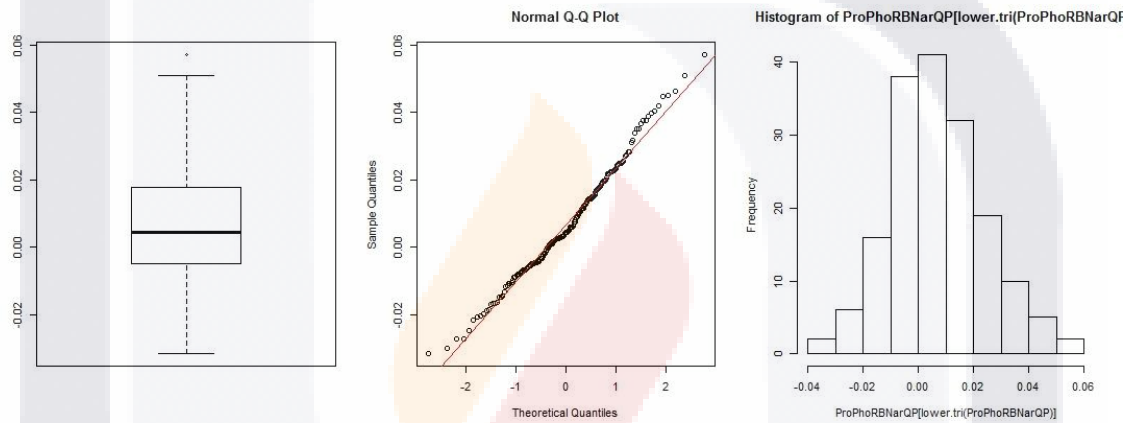


Figure 3.3.1.5 ProRB::NarQP Mlp matrix mean values distribution A) Box plot from lower triangle pair-wise Mlp mean values. B) Q-Q plot from lower triangle pair-wise Mlp mean values, the red line suggests normally distributed data. C) Histogram from lower triangle pair-wise Mlp mean values frequency.

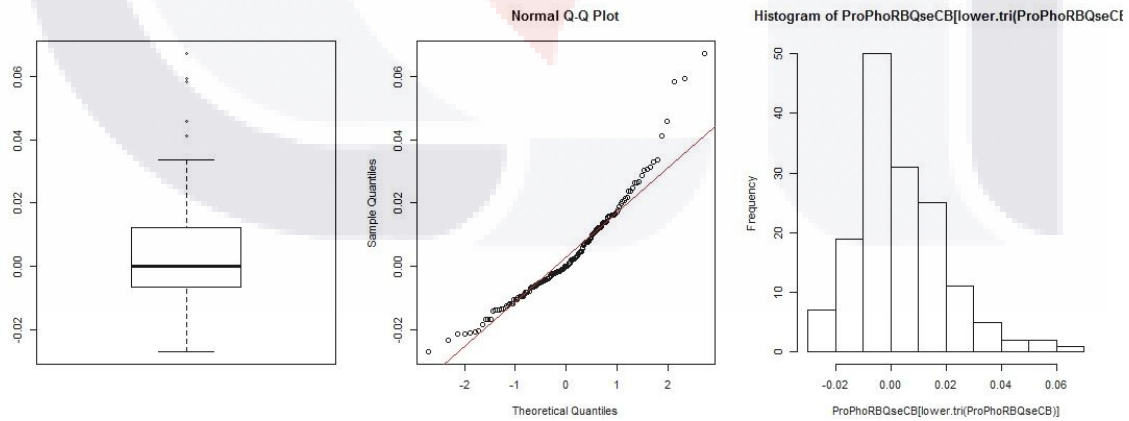


Figure 3.3.1.6 ProRB::QseCB Mlp matrix mean values distribution A) Box plot from lower triangle pair-wise Mlp mean values. B) Q-Q plot from lower triangle pair-wise Mlp mean values, the red line suggests normally distributed data. C) Histogram from lower triangle pair-wise Mlp mean values frequency.

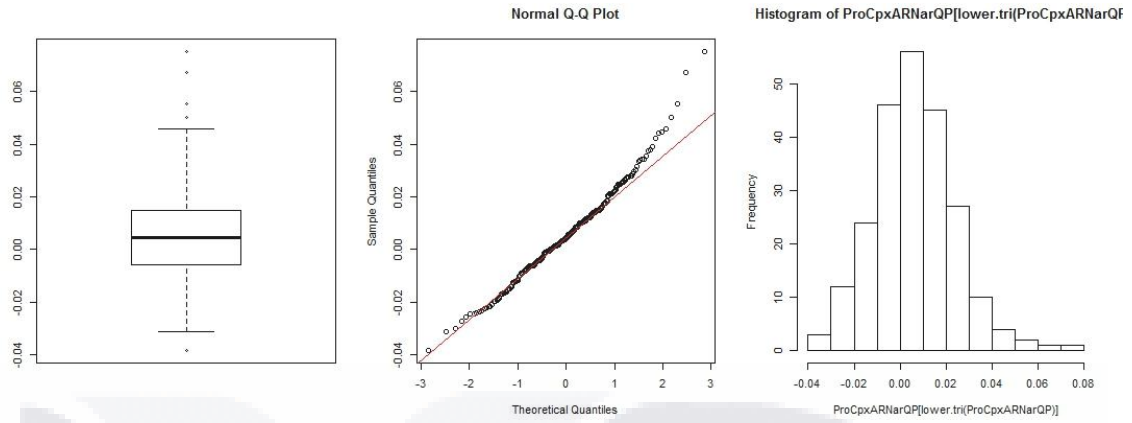


Figure 3.3.1.7 CpxAR::NarQP Mlp matrix mean values distribution A) Box plot from lower triangle pair-wise Mlp mean values. B) Q-Q plot from lower triangle pair-wise Mlp mean values, the red line suggests normally distributed data. C) Histogram from lower triangle pair-wise Mlp mean values frequency.

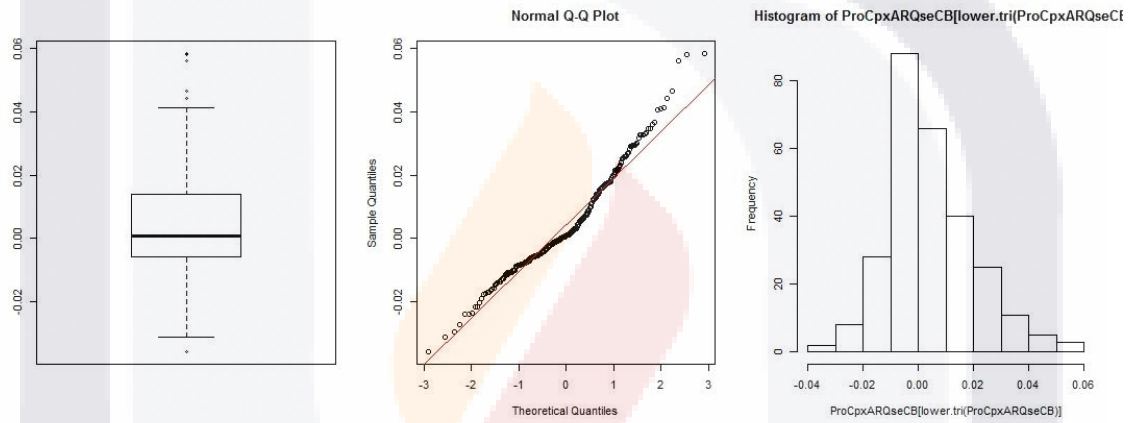


Figure 3.3.1.8 CpxAR::QseCB Mlp matrix mean values distribution A) Box plot from lower triangle pair-wise Mlp mean values. B) Q-Q plot from lower triangle pair-wise Mlp mean values, the red line suggests normally distributed data. C) Histogram from lower triangle pair-wise Mlp mean values frequency.

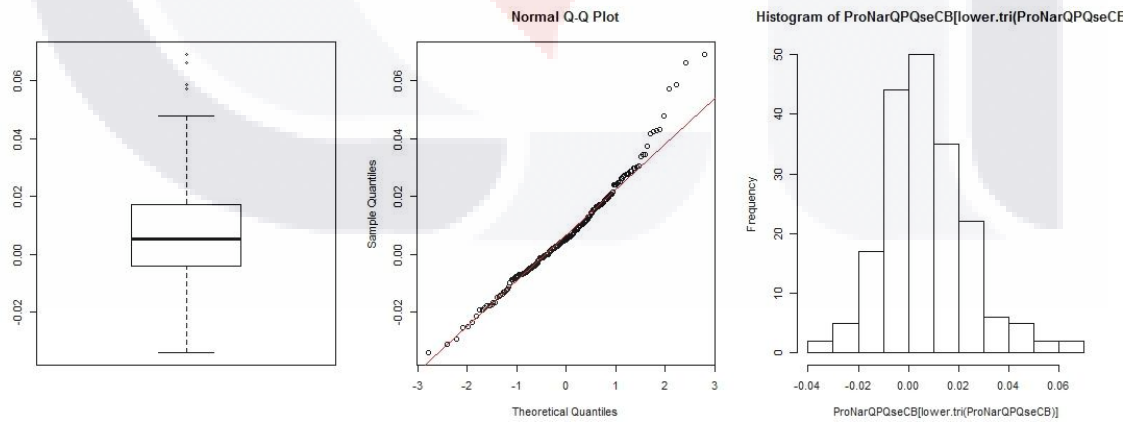


Figure 3.3.1.9 NarQP::QseCB Mlp matrix mean values distribution A) Box plot from lower triangle pair-wise Mlp mean values. B) Q-Q plot from lower triangle pair-wise Mlp mean values, the red line suggests normally distributed data. C) Histogram from lower triangle pair-wise Mlp mean values frequency.

Pattern distribution among systems forms clusters of residues with varying Mlp distance values, moreover, cluster formation and distance values are pairwise specific. Pairwise comparisons, keeping one of the systems fixed, allowed us to identify three distinct sets of

high distance Mlp values among paralogue TCSs' interface matrices. One set of co-evolving residues is present in all three pairwise comparisons, hence it was labeled orthologue interface specificity core cluster (OISCC). The identified OISCC of the TCSs analyzed herein have 10, 7, 5, and 5 residue pairs, from systems CpxA-CpxR, NarQ-NarP, PhoR-PhoB, and QseC-QseB, respectively (Figures 3.3.1.1 and 3.3.1.11 and Tables 3.3.1.1 and 3.3.1.2). The second set is present in more than one, though not in all the matrices of a given TCS, and defined as interface specificity accessory cluster (ISAC); whereas the last is only present in individual datasets, and was named as interface specificity accessory residues (ISAR).

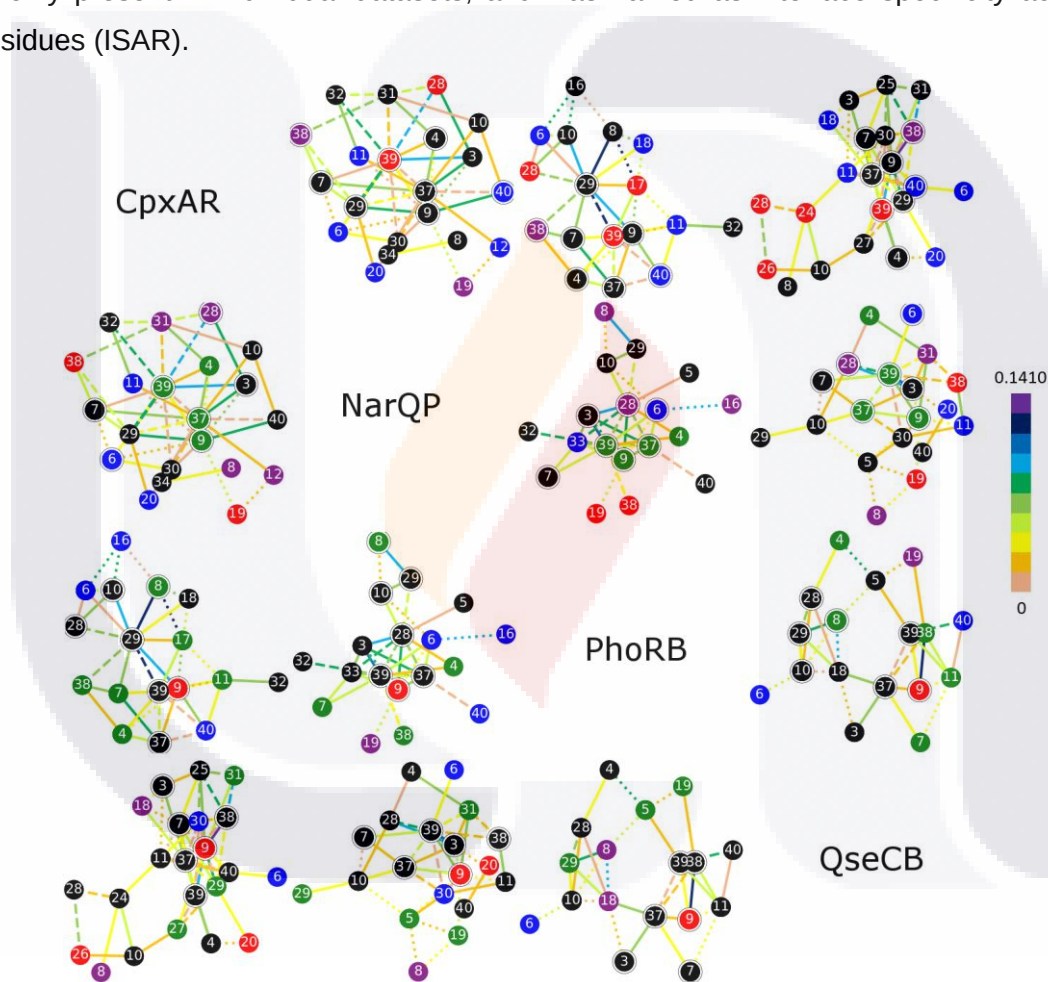


Figure 3.3.1.10 Pairwise undirected graphs of interaction interfaces' residue network. Mlp distance value are depicted as colored edges according to absolute distance value classes, and are either dotted, solid, or dashed, whether they are HK intraprotein, HK-RR interprotein, or RR intraprotein interactions, respectively. Nodes are represented as circles with relative positions and colored by chemical properties as green, purple, blue, red, and black, corresponding to polar, neutral, basic, acidic, and hydrophobic, respectively. Chemical properties were taken from the sequence logo of *App* TCSs, and are the same for each row. Nodes belonging to the OISCC are highlighted by a white ring. Positions are in accordance to the amino acid alignment from *App* TCSs to the model interaction interface (Table 3.2.1).

Independently, the four pairwise tested datasets share two residue pairs which are present among OISCCs, namely residue pairs 9-37 and 9-39. The chemical identities from those

residue pairs differ in all but PhoR-PhoB's and QseC-QseB's position 9-37. Interestingly enough the co-evolution distance value among paralogue systems is also different in 9-37 0.05213, 0.01817, 0.00682, 0.0703, 0.04531, and 0.02499, for pairwise comparisons CpxAR-NarQP, CpxAR-PhoRB, CpxAR-QseCB, NarQP-PhoRB, NarQP-QseCB, and PhoRB-QseCB, respectively. While for the 9-39 pair the values are 0.03873, 0.06119, 0.09608, 0.02246, 0.05735, and 0.03489, for CpxAR-NarQP, CpxAR-PhoRB, CpxAR-QseCB, NarQP-PhoRB, NarQP-QseCB, and PhoRB-QseCB, respectively. Moreover, the second lowest value of 9-37 is from a comparison where the residues have the same identities (Figures 3.3.1.1 and 3.3.1.11 and Tables 3.3.1.1 and 3.3.1.2).

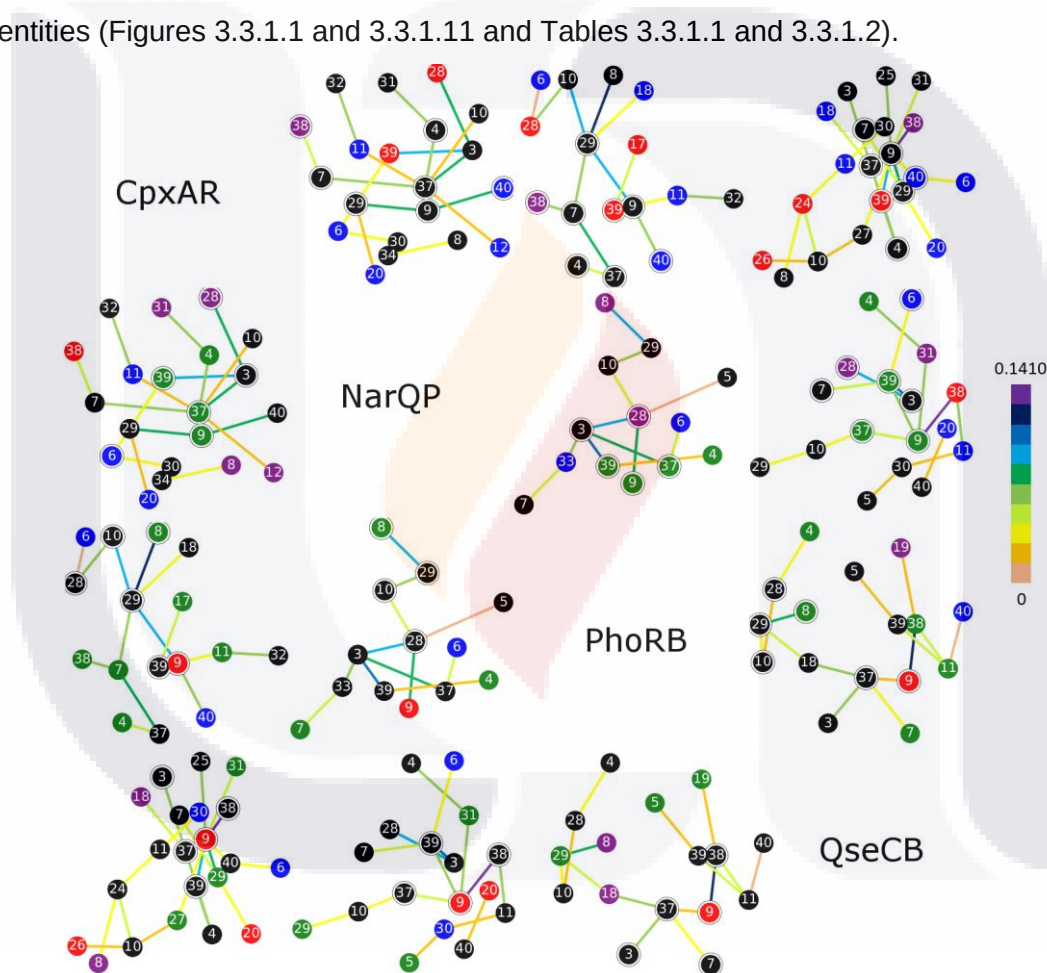


Figure 3.3.1.11 Pairwise maximum spanning tree interaction interfaces' residue network. Mlp distance value are depicted as colored edges according to absolute distance value classes, and are either dotted, solid, or dashed, if they are HK intraprotein, HK-RR interprotein, or RR intraprotein interactions, respectively. Nodes are represented as circles with relative positions; and colored by chemical properties as green, purple, blue, red, and black corresponding to polar, neutral, basic, acidic, and hydrophobic, respectively. Chemical properties were taken from the sequence logo of *App* TCSs, and are the same for each row. Nodes belonging to the OISCC are highlighted by a white ring. Positions are in accordance to the amino acid alignment from *App* TCSs to the model interaction interface as in (Figures 3.2.1, 3.3.1.2, and 3.3.1.2, and Table 3.2.1).

CpxA::CpxR					NarQ::NarP					PhoR::PhoB					QseC::QseB				
x	y	aa _x	aa _y	$\overline{MI}_p$	x	y	aa _x	aa _y	$\overline{MI}_p$	x	y	aa _x	aa _y	$\overline{MI}_p$	x	y	aa _x	aa _y	$\overline{MI}_p$
-	-	-	-	-	3	28	LEU	GLN	0.08561	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	3	37	LEU	ILE	0.05644	-	-	-	-	-	3	3	LEU	PHE	0.03790
-	-	-	-	-	3	39	LEU	PRO	0.09944	-	-	-	-	-	-	-	-	-	
4	37	THR	PHE	0.05611	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
4	39	THR	ASP	0.03839	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	6	39	LYS	PRO	0.03588	-	-	-	-	-	-	-	-	-	
7	29	LEU	LEU	0.05567	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
7	37	LEU	PHE	0.05838	-	-	-	-	-	-	-	-	-	-	7	3	VAL	PHE	0.03455
7	38	LEU	ASN	0.04944	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
7	39	LEU	ASP	0.03091	7	39	ILE	PRO	0.03899	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	8	29	TYR	ASN	0.09586	-	-	-	-	
9	29	ALA	LEU	0.07788	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
9	37	ALA	PHE	0.02571	9	37	GLN	ILE	0.05591	9	37	GLU	PHE	0.03782	9	3	GLU	PHE	0.02571
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	3	GLU	ALA	0.13323
9	39	ALA	ASP	0.06533	9	39	GLN	PRO	0.03951	9	39	GLU	PRO	0.03951	9	3	GLU	LEU	0.06277
9	40	ALA	ARG	0.06161	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	10	28	THR	PHE	0.04622	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	10	29	THR	ASN	0.05976	-	-	-	-	
-	-	-	-	-	28	3	GLN	LEU	0.08561	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	28	10	PHE	THR	0.04622	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	29	8	ASN	TYR	0.09586	-	-	-	-	
29	7	LEU	LEU	0.05567	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
29	9	LEU	ALA	0.07788	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	29	10	ASN	THR	0.05976	-	-	-	-	
29	38	LEU	ASN	0.04873	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
29	39	LEU	ASP	0.08776	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	37	3	ILE	LEU	0.05644	-	-	-	-	-	37	3	PHE	LEU	0.04633
37	4	PHE	THR	0.05611	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
37	7	PHE	LEU	0.05838	-	-	-	-	-	-	-	-	-	-	37	7	PHE	VAL	0.03455
37	9	PHE	ALA	0.02571	37	9	ILE	GLN	0.05591	37	9	PHE	GLU	0.03782	37	9	PHE	GLU	0.02571
37	40	PHE	ARG	0.01480	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
38	7	ASN	LEU	0.04944	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	38	9	ALA	GLU	0.13323
38	29	ASN	LEU	0.04873	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	39	3	PRO	LEU	0.09944	-	-	-	-	-	-	-	-	-	
39	4	ASP	THR	0.03839	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	-	39	6	PRO	LYS	0.03588	-	-	-	-	-	-	-	-	-	
39	7	ASP	LEU	0.03091	39	7	PRO	ILE	0.03899	-	-	-	-	-	-	-	-	-	
39	9	ASP	ALA	0.06533	39	9	PRO	GLN	0.03951	39	9	PRO	GLU	0.03951	39	9	LEU	GLU	0.06277
40	9	ARG	ALA	0.06161	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Table 3.3.1.1 $\overline{MI}_p$ mean values of orthologue interface's specificity positions Amino acid pairs clusters from each two-component system compared. The numeration is discontinuous as it only reflects those amino acid pair that were present in all pairwise test comparisons. All positions belong to the interaction interface (Table 3.2.1). Amino acid identities and chemical properties were taken from Figure 3.3.1.2. Average mutual information ($\overline{MI}_p$) of pair comparisons was calculated from their corresponding dataset.

CpxA::CpxR				NarQ::NarP				PhoR::PhoB				QseC::QseB			
x	y	aa _x	aa _y	x	y	aa _x	aa _y	x	y	aa _x	aa _y	x	y	aa _x	aa _y
-	-	-	-	3	28	LEU	GLN	-	-	-	-	-	-	-	-
-	-	-	-	3	37	LEU	SER	-	-	-	-	3	37	LEU	PHE
-	-	-	-	3	39	LEU	THR	-	-	-	-	-	-	-	-
4	37	ALA	PHE	-	-	-	-	-	-	-	-	-	-	-	-
4	39	ALA	ASP	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	6	39	LYS	THR	-	-	-	-	-	-	-	-
7	29	LEU	LEU	-	-	-	-	-	-	-	-	-	-	-	-
7	37	LEU	PHE	-	-	-	-	-	-	-	-	7	37	ILE	PHE
7	38	LEU	ASN	-	-	-	-	-	-	-	-	-	-	-	-
7	39	LEU	ASP	7	39	ILE	THR	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	8	29	TYR	PHE	-	-	-	-
9	29	ALA	LEU	-	-	-	-	-	-	-	-	-	-	-	-
9	37	ALA	PHE	9	37	SER	SER	9	37	GLU	PHE	9	37	GLU	PHE
-	-	-	-	-	-	-	-	-	-	-	-	9	38	GLU	ALA
9	39	ALA	ASP	9	39	SER	THR	9	39	GLU	PRO	9	39	GLU	LEU
9	40	ALA	ARG	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	10	28	LEU	LEU	-	-	-	-
-	-	-	-	-	-	-	-	10	29	LEU	PHE	-	-	-	-
-	-	-	-	28	3	SER	LEU	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	28	10	LEU	LEU	-	-	-	-
-	-	-	-	-	-	-	-	29	8	PHE	TYR	-	-	-	-
29	7	LEU	LEU	-	-	-	-	-	-	-	-	-	-	-	-
29	9	LEU	ALA	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	29	10	PHE	LEU	-	-	-	-
29	38	LEU	ASN	-	-	-	-	-	-	-	-	-	-	-	-
29	39	LEU	ASP	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	37	3	SER	LEU	-	-	-	-	37	3	PHE	LEU
37	4	PHE	ALA	-	-	-	-	-	-	-	-	-	-	-	-
37	7	PHE	LEU	-	-	-	-	-	-	-	-	37	7	PHE	ILE
37	9	PHE	ALA	37	9	SER	SER	37	9	PHE	GLU	37	9	PHE	GLU
37	40	PHE	ARG	-	-	-	-	-	-	-	-	-	-	-	-
38	7	ASN	LEU	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	38	9	ALA	GLU
38	29	ASN	LEU	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	39	3	THR	LEU	-	-	-	-	-	-	-	-
39	4	ASP	ALA	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	39	6	THR	LYS	-	-	-	-	-	-	-	-
39	7	ASP	LEU	39	7	THR	ILE	-	-	-	-	-	-	-	-
39	9	ASP	ALA	39	9	THR	SER	39	9	PRO	GLU	39	9	LEU	GLU
40	9	ARG	ALA	-	-	-	-	-	-	-	-	-	-	-	-

Table 3.3.1.2 Orthologue interface specificity core clusters in *A. pleuropneumoniae* Putative orthologue interface specificity cluster in *A. pleuropneumoniae*, responsible of TCS system wide discrimination. Allowing cognate specific recognition and non-cognate components insulation. The numeration is discontinuous and only reflects those amino acid pair that were present in all pairwise test comparisons. All positions belong to the interaction interface (Table 3.2.1). Amino acid identities and chemical properties were taken from Figure 3.3.1.3.

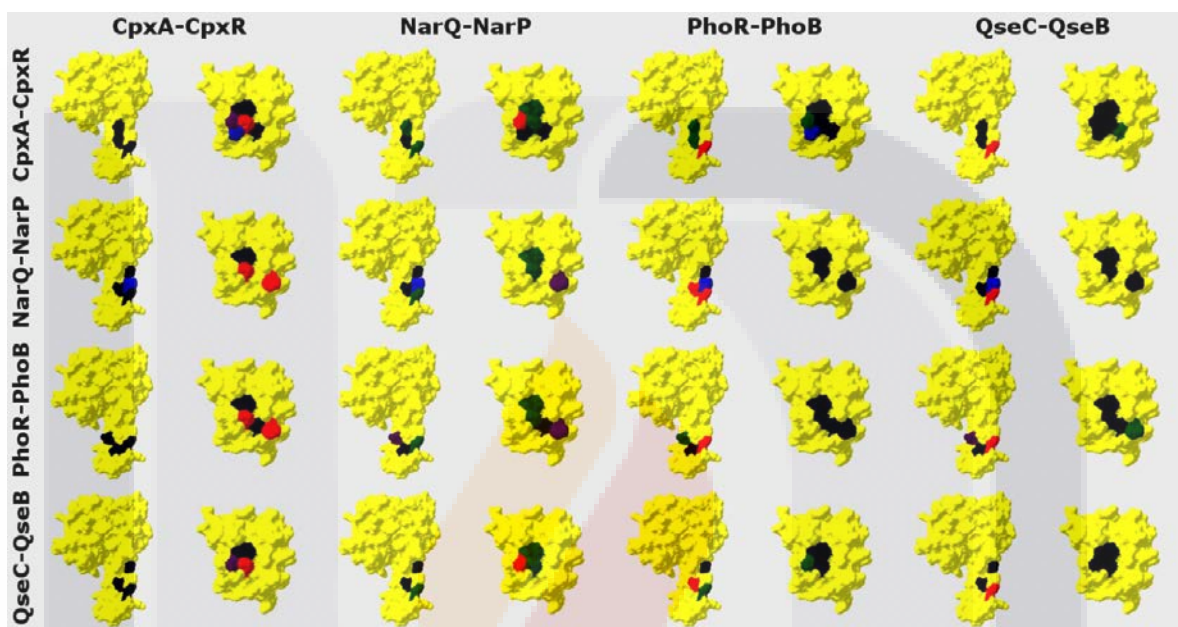


Figure 3.3.1.12 OISCC spatial localization Protein surface shown in yellow, chemical properties depicted as follows: polar, neutral, basic, acidic, and hydrophobic, appear as green, purple, blue, red, and black, respectively. Amino acid localizations are in accordance to the relative positions from App TCSs' alignment to the HK853-RR468 complex (Figures 3.2.1 and 3.3.1.10 and Tables 3.2.1 and 3.3.1.2).

3.3.2 Interface coupling index

The interface coupling index (ICI) estimates both co-evolution and close contact for residue pairs, the 99 percent cutoff forms clusters for each system (Figure 3.3.2.1). The pattern distribution among systems forms clusters of neighboring residue pairs with varying ICI values (Figure 3.3.2.2 and Table 3.3.2.1). Cluster formation is system specific and amino acid chemical identity and frequency (Figures 3.3.2.3 and 3. 3.2.4). Position 35 is the only chemical identity completely conserved throughout systems, and is shared along OmpR and NarL response regulator families. OmpR family's response regulators share the consensus PAKPF on positions 33 to 37 (Figures 3.3.2.3 and 3. 3.2.4), which are spatially nearby to the phosphoacceptor aspartate in the RR's tridimensional model. Even though some residue positions are completely conserved in different systems' PII, its individual or

cumulative contribution cannot be assessed directly by methods that rely on mutual information theory, for it is an inherent liability in its calculations.

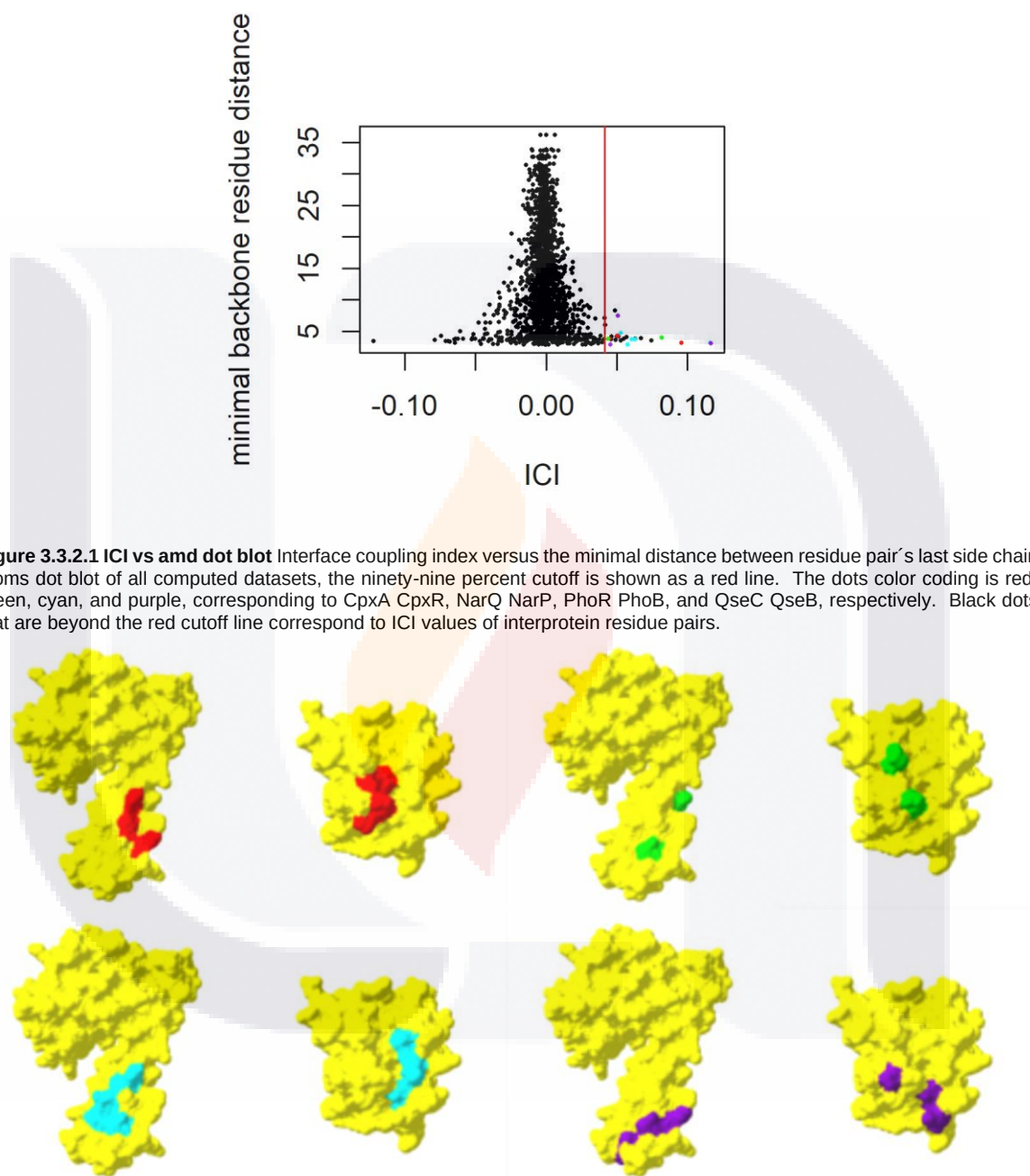


Figure 3.3.2.1 ICI vs and dot blot Interface coupling index versus the minimal distance between residue pair's last side chain atoms dot blot of all computed datasets, the ninety-nine percent cutoff is shown as a red line. The dots color coding is red, green, cyan, and purple, corresponding to CpxA CpxR, NarQ NarP, PhoR PhoB, and QseC QseB, respectively. Black dots that are beyond the red cutoff line correspond to ICI values of interprotein residue pairs.

Figure 3.3.2.2 OISCC spatial localization with ICI Spatial localization of the determined OISCC in the surface from histidine kinases and response regulators relative to the prototypical interaction interface, all systems present in *Actinobacillus pleuropneumoniae* are depicted (A) CpxA CpxR, (B) NarQ NarP, (C) PhoR PhoB, and (D) QseC QseB. The exact chemical composition and ICI values for all OISCC can be found in Table 3.3.2.1. Response regulators are not in scale in relation to histidine kinases and are shown only to depict interface correlation.

	MSA x	MSA y	AAx	AAy	PII a	PII y	AAx	AAy	ICI	Distance nm
CpxAR	384	778	T	F	4	37	A	F	0.0450	3.7495
	388	778	L	F	7	37	L	F	0.0509	4.2624
	391	780	A	D	9	39	A	D	0.0957	3.1714
	391	781	A	R	9	40	A	R	0.0504	4.2983
NarQP	731	1137	L	D	3	36	L	D	0.0818	3.9912
	740	1140	I	P	10	39	I	T	0.0437	3.8165
PhoRB	535	937	T	I	4	25	T	-	0.0629	3.6727
	536	937	V	I	5	25	V	-	0.0530	4.6903
	540	940	Y	M	8	27	Y	M	0.1161	3.1885
	543	944	T	N	10	29	L	F	0.0577	2.8786
	561	940	A	M	17	27	S	M	0.0607	3.7061
QseCB	344	727	E	A	9	38	E	A	0.1167	3.0223
	345	637	V	G	10	29	I	G	0.0456	2.8786
	357	640	D	K	14	30	D	K	0.0510	7.4477

Table 3.3.2.1 OISCC amino acid pairs

The conserved histidine from HisKA and HisKA_3 domain families align differently to the *T. maritima*'s co-crystal HK258-RR245, thus the relative position of NarQ/NarX residues in the inferred PII differs to those from the other 3 HKs that belong to HisKA family. In HisKAs the histidine appears on position 1 (Grebe & Stock, 1999), while on HisKA_3 it is in position 2 (Figures 3.4.3 and 3.4.4).

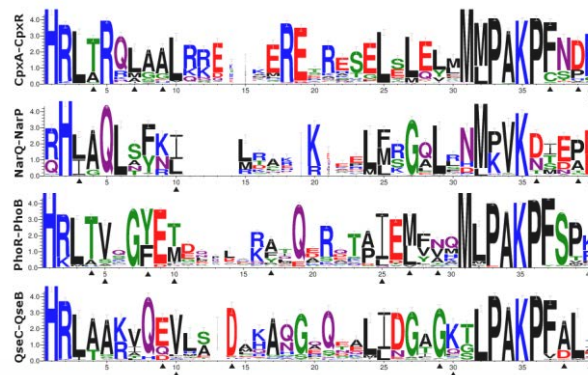


Figure 3.3.2.3 Sequence logo from parental TCSs' MSAs Sequence logo from Mlp matrices position corresponding to MSA from TCSs databases (Table 3.2.1), amino acid frequencies for each position sorted by chemical properties: polar green, neutral purple, basic blue, acidic red and hydrophobic black. A) CpxA::CpxR, B) NarQ::NarP, C) PhoR::PhoB and D) QseC::QseB.



Figure 3.3.2.4 Sequence logo from non-hybrid *A. pleuropneumoniae*'s TCSs MSAs Prototypical interaction interface relative positions from *A. pleuropneumoniae* according to HK853-RR468 (Table 3.2.1). Triangles below y-axes highlight positions that belong to the orthologue interaction specificity core (Figures 3.2.1 and 3.3.2.2). Amino acid frequencies are in bits, histidine kinase's positions go from 1 to 23 and response regulator from 24 to 40. Polar, neutral, basic, acidic, and hydrophobic, amino acid chemical properties appear as green, purple, blue, red, and black, respectively.

DISCUSSION

Two-component systems' signal transduction is based in the formation of a transient complex between Histidine kinases (HKs) and Response regulators (RRs), where an interaction interface forms between heterodimers. Pathway specificity is achieved by different mechanisms: HK phosphatase activity over its cognate RR due to non-cognate activation (Fisher, Jiang, Wanner, & Walsh, 1995; Gao & Stock, 2010b; Huynh & Stewart, 2011a), substrate competition between cognate and non-cognate RRs (Huynh, Chen, & Stewart, 2015b; Huynh & Stewart, 2011a; Laub & Goulian, 2007; Siryaporn, Perchuk, Laub, & Goulian, 2010a; Typas & Sourjik, 2015), spatial-temporal localization from cognate components or regulatory elements (Batchelor & Goulian, 2006; Jensen, Wang, & Shapiro, 2002b; Ryan & Shapiro, 2003; Sommer, Koler, Frank, Sourjik, & Vaknin, 2013; Sourjik & Armitage, 2010; Ventura & Sourjik, 2011), and the primary mechanism to which all

TESIS TESIS TESIS TESIS TESIS

aforementioned mechanisms rely on molecular recognition between cognate HK and RR (Capra & Laub, 2012; Capra et al., 2010; Capra, Perchuk, Skerker, & Laub, 2012; Podgornaia, Casino, Marina, & Laub, 2013; Podgornaia & Laub, 2013; Skerker et al., 2008). Molecular recognition is the intrinsic ability of HKs and RRs to recognize its cognate partner through interface interactions. Hence, it was rationalized that orthologue cognate interface composition can be thoroughly analyzed for co-evolving and close contact amino acid residues, inferring TCS system-wide insulation and intra-system selectivity. TCS specificity is then, the capability of a system to effectively recognize its cognate component over non-cognate substrates.

Understanding precisely how proteins discriminate between cognate and non-cognate partners at the molecular level has remained elusive. Solid evidence supports molecular recognition as the primary force for cognate proteins' specific recognition, though the fine underlying effect of the residues that results in specificity has not been thoroughly established. Measuring MI over multiple sequence alignments built by great numbers of protein sequences from diverse paralogue proteins from a single family creates a sample selection bias, in which the composition of the MSA overrepresents some proteins families relative to others. Therefore, the results are suitable for the whole protein family, though lack the sensitivity to distinguish differences among paralogous systems.

The approaches here envisioned attempt to pinpoint specificity determining residues in the interaction interface throughout two-component systems' paralogous families. These approaches share the same initial preprocessing, relying on MSAs assembled only by orthologous TCSs' sequences and a Prototypical interaction interface (PII) obtained from structural data. The co-evolution values were computed from whole HK-RR concatenated sequences and correlated with structural information, to comprise all information available and preserve the *in vivo* relationship between interacting proteins. After these initial steps both methodologies branch and make use of data acquired in different ways. The Specificity determining matrices of co-evolving interaction interface method uses a combinatorial strategy of pairwise testing of matrices to predict important residues in the interface, whereas the Interface coupling index uses the minimal distance among residues in the interface to implement an standardize parameter that infer relevant residues in the interface.

4.1 Pair wise testing

High-distance values in matrix pairwise comparison between paralogous TCSs interaction interfaces can be defined as residues that confer specificity to orthologous proteins and

insulates paralogous system components. While, low-distance values belong to highly variable or redundant residue positions, not required for cognate proteins recognition nor non-cognate insulation. Matrices positions of co-evolving amino acids from the HK-RR interaction interface between paralogous TCSs form specificity clusters of spatial neighboring residues with varying Mlp distance values (Figures 3.3.1.1 and 3.3.1.12 and Tables 3.2.1 and 3.3.1.1). Clusters composition is pairwise specific and data analysis suggests interaction interface specificity is driven by the cumulative effect of neighboring residues, also the conformation of structural pockets of co-evolving residue pairs must be taken into consideration (Figure 3.3.1.12).

The results of pairwise matrix comparison of co-evolution distance values and interface amino acid position, suggest that interface residues can be classified in three components which confer different classes of specificity between cognate TCSs proteins. System wide insulation is achieved by high distance Mlp value residue positions that are present in all pairwise compared matrices, these are defined as Orthologue interface specificity core cluster (Figures 3.3.1.1, 3.3.1.2, and 3.3.1.3). This cluster endows the cognate interaction interface of intrasystem specificity recognition and system-wide paralogue intersystem insulation. OISCCs act as a one-to-all insulation and one-to-one selective recognition component. The Interface specificity accessory cluster provides the second tier of specificity recognition, this encloses high distance Mlp value residue positions that are present in two or more pairwise comparisons but not in all, this cluster provides one-to-many discrimination. The last tier, the Interface specificity accessory residues is made of high distance Mlp residues that are only present between two given TCSs, and confers one-to-one intra-system selective recognition.

These OISCCs vary in quantity, tridimensional localization, chemical identities, and co-evolution values of its residue pairs, emerging as fingerprints for molecular recognition. The synergic and context dependent interaction from these three interface components leads to an exquisite molecular recognition that discriminates between cognate and non-cognate substrates. Here, we suggest that the gross system-wide insulating capability relies on the OISCC, as it is rationalized that changes in the ISAC or ISAR composition would not thoroughly disrupt molecular recognition or discrimination among substrates. Nevertheless, the second and third interface components further contribute to improve the molecular recognition capability, fine-tuning cognate specificity recognition and non-cognate

insulation from the interaction interface, a view shared by Capra *et al*, 2010 (Capra *et al*., 2010).

Even though these assumptions seemed to hold true at the time, once the datasets were further processed with the new algorithm that takes into account all positions independently of its occupancy two unexpected caveats arose. The second and third tiers of discrimination disappeared and the number of positions with high distance values increased drastically. After careful review of the new results, it was clear that the formation of the second and third tier clusters was an artifact due to the way in which the mutual information was computed by the first algorithm, only completely occupied positions were used for the calculations. Once that all positions were available in the matrices pairwise comparison no longer had unpopulated matrix's positions, hence these two tiers never existed. The drastic increase in the number of high distance value positions arises from the fact that all positions are now available for calculation, but due to the way in which the distance value is calculated. The values used was the weighted mutual information (Equation 2.3.6.2), therefore a lot of the positions from the PII (Figure 3.2.1 and Table 3.2.1) that were not completely occupied showed low wMIp values. Once these values were used to run calculation (Equation 2.3.6.1) a lot of positions that previously did not had a value now had high distance values, making worthless the approach with the new algorithm written. Even though, some of the results were in accordance with previously published findings, most of the assumptions were then devoid of sense. Thus, the approach was deprecated and a new approach had to be envisioned.

4.2 Interaction interfaces' discriminating components

High ICI values are dealt as interaction interface's residue pairs that confer specificity to orthologous cognate proteins and insulate paralogous system components. While, low values are highly variable, redundant, or non-contacting residue positions, not required for cognate protein recognition nor non-cognate insulation. Specificity clusters are formed by spatial neighboring residue pairs in the interaction interface with varying ICI values (Figure 3.3.2.2 and Table 3.3.2.1). Clusters composition is system specific and data suggests specificity in the interaction interface is driven by a cumulative effect of the neighboring residues through electrostatic, size, hydrophilic, and/or hydrophobic interaction and/or structural pockets with inter- and intraprotein associations (Figures 3.3.2.3 and 3.3.2.4 and Table 3.3.2.1). This cumulative effect, in order to achieve effective pathway insulation is clearly seen in the single, double and triple point mutation made to PhoR from *Caulobacter*

crescentus A218T, S219V and F223Y, these mutations disrupt system wide insulation between PhoR phosphorylating non cognate response regulator together with cognate PhoB (Capra et al., 2012). The involvement of intraprotein residues in specificity is outside the scope of the present work, though it would be interesting to assess whether these belong to structural pockets in which all components interact or if their association are non-essential, this will be worked upon in the future.

System wide discrimination is achieved by high ICI value residue pairs which are defined as the Orthologue interface specificity core (Table 3.3.2.1), the concept was borrowed from the previous approach because it accurately described the role of these specificity endowing pair residues. This clusters endow the cognate interaction interface of intrasystem specific molecular recognition and system-wide paralogue intersystem insulation, acting as a one-to-all insulation and one-to-one selective recognition component. OISCCs composition vary in quantity, spatial location, chemical identity, and ICI values of residue pairs, emerging as fingerprints for molecular recognition (Figure 3.3.2.2 and Table 3.3.2.1). The synergic and context dependent interaction from these interface components leads to an exquisite molecular recognition that discriminates between cognate and non-cognate substrates. Hence is suggested that the gross system-wide insulating capability relies on the OISCC composition among paralogue systems.

4.3 System-wide discrimination

Predicted pairs of residues with high ICI values are located on the $\alpha 1$ and $\alpha 5$ - $\beta 5$ loop from the RRs and $\alpha 1$ and $\alpha 1$ - $\alpha 2$ loop from the HKs, these findings are in accordance with previously published experimental and computational data (Capra et al., 2010; Podgornaia et al., 2013; Siryaporn, Perchuk, Laub, & Goulian, 2010b; Skerker et al., 2008; Weigt, White, Szurmant, Hoch, & Hwa, 2009). OISCCs pairwise comparison among paralogous systems was achieved by keeping interaction interface relative positions constant during analysis (Table 3.2.1). The prototypical interaction interface composition of each system can be discerned in the parental sequence logos (Figure 3.3.2.4), enabling a detailed analysis of the specificity determining residues, their relation and interaction among them (Table 3.3.2.1). Making it clearer how OISCC composition endows proteins the capability - through molecular recognition- to discriminate between orthologue and paralogue proteins. Furthermore, most of the OISCC residue pairs form clusters of spatially neighboring residues different from each other with minimal surface overlap, conforming unique fingerprints in the interaction interfaces (Figure 3.3.2.2).

Hereinafter we review the edges and networks from the systems that were analyzed, the first network that will be described is the PhoRB system for which previous reports have already described most of its edges, afterwards we will deal with those systems which edges have not been reported before. PhoRB's edge T10-N29 is in accordance with what has been reported in (Capra et al., 2012; Podgornaia et al., 2013). Additionally, three residues were identified that have been reported to be involved in molecular recognition T4, V5, and Y8 (Capra et al., 2012), and their corresponding residue pairs I25, I25, and M27, respectively (Table 3.3.2.1). These could be involved in a hydrophobic pocket as suggested by Podgornaia *et al.* 2013 (Podgornaia et al., 2013). One last pair was also inferred A17-M27 that to the best of our knowledge has not been reported before. The ICI capability to highlight residues pairs that have been previously reported to be important for PhoRB system's molecular recognition either through structural or biochemical assays, shows the reliability of this approach and deposits confidence on those residue pairs determined for the systems that have not been thoroughly studied before.

CpxAR's OISCC is comprised of the following edges T4-F37, L7-F37, A9-D39, and A9-R40, the first two edges most probably conform a hydrophobic pocket (Table 3.3.2.1). The last two edges might be related to a size compatibility interaction, since neighboring residues on the RR might be involved in an electrostatic interaction or among spatially near charged residues from the interface, namely HK's positions 11, 12, and 13, and RR's 39 and 40 (Figure 3.3.2.4 and Table 3.3.2.1). In a published report, mutated position P109D amongst others rewired OmpR allowing it to be phosphorylated by CpxA, likewise, mutations E257A, V250T and Y254L inserted to RstB allowed it to phosphorylate CpxR to some extent (Capra et al., 2010), these correspond to relative positions 39, 9, 4, and 7, respectively (Figure 3.3.2.4 and Table 3.3.2.1). All of these positions are present in the predicted OISCC, hence to some extent this methodology can predict residue pairs that are important for molecular recognition.

NarQP has the lowest number of edges with only two L3-D36 and I7-39P (Table 3.3.2.1), both pairings are not so favorable according to Procaccini *et al.* 2011 but have average compatibility values according to Biro 2006 (Biro, 2006; Procaccini, Lunt, Szurmant, Hwa, & Weigt, 2011). There are two reasonable causes for the identification of only two pairs and that both of them appear to be unfavorable, first of all that the prototypical interaction interface does not accurately represent the NarL/FixJ subfamily and the second one that is a consequence of the first, the use of distance values between residues that come from an

inaccurate framework. Since the distance values used were computed from a PhoR-PhoB-like structure in complex. Tridimensional modelling will help to further enhance the selection of a proper prototypical interaction interface, this is currently being explored. Finally, QseCB's network has 3 edges E9-A38, V10-G29, and D14-K30, the first edge has the highest ICI value calculated for all datasets (Table 3.3.2.1). Noteworthy is that this network also has the edge with the highest distance between residues (D14-K30), the fact that a pair with such a great distance was detected, highlights the strength of the present approach (Figure 3.3.2.3). Moreover, these two residues have opposite charges which makes it highly probable that this two are nearer in the real complex during signal transduction, strengthening molecular recognition between cognate substrates (Figure 3.3.2.2 D). A shorter distance between these residues should change the overall docking upon complex formation and the distance between the histidine and aspartate that catalyze phosphorylation, though assessing this is well beyond the aim of this study.

Some of the aforementioned pairs have been identified before (Capra et al., 2010, 2012; Skerker et al., 2008), though this is the first time that residue pairs have been determined for distinct systems. Previously unreported co-evolving positions are due to us further dissecting the TCS family in orthologues and asses them as paralogues. Besides, circumscribing our method to residues which are located on a previously defined PII allowed to better describe networks unique to its systems. Furthermore, most of the specificity residues found on the RRs belong to the $\alpha 5$ - $\beta 5$ loop which has been reported to switch phosphotransfer specificity almost completely (Capra et al., 2010).

It is suggested that changes to components of these networks will substantially hinder the overall molecular recognition capability of any system, rendering the system unable to discriminate cognate from non-cognate proteins. Moreover, changes to a given network will affect directly a cognate system, but it is possible that the whole network becomes unstable and considerable cross-talk arises among paralogue systems (Capra et al., 2010, 2012; Guckes et al., 2013; Siryaporn et al., 2010b).

Even though, it would be interesting to analyze the ICI value and relations among particular amino acid pairs, their individual contribution to specificity and discrimination *in vivo* cannot be inferred by the current method. Besides, their individual contribution is highly context dependent and could be synergetic, non-additive, or epistatic. Though the defined networks could be used as an *a priori* inference to assess the *in vivo* context dependency and synergetic contribution of vertex's edges and its network. Major molecular recognition

TESIS TESIS TESIS TESIS TESIS

disruption can be achieved by substituting residue pairs with opposing chemical properties, same charge, or greater size. Hence, replacing any of these residues will lead to intersystem promiscuous paralogue interactions, such behavior has been reported for PhoR (Capra et al., 2012). On the other hand, replacing residue pairs that match another paralogue system's OISCC can result in rewiring specificity, though the exact outcome will depend on the residues being substituted due to context dependency and interactions to other components of the interaction interface (Capra et al., 2010, 2012; Podgornaia et al., 2013; Skerker et al., 2008). Withal, HK and RR interface mutants from the literature highlight the importance of co-evolving HK-RR interaction interfaces to the system-wide discrimination, avoiding detrimental cross-talk between paralogue TCSs or eliciting slight cross-talk as functional or salvaging mechanism from deleterious loss of HKs or RRs (Capra et al., 2010, 2012; Guckes et al., 2013; Huynh et al., 2015a; Noriega et al., 2010; Noriega, Schmidt, Gray, Chen, & Stewart, 2008; Siryaporn et al., 2010b).

The utmost rewiring of a system would require the cumulative effect from more interacting residues than those present in the system's OISCC that might be present in lower percentiles, though lowering the filtering threshold would increase the detection of false positives with the present methodology. Moreover, residues which are completely conserved among orthologue interfaces should not be overlooked, although its contribution cannot be directly measured by methods such as the present, *in vivo* experiments should identify its contribution both to intrasystem selectivity and intersystem insulation (Figure 3.3.2.4). Taking everything into consideration gross discrimination can be overridden and specificity significantly disrupted by substituting all component of a given OISCC network (Table 3.3.2.1).

Even though this might be a strong predictive method for OmpR/PhoB systems, accurate prediction for the NarL/FixJ subfamily might be precluded by the fact that no tridimensional interacting co-crystal structure has been solved for any TCS subfamily other than OmpR/PhoB. Thus, slight structural differences between HK and RR from this subfamily might lead to an alternative docking and a different interaction interface (Podgornaia et al., 2013). This drawback can be circumvented by modelling interacting proteins from this particular subfamily and using optimization algorithms to find approximate solutions for the interaction interface (Song et al., 2021).

In order to obtain a database greater than 125 sequences for NarQ-NarP in which random MI subsided (Dunn et al., 2008), we merged MSA of concatenated co-operonic sequence

from both NarQ-NarP and NarX-NarL systems. Since NarQ can phosphorylate both NarL and NarP (RRs), the rationale was that the interaction interfaces from both are conserved enough to allow recognition and phosphotransfer from both cognate and non-cognate HKs, NarQ-NarP and NarQ-NarL, respectively (Casino et al., 2009; Noriega et al., 2010, 2008; Podgornaia et al., 2013). Hence, this interaction interface network can be used as an approximate solution to residues that confer intrasystem specificity and intersystem insulation.

4.4 TCSs' discrimination in *A. pleuropneumoniae*

Two-component systems in *A. pleuropneumoniae* (*App*) have not been thoroughly studied, though they show interesting features. Aside QseC-QseB and NarQ-NarP systems being reported in response to catecholamines (L. Li et al., 2012; Liu et al., 2015) and CpxA-CpxR contributing to biofilm formation and virulence (H. Li et al., 2018; Yan et al., 2020) much is still unknown of the *App*'s TCS network. Moreover, *App* is a pathogenic bacterium with high prevalence that forms monospecies and multispecies biofilms (Hathroubi et al., 2015; Labrie et al., 2010; A. Loera-Muro et al., 2013, 2021; V. M. Loera-Muro et al., 2013) and achieves host colonization, establishment of infection, essential nutrients acquisition, metabolic adaptation, and persistence within the host, with just a limited set of environment sensory systems (Bossé et al., 2002; Chiers et al., 2010; Deslandes et al., 2010; Deslandes, Nash, Harel, Coulton, & Jacques, 2007; Jacques, 2004; Liu et al., 2015).

The OISCCs discussed in the previous section can be used to infer the species-specific interface discrimination of *App* (Figure 3.3.2.4 and Table 3.3.2.1), sequences from all sequenced *App* strains were used to build MSAs and retrieve the species specific PII. Once the species-specific PIIs are compared to the parental data, species-specific OISCCs resemble parental residue composition (Figures 3.3.2.3 and 3.3.2.4), albeit particular. Therefore, the interaction interface discriminatory capability is system dependent and species-specific, all interfaces are going to be differentially insulated, however they are still capable of system wide insulation of non-cognate substrates. Furthermore, the cognate substrate's interface is going to be specifically recognized over other interfaces, primarily due to the OISCC composition, though not solely, since additional associations might also fine tune the interface's discriminatory capability. Interface composition is the source of TCS's specificity, discriminating cognate from non-cognate interfaces through selectively recognizing the former and non-selective insulation of the latter.

TESIS TESIS TESIS TESIS TESIS

In most *App* strains' response regulator protein PhoB lacks critical residues from the acidic pocket that coordinates Mg²⁺ or other divalent ions that catalyze the phosphorylation of the conserved aspartate from the phosphorylation site (Lukat et al., 1990; Needham et al., 1993). This should disrupt PhoB's ability to phosphorylate, hindering the downstream signaling cascade of phosphate starvation response mediated by the PhoR-PhoB system with inorganic phosphate (Pi) as its specific stimulus (Crépin et al., 2011; Lamarche et al., 2008; Santos-Beneit, 2015). Likewise, the interaction interface lacks a residue which confers specificity, relative position I25 that corresponds to RR468's L14 from *T. maritima* (Figures 3.3.2.3C and 3.3.2.4C and Tables 3.2.1 and 3.3.2.1) (Podgornaia et al., 2013). In most *App* strains, the PhoR-PhoB OISCC is missing residue 25 that pairs with residues 4 and 5, and two other strains are also missing residue 27 that pairs with residue 8, both occurrences compromise the OISCC integrity, though is not clear to what extent. Only strains 1 4074, 18-1342, G1-9626, S8, and S4074 have fully functional PhoB proteins according to their coding sequences, thus a completely functional PhoR-PhoB signaling pathway and its OISCC. No compensatory changes were detected in the OISCC from strains that have truncated versions of PhoB, the composition is similar to that of the parental interaction interface and is the same for all observed strains (Figures 3.3.2.3C and 3.3.2.4C). Therefore, intersystem discrimination of non-cognate substrates is maintained over salvaging cross-talk (Capra et al., 2012; Rowland & Deeds, 2014), and for an event to occur that compensates for a loss in an already stable network might take a considerable amount of time. PhoR-PhoB is extremely important to bacteria, thus *App* must have evolved a molecular strategy to withstand for the apparent loss of this critical function. Hence, it is suggested that the PhoR-PhoB salvaging strategy could be in an upper or a lower signaling pathway hierarchy, further biological research will be required to clarify this.

The PhoB-PhoR's OISCC in *App* is similar to the parental one with some remarkable features, the lack of residue 25 and that some residue pairs' chemical identities changed (Table 3.3.2.1). Both of these pairs have a good compatibility in terms of size, charge, and hydropathicity (Biro, 2006). Implications of missing OISCC's residue 25 have been discussed in the previous paragraph. PhoR-PhoB is the most compromised of all observed OISCC from *App*.

CpxR protein from *App*'s strains 2 4226, 2 S1536, 3 JL03, and NCTC10976 have an 18 amino acids gap (Figure 3.1.9). None of the missing residues belong to the interaction interface from CpxA-CpxR, yet the gap starts after CpxR P33, this could alter the overall

tridimensional structure of the protein and modify the docking properties and relative position of subsequent residues from the PII (Figure 3.2.1C and Table 3.2.1). The discrimination capability of this system might be hindered to some extent, though not thoroughly compromised. The only change present in CpxA CpxR's OISCC is the change of residue 4 in the pair A4-F37 (Table 3.2.1), this substitution improves hydropathicity and is little bit better in terms of size and charge (Biro, 2006).

The CpxA-CpxR system has been reported to be involved in the sensing of diverse stimuli, such as cell envelope integrity, adhesion, central metabolism, and misfolded proteins either from the inner membrane or the periplasmic space in Gram negative bacteria (Hunke et al., 2012; Wolfe, 2010; Wolfe et al., 2008), in *App* has been reported to be involved in biofilm formation, stress resistance, growth, and virulence (H. Li et al., 2018; Yan et al., 2020). Although the discrimination capability of the CpxA-CpxR might be hampered, strains 2 S1536 and 3 JL03 are prevalent swine pathogens in Europe and China, respectively (Jacques, 2004; Xu et al., 2010, 2008).

NCTC11383 is the only strain that presents a fragmentary NarQ sensor protein and it is missing the conserved histidine, leading to a non-functional HK and a hampered upstream signaling pathway. NarQP's is the smallest of all observed OISCCs with only two pairs and just one residue substitution P39T (Figure 3.3.2.4B and Table 3.2.1), this pair (I10-T39) is about the same in terms of residue compatibility values (Biro, 2006).

QseC and QseB proteins in *App* do not have any noteworthy features, a fully functional QseCB two-component system. This system's OISCC presents a single substitution V10I on the I10-G29 pair (Figure 3.3.2.4D and Table 3.2.1), this substitution is less favorable than the parental one, though it still has fairly good size and hydropathicity compatibility values (Biro, 2006; Procaccini et al., 2011).

Four clusters of residue pairs (OISCC) were determined for each of the assessed systems, as well as its exact chemical composition and its putative spatial location in the protein complex (Table 3.2.1). Each cluster has a unique shape constituted by spatial neighboring residue pairs (Figure 4), these shapes are formed in the interaction interface and have minimal overlap among them. Effectively allowing us to dissect those residue pairs that confer each of the systems with its capability to discriminate cognate of non-cognate substrates and which is species specific and quasi system wide.

CONCLUSIONS

This study uses a computational biology approach that relied on the Interface coupling index to assess *App*'s two-components discrimination through molecular recognition. Molecular recognition has been recognized as the primary mechanism that confers specificity to two component systems signal transduction networks (Capra et al., 2010, 2012; Dago et al., 2012; Podgornaia et al., 2013; Procaccini et al., 2011; Skerker et al., 2008). Hence, molecular recognition depends on the formation of an interface between interacting proteins in a transient complex. Two-component systems' specificity, selectively recognizes a cognate and effectively insulates non-cognate components. In this work it is suggested that specificity is achieved by a network of residue pairs named Orthologue interface specificity core cluster (OISCC), that is the primary component of specificity and is constituted by residue pairs with varying degree of ICI values. Furthermore, we were able to describe previously unreported residue position that confer specificity among paralogous systems. Most of the specificity conferring residues in the RRs are located on the $\alpha 5$ - $\beta 5$ loop and together with its paired residues on the HK's $\alpha 1$ and $\alpha 2$ form clusters of spatially neighboring residues in the interface.

These residues form individual networks that grant paralogous systems the capability of specifically recognizing cognate over non-cognate proteins. With these we can identify in more detail the specificity endowing networks among paralogues, and how changes to one or more of its components might switch its capability to discriminate components. These changes can either be of size, charge, or hydrophathy (Biro, 2006), though the exact resulting changes to specificity can only be assessed biologically.

Amino acid pairs from specificity interaction interface components were calculated from datasets built from approximately 2500 genomes available at the time. The datasets were built only of orthologue sequences, thus describing the whole orthologue's family specificity. Therefore, this systematic procedure serves to infer species-specific system wide TCSs' specificity in bacteria. The interface specificity from the components of the network is context dependent and species-specific, and will depend on the total number, type, architecture, and phylogeny of the systems coded. We suggest that the whole interaction interface is being used in what could resemble a spatial permutational process, this could explain how bacteria such as *Myxococcus xanthus* and *Geobacter sulfurreducens* achieve system wide discrimination with a number of TCSs around the hundreds. The interaction interface works as a tridimensional puzzle that relies on interface complementarity and

physicochemical properties to achieve specific recognition of cognate and insulation of non-cognate components.

Though, the relevance of the interaction interface specificity components is at the time merely theoretical, it can be used as an *a priori* approach to infer species-specific system-wide prototypical TCSs in vivo specificity. Further contributing to the current knowledge about molecular recognition among TCSs' proteins, and its implication in cognate specific recognition and non-cognate discrimination among orthologue and paralogue families.

Even though *A. pleuropneumoniae* has a small number of two-component system proteins ten in all sequenced genomes, which account for all five signaling pathways of which three are cognate and two orphans. It is a pathogenic bacterium that has a great prevalence both in its host and in the environment, with its ability to form monospecies and multispecies biofilms in order to survive. In addition, it has different degrees of virulence, causing from severe acute pleuropneumonia but also has been found as a chronic disease. All of these being achieved with these few environmental signal transduction systems.

Particularly in *A. pleuropneumoniae* TCSs specificity is maintained over salvaging cross-talk. It is thought that once a system has been effectively insulated, the loss of function of one of the components does not lead right away to a compensatory mechanism that rewires specificity. Therefore, TCSs' interaction interface plasticity in *App* can be discarded and the compensation for loss of function might come from components outside the studied network. Therefore, the stability of the whole network is greater than the function of one of the components, i.e., PhoR-PhoB and CpxA-CpxR which have putative critical functions on environmental queues.

Bacterial two-component system networks are comprised of phosphotransfer and phosphorelay systems, our current methodology is unable to describe hybrid or unorthodox interaction interfaces. This arises as a liability to describe system-wide signal transduction networks and highlights the lack of work done over hybrid, unorthodox, or components from seldom found families. Solved crystals or accurate models of interacting proteins from these systems will clarify the interaction interface from different TCSs architectures and families, solving docking properties from HK-RR heterodimer formation and allowing the assessment of differences and similarities among interaction interfaces. AlphaFold (Jumper et al., 2021; Senior et al., 2020; J. Yang et al., 2020), a recently published algorithm that predicts the tridimensional structure of protein based solely in its amino acid sequence will allow to infer

the interaction interface of any protein complex and enhance the performance of algorithms that use structural data to determine specificity components, such as the one here described.

Despite some drawbacks, there are also strengths to this methodology. The exponential growth in the total number of available genomes from organisms completely or partially sequenced and the availability of tools that predict protein structure, will allow the creation of datasets big enough to even describe or model putative interacting interfaces from TCSs that are rarely distributed, such as hybrid HKs, HPt, CheA, LytTR, PleD, PrrA or YesN, just to mention a few. The forgoing coupled to an algorithm that takes in to account variables such as: size volume, charge, and hydrophaticity compatibility and tridimensional docking complementarity; will provide more reliable results in the assessment of interaction interface specificity components. Enhancing our current understanding of the capability of the systems to discriminate among cognate and non-cognate components, allowing to model even more complex TCS system wide networks from prokaryote or archaeal organisms. Moreover, this methodology is not circumscribed to two-component systems' protein interfaces but it can also be used to analyze specificity interface components from other transient interacting protein, whether they are homomers or heteromers. It could be of use in the study of signal transduction network discrimination among components and the rewiring of individual ones. Biotechnology would also benefit from this kind information, by changing environmental signals from metabolites of interest to more amenable production conditions. Systems biology can also benefit from this approach to engineer highly specific or promiscuous synthetic signal transduction networks. Using specificity residue pairs to rationally wire signals to discrete outputs, designing novel phosphotransfer targets, or phosphorelay circuitry, thus, allowing synthetic one-to-one, one-to-many, many-to-one or many-to-many stimuli response hubs design.

BIBLIOGRAPHY

Alvarez, A. F., Rodriguez, C., & Georgellis, D. (2013). Ubiquinone and menaquinone electron carriers represent the yin and yang in the redox regulation of the ArcB sensor kinase. *Journal of Bacteriology*, 195(13), 3054–3061. doi: 10.1128/jb.00406-13

Auger, E., Deslandes, V., Ramjeet, M., Contreras, I., Nash, J. H. E., Harel, J., ... Jacques, M. (2009). Host-Pathogen Interactions of *Actinobacillus pleuropneumoniae* with Porcine Lung and Tracheal Epithelial Cells. *Infection and Immunity*, 77(4), 1426–1441. doi: 10.1128/iai.00297-08

- Baltes, N., N'diaye, M., Jacobsen, I. D., Maas, A., Buettner, F. F. R., & Gerlach, G.-F. (2005). Deletion of the Anaerobic Regulator HlyX Causes Reduced Colonization and Persistence of *Actinobacillus pleuropneumoniae* in the Porcine Respiratory Tract. *Infection and Immunity*, 73(8), 4614–4619. doi: 10.1128/iai.73.8.4614-4619.2005
- Batchelor, E., & Goulian, M. (2006). Imaging OmpR localization in *Escherichia coli*. *Molecular Microbiology*, 59(6), 1767–1778. doi: 10.1111/j.1365-2958.2006.05048.x
- Biro, J. (2006). Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. *Theoretical Biology & Medical Modelling*, 3(1), 15–15. doi: 10.1186/1742-4682-3-15
- Bossé, J. T., Chaudhuri, R. R., Li, Y., Leanse, L. G., Crespo, R. F., Coupland, P., ... Langford, P. R. (2016). Complete Genome Sequence of MIDG2331, a Genetically Tractable Serovar 8 Clinical Isolate of *Actinobacillus pleuropneumoniae*. *Genome Announcements*, 4(1), e01667-15. doi: 10.1128/genomea.01667-15
- Bossé, J. T., Janson, H., Sheehan, B. J., Beddek, A. J., Rycroft, A. N., Kroll, J. S., & Langford, P. R. (2002). *Actinobacillus pleuropneumoniae*: pathobiology and pathogenesis of infection. *Microbes and Infection*, 4(2), 225–235. doi: 10.1016/s1286-4579(01)01534-9
- Bourret, R. B. (2010). Receiver domain structure and function in response regulator proteins. *Current Opinion in Microbiology*, 13(2), 142–149. doi: 10.1016/j.mib.2010.01.015
- Bourret, R. B., & Silversmith, R. E. (2010). Two-component signal transduction. *Current Opinion in Microbiology*, 13(2), 113–115. doi: 10.1016/j.mib.2010.02.003
- Buettner, F. F. R., Bendalla, I. M., Bossé, J. T., Meens, J., Nash, J. H. E., Härtig, E., ... Gerlach, G. (2009). Analysis of the *Actinobacillus pleuropneumoniae* HlyX (FNR) regulon and identification of iron-regulated protein B as an essential virulence factor. *PROTEOMICS*, 9(9), 2383–2398. doi: 10.1002/pmic.200800439
- Buettner, F. F. R., Bendallah, I. M., Bosse, J. T., Dreckmann, K., Nash, J. H. E., Langford, P. R., & Gerlach, G.-F. (2008). Analysis of the *Actinobacillus pleuropneumoniae* ArcA Regulon Identifies Fumarate Reductase as a Determinant of Virulence. *Infection and Immunity*, 76(6), 2284–2295. doi: 10.1128/iai.01540-07
- Buettner, F. F. R., Maas, A., & Gerlach, G.-F. (2008). An *Actinobacillus pleuropneumoniae* arcA deletion mutant is attenuated and deficient in biofilm formation. *Veterinary Microbiology*, 127(1–2), 106–115. doi: 10.1016/j.vetmic.2007.08.005
- Cai, S. J., & Inouye, M. (2002). EnvZ-OmpR Interaction and Osmoregulation in *Escherichia coli* *. *Journal of Biological Chemistry*, 277(27), 24155–24161. doi: 10.1074/jbc.m110715200
- Capra, E. J., & Laub, M. T. (2012). Evolution of Two-Component Signal Transduction Systems. *Annual Review of Microbiology*, 66(1), 325–347. doi: 10.1146/annurev-micro-092611-150039
- Capra, E. J., Perchuk, B. S., Lubin, E. A., Ashenberg, O., Skerker, J. M., & Laub, M. T. (2010). Systematic Dissection and Trajectory-Scanning Mutagenesis of the Molecular Interface That

- Ensures Specificity of Two-Component Signaling Pathways. *PLoS Genetics*, 6(11), e1001220. doi: 10.1371/journal.pgen.1001220
- Capra, E. J., Perchuk, B. S., Skerker, J. M., & Laub, M. T. (2012). Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. *Cell*, 150(1), 222–232. doi: 10.1016/j.cell.2012.05.033
- Cases, I., & Lorenzo, V. de. (2005). Promoters in the environment: transcriptional regulation in its natural context. *Nature Reviews Microbiology*, 3(2), 105–118. doi: 10.1038/nrmicro1084
- Casino, P., Rubio, V., & Marina, A. (2009). Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell*, 139(2), 325–336. doi: 10.1016/j.cell.2009.08.032
- Cheung, J., & Hendrickson, W. A. (2010). Sensor domains of two-component regulatory systems. *Current Opinion in Microbiology*, 13(2), 116–123. doi: 10.1016/j.mib.2010.01.016
- Chiers, K., Waele, T. D., Pasmans, F., Ducatelle, R., & Haesebrouck, F. (2010). Virulence factors of *Actinobacillus pleuropneumoniae* involved in colonization, persistence and induction of lesions in its porcine host. *Veterinary Research*, 41(5), 65. doi: 10.1051/vetres/2010037
- Clarke, M. B., Hughes, D. T., Zhu, C., Boedeker, E. C., & Sperandio, V. (2006). The QseC sensor kinase: A bacterial adrenergic receptor. *Proceedings of the National Academy of Sciences*, 103(27), 10420–10425. doi: 10.1073/pnas.0604343103
- Crépin, S., Chekabab, S.-M., Bihan, G. L., Bertrand, N., Dozois, C. M., & Harel, J. (2011). The Pho regulon and the pathogenesis of *Escherichia coli*. *Veterinary Microbiology*, 153(1–2), 82–88. doi: 10.1016/j.vetmic.2011.05.043
- Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, 14(6), 1188–1190. doi: 10.1101/gr.849004
- Dago, A. E., Schug, A., Procaccini, A., Hoch, J. A., Weigt, M., & Szurmant, H. (2012). Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *Proceedings of the National Academy of Sciences*, 109(26), E1733–E1742. doi: 10.1073/pnas.1201301109
- Deslandes, V., Denicourt, M., Girard, C., Harel, J., Nash, J. H., & Jacques, M. (2010). Transcriptional profiling of *Actinobacillus pleuropneumoniae* during the acute phase of a natural infection in pigs. *BMC Genomics*, 11(1), 98–98. doi: 10.1186/1471-2164-11-98
- Deslandes, V., Nash, J. H., Harel, J., Coulton, J. W., & Jacques, M. (2007). Transcriptional profiling of *Actinobacillus pleuropneumoniae* under iron-restricted conditions. *BMC Genomics*, 8(1), 72. doi: 10.1186/1471-2164-8-72
- Dubreuil, J. D., Jacques, M., Mittal, K. R., & Gottschalk, M. (2000). *Actinobacillus pleuropneumoniae* surface polysaccharides: their role in diagnosis and immunogenicity. *Animal Health Research Reviews*, 1(2), 73–93. doi: 10.1017/s1466252300000074

- Dunn, S. D., Wahl, L. M., & Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, 24(3), 333–340. doi: 10.1093/bioinformatics/btm604
- Fisher, S. L., Jiang, W., Wanner, B. L., & Walsh, C. T. (1995). Cross-talk between the Histidine Protein Kinase VanS and the Response Regulator PhoB. *Journal of Biological Chemistry*, 270(39), 23143–23149. doi: 10.1074/jbc.270.39.23143
- Fodor, A. A., & Aldrich, R. W. (2004). On Evolutionary Conservation of Thermodynamic Coupling in Proteins*. *Journal of Biological Chemistry*, 279(18), 19046–19050. doi: 10.1074/jbc.m402560200
- Foote, S. J., Bossé, J. T., Bouevitch, A. B., Langford, P. R., Young, N. M., & Nash, J. H. E. (2007). The complete genome sequence of *Actinobacillus pleuropneumoniae* L20 (serotype 5b). *Journal of Bacteriology*, 190(4), 1495–1496. doi: 10.1128/jb.01845-07
- Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. *Current Opinion in Microbiology*, 13(2), 150–159. doi: 10.1016/j.mib.2010.01.005
- Gao, R., & Stock, A. M. (2010a). Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Current Opinion in Microbiology*, 13(2), 160–167. doi: 10.1016/j.mib.2009.12.009
- Gao, R., & Stock, A. M. (2010b). Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Current Opinion in Microbiology*, 13(2), 160–167. doi: 10.1016/j.mib.2009.12.009
- Georgellis, D., Kwon, O., Lin, E. C., Wong, S. M., & Akerley, B. J. (2001). Redox signal transduction by the ArcB sensor kinase of *Haemophilus influenzae* lacking the PAS domain. *Journal of Bacteriology*, 183(24), 7206–7212. doi: 10.1128/jb.183.24.7206-7212.2001
- Goulian, M. (2009). Two-component signaling circuit structure and properties. *Current Opinion in Microbiology*, 13(2), 184–189. doi: 10.1016/j.mib.2010.01.009
- Grebe, T. W., & Stock, J. B. (1999). The Histidine Protein Kinase Superfamily. *Advances in Microbial Physiology*, 41, 139–227. doi: 10.1016/s0065-2911(08)60167-8
- Green, J., & Baldwin, M. L. (1997). HlyX, the FNR homologue of *Actinobacillus pleuropneumoniae*, is a [4Fe–4S]-containing oxygen-responsive transcription regulator that anaerobically activates FNR-dependent Class I promoters via an enhanced AR1 contact. *Molecular Microbiology*, 24(3), 593–605. doi: 10.1046/j.1365-2958.1997.3801737.x
- Guckes, K. R., Kostakioti, M., Breland, E. J., Gu, A. P., Shaffer, C. L., Martinez, C. R., ... Hadjifrangiskou, M. (2013). Strong cross-system interactions drive the activation of the QseB response regulator in the absence of its cognate sensor. *Proceedings of the National Academy of Sciences*, 110(41), 16592–16597. doi: 10.1073/pnas.1315320110

- Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15), 2714–2723. doi: 10.1002/elps.1150181505
- Hathroubi, S., Fontaine-Gosselin, S.-È., Tremblay, Y. D. N., Labrie, J., & Jacques, M. (2015). Sub-inhibitory concentrations of penicillin G induce biofilm formation by field isolates of *Actinobacillus pleuropneumoniae*. *Veterinary Microbiology*, 179(3–4), 277–286. doi: 10.1016/j.vetmic.2015.06.011
- Hughes, D. T., Clarke, M. B., Yamamoto, K., Rasko, D. A., & Sperandio, V. (2009). The QseC Adrenergic Signaling Cascade in Enterohemorrhagic *E. coli* (EHEC). *PLoS Pathogens*, 5(8), e1000553. doi: 10.1371/journal.ppat.1000553
- Hunke, S., Keller, R., & Müller, V. S. (2012). Signal integration by the Cpx-envelope stress system. *FEMS Microbiology Letters*, 326(1), 12–22. doi: 10.1111/j.1574-6968.2011.02436.x
- Huynh, T. N., Chen, L.-L., & Stewart, V. (2015a). Sensor–response regulator interactions in a cross-regulated signal transduction network. *Microbiology*, 161(7), 1504–1515. doi: 10.1099/mic.0.000092
- Huynh, T. N., Chen, L.-L., & Stewart, V. (2015b). Sensor–response regulator interactions in a cross-regulated signal transduction network. *Microbiology*, 161(7), 1504–1515. doi: 10.1099/mic.0.000092
- Huynh, T. N., Noriega, C. E., & Stewart, V. (2010). Conserved mechanism for sensor phosphatase control of two-component signaling revealed in the nitrate sensor NarX. *Proceedings of the National Academy of Sciences of the United States of America*, 107(49), 21140–21145. doi: 10.1073/pnas.1013081107
- Huynh, T. N., & Stewart, V. (2011a). Negative control in two-component signal transduction by transmitter phosphatase activity. *Molecular Microbiology*, 82(2), 275–286. doi: 10.1111/j.1365-2958.2011.07829.x
- Huynh, T. N., & Stewart, V. (2011b). Negative control in two-component signal transduction by transmitter phosphatase activity: Transmitter phosphatase in two-component signalling. *Molecular Microbiology*, 82(2), 275–286. doi: 10.1111/j.1365-2958.2011.07829.x
- Jacques, M. (2004). Surface polysaccharides and iron-uptake systems of *Actinobacillus pleuropneumoniae*. *Canadian Journal of Veterinary Research = Revue Canadienne de Recherche Veterinaire*, 68(2), 81–85.
- Jensen, R. B., Wang, S. C., & Shapiro, L. (2002a). Dynamic localization of proteins and DNA during a bacterial cell cycle. *Nature Reviews Molecular Cell Biology*, 3(3), 167–176. doi: 10.1038/nrm758
- Jensen, R. B., Wang, S. C., & Shapiro, L. (2002b). Dynamic localization of proteins and DNA during a bacterial cell cycle. *Nature Reviews Molecular Cell Biology*, 3(3), 167–176. doi: 10.1038/nrm758

- Johansson, M. U., Zoete, V., Michielin, O., & Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics*, 13(1), 173–173. doi: 10.1186/1471-2105-13-173
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. doi: 10.1038/s41586-021-03819-2
- Jung, K., Fried, L., Behr, S., & Heermann, R. (2012). Histidine kinases and response regulators in networks. *Current Opinion in Microbiology*, 15(2), 118–124. doi: 10.1016/j.mib.2011.11.009
- Labrie, J., Pelletier-Jacques, G., Deslandes, V., Ramjeet, M., Auger, E., Nash, J. H. E., & Jacques, M. (2010). Effects of growth conditions on biofilm formation by *Actinobacillus pleuropneumoniae*. *Veterinary Research*, 41(1), 1–17. doi: 10.1051/vetres/2009051
- Lamarche, M. G., Wanner, B. L., Crépin, S., & Harel, J. (2008). The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *FEMS Microbiology Reviews*, 32(3), 461–473. doi: 10.1111/j.1574-6976.2008.00101.x
- Laub, M. T., Biondi, E. G., & Skerker, J. M. (2007). Phosphotransfer profiling: systematic mapping of two-component signal transduction pathways and phosphorelays. *Methods in Enzymology*, 423, 531–548. doi: 10.1016/s0076-6879(07)23026-5
- Laub, M. T., & Goulian, M. (2007). Specificity in Two-Component Signal Transduction Pathways. *Genetics*, 41(1), 121–145. doi: 10.1146/annurev.genet.41.042007.170548
- Li, G., Xie, F., Zhang, Y., & Wang, C. (2012). Draft genome sequence of *Actinobacillus pleuropneumoniae* serotype 7 strain S-8. *Journal of Bacteriology*, 194(23), 6606–6607. doi: 10.1128/jb.01650-12
- Li, H., Liu, F., Peng, W., Yan, K., Zhao, H., Liu, T., ... Bei, W. (2018). The CpxA/CpxR Two-Component System Affects Biofilm Formation and Virulence in *Actinobacillus pleuropneumoniae*. *Frontiers in Cellular and Infection Microbiology*, 8, 72. doi: 10.3389/fcimb.2018.00072
- Li, L., Xu, Z., Zhou, Y., Li, T., Sun, L., Chen, H., & Zhou, R. (2011). Analysis on *Actinobacillus pleuropneumoniae* LuxS regulated genes reveals pleiotropic roles of LuxS/AI-2 on biofilm formation, adhesion ability and iron metabolism. *Microbial Pathogenesis*, 50(6), 293–302. doi: 10.1016/j.micpath.2011.02.002
- Li, L., Xu, Z., Zhou, Y., Sun, L., Liu, Z., Chen, H., & Zhou, R. (2012). Global Effects of Catecholamines on *Actinobacillus pleuropneumoniae* Gene Expression. *PLoS ONE*, 7(2), e31121. doi: 10.1371/journal.pone.0031121
- Li, L., Zhu, J., Yang, K., Xu, Z., Liu, Z., & Zhou, R. (2014). Changes in gene expression of *Actinobacillus pleuropneumoniae* in response to anaerobic stress reveal induction of central metabolism and biofilm formation. *Journal of Microbiology*, 52(6), 473–481. doi: 10.1007/s12275-014-3456-y

- Liu, J., Hu, L., Xu, Z., Tan, C., Yuan, F., Fu, S., ... Bei, W. (2015). Actinobacillus pleuropneumoniae two-component system QseB/QseC regulates the transcription of PilM, an important determinant of bacterial adherence and virulence. *Veterinary Microbiology*, 177(1–2), 184–192. doi: 10.1016/j.vetmic.2015.02.033
- Loera-Muro, A., Avelar-González, F. J., Loera-Muro, V. M., Jacques, M., & Guerrero-Barrera, A. L. (2013). Presence of <i>Actinobacillus pleuropneumoniae</i>, <i>Streptococcus suis</i>, <i>Pasteurella multocida</i>, <i>Bordetella bronchiseptica</i>, <i>Haemophilus parasuis</i> and <i>Mycoplasma hyopneumoniae</i> in upper respiratory tract of swine in farms from Aguascalientes, Mexico. *Open Journal of Animal Sciences*, 03(02), 132–137. doi: 10.4236/ojas.2013.32020
- Loera-Muro, A., Ramírez-Castillo, F. Y., Moreno-Flores, A. C., Martin, E. M., Avelar-González, F. J., & Guerrero-Barrera, A. L. (2021). Actinobacillus pleuropneumoniae Surviving on Environmental Multi-Species Biofilms in Swine Farms. *Frontiers in Veterinary Science*, 8, 722683. doi: 10.3389/fvets.2021.722683
- Loera-Muro, V. M., Jacques, M., Tremblay, Y. D. N., Avelar-González, F. J., Muro, A. L., Ramírez-López, E. M., ... Guerrero-Barrera, A. L. (2013). Detection of Actinobacillus pleuropneumoniae in drinking water from pig farms. *Microbiology*, 159(Pt_3), 536–544. doi: 10.1099/mic.0.057992-0
- Lukat, G. S., Stock, A. M., & Stock, J. B. (1990). Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. *Biochemistry*, 29(23), 5436–5442. doi: 10.1021/bi00475a004
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., ... Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, 39(Database issue), D225–D229. doi: 10.1093/nar/gkq1189
- Martinez-Guerrero, C. E., Ciria, R., Abreu-Goodger, C., Moreno-Hagelsieb, G., & Merino, E. (2008). GeConT 2: gene context analysis for orthologous proteins, conserved domains and metabolic pathways. *Nucleic Acids Research*, 36(suppl_2), W176–W180. doi: 10.1093/nar/gkn330
- Miyashiro, T., & Goulian, M. (2007). Two-Component Signaling Systems, Part B. *Methods in Enzymology*, 423, 458–475. doi: 10.1016/s0076-6879(07)23022-8
- Needham, J. V., Chen, T. Y., & Falke, J. J. (1993). Novel ion specificity of a carboxylate cluster Mg(II) binding site: strong charge selectivity and weak size selectivity. *Biochemistry*, 32(13), 3363–3367. doi: 10.1021/bi00064a020
- Ng, W.-L., & Bassler, B. L. (2009). Bacterial quorum-sensing network architectures. *Annual Review of Genetics*, 43(1), 197–222. doi: 10.1146/annurev-genet-102108-134304
- Noriega, C. E., Lin, H., Chen, L., Williams, S. B., & Stewart, V. (2010). Asymmetric cross-regulation between the nitrate-responsive NarX–NarL and NarQ–NarP two-component regulatory systems from Escherichia coli K-12. *Molecular Microbiology*, 75(2), 394–412. doi: 10.1111/j.1365-2958.2009.06987.x

- Noriega, C. E., Schmidt, R., Gray, M. J., Chen, L.-L., & Stewart, V. (2008). Autophosphorylation and Dephosphorylation by Soluble Forms of the Nitrate-Responsive Sensors NarX and NarQ from *Escherichia coli* K-12. *Journal of Bacteriology*, 190(11), 3869–3876. doi: 10.1128/jb.00092-08
- Peña-Sandoval, G. R., & Georgellis, D. (2010). The ArcB Sensor Kinase of *Escherichia coli* Autophosphorylates by an Intramolecular Reaction. *Journal of Bacteriology*, 192(6), 1735–1739. doi: 10.1128/jb.01401-09
- Pereira, M. F., Rossi, C. C., Carvalho, F. M. de, Almeida, L. G. P. de, Souza, R. C., Vasconcelos, A. T. R. de, & Bazzolli, D. M. S. (2015). Draft Genome Sequences of Six *Actinobacillus pleuropneumoniae* Serotype 8 Brazilian Clinical Isolates: Insight into New Applications. *Genome Announcements*, 3(2), e01585-14. doi: 10.1128/genomea.01585-14
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., ... Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70–82. doi: 10.1002/pro.3943
- Podgornaia, A. I., Casino, P., Marina, A., & Laub, M. T. (2013). Structural Basis of a Rationally Rewired Protein-Protein Interface Critical to Bacterial Signaling. *Structure*, 21(9), 1636–1647. doi: 10.1016/j.str.2013.07.005
- Podgornaia, A. I., & Laub, M. T. (2013). Determinants of specificity in two-component signal transduction. *Current Opinion in Microbiology*, 16(2), 156–162. doi: 10.1016/j.mib.2013.01.004
- Procaccini, A., Lunt, B., Szurmant, H., Hwa, T., & Weigt, M. (2011). Dissecting the Specificity of Protein-Protein Interaction in Bacterial Two-Component Signaling: Orphans and Crosstalks. *PLoS ONE*, 6(5), e19729. doi: 10.1371/journal.pone.0019729
- Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., ... Schneider, V. A. (2021). Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV). *Genome Research*, 31(1), 159–169. doi: 10.1101/gr.266932.120
- Rowland, M. A., & Deeds, E. J. (2014). Crosstalk and the evolution of specificity in two-component signaling. *Proceedings of the National Academy of Sciences*, 111(15), 5550–5555. doi: 10.1073/pnas.1317178111
- Ryan, K. R., & Shapiro, L. (2003). TEMPORAL AND SPATIAL REGULATION IN PROKARYOTIC CELL CYCLE PROGRESSION AND DEVELOPMENT. *Biochemistry*, 72(1), 367–394. doi: 10.1146/annurev.biochem.72.121801.161824
- Salazar, M. E., & Laub, M. T. (2015). Temporal and evolutionary dynamics of two-component signaling pathways. *Current Opinion in Microbiology*, 24, 7–14. doi: 10.1016/j.mib.2014.12.003
- Santos-Beneit, F. (2015). The Pho regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology*, 6, 402. doi: 10.3389/fmicb.2015.00402

- Sárközi, R., Makrai, L., & Fodor, L. (2015). Identification of a proposed new serovar of *Actinobacillus Pleuropneumoniae*: Serovar 16. *Acta Veterinaria Hungarica*, 63(4), 444–450. doi: 10.1556/004.2015.041
- Sawitzki, G. (2009). *Computational Statistics An Introduction to R*. 233–235. doi: 10.1201/b17179-7
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706–710. doi: 10.1038/s41586-019-1923-7
- Seshasayee, A. S., Bertone, P., Fraser, G. M., & Luscombe, N. M. (2006). Transcriptional regulatory networks in bacteria: from input signals to output responses. *Current Opinion in Microbiology*, 9(5), 511–519. doi: 10.1016/j.mib.2006.08.007
- Sievers, F., & Higgins, D. G. (2020). The Clustal Omega Multiple Alignment Package. *Methods in Molecular Biology (Clifton, N.J.)*, 2231, 3–16. doi: 10.1007/978-1-0716-1036-7_1
- Siryaporn, A., & Goulian, M. (2008). Cross-talk suppression between the CpxA–CpxR and EnvZ–OmpR two-component systems in *E. coli*. *Molecular Microbiology*, 70(2), 494–506. doi: 10.1111/j.1365-2958.2008.06426.x
- Siryaporn, A., Perchuk, B. S., Laub, M. T., & Goulian, M. (2010a). Evolving a robust signal transduction pathway from weak cross-talk. *Molecular Systems Biology*, 6(1), 452–452. doi: 10.1038/msb.2010.105
- Siryaporn, A., Perchuk, B. S., Laub, M. T., & Goulian, M. (2010b). Evolving a robust signal transduction pathway from weak cross-talk. *Molecular Systems Biology*, 6(1), 452–452. doi: 10.1038/msb.2010.105
- Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., & Laub, M. T. (2008). Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell*, 133(6), 1043–1054. doi: 10.1016/j.cell.2008.04.040
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., & Laub, M. T. (2005). Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. *PLoS Biology*, 3(10), e334. doi: 10.1371/journal.pbio.0030334
- Sommer, E., Koler, M., Frank, V., Sourjik, V., & Vaknin, A. (2013). The Sensory Histidine Kinases TorS and EvgS Tend to Form Clusters in *Escherichia coli* Cells. *PLoS ONE*, 8(10), e77708. doi: 10.1371/journal.pone.0077708
- Song, B., Luo, X., Luo, X., Liu, Y., Niu, Z., & Zeng, X. (2021). Learning spatial structures of proteins improves protein-protein interaction prediction. *Briefings in Bioinformatics*, 23(2). doi: 10.1093/bib/bbab558
- Sourjik, V., & Armitage, J. P. (2010). Spatial organization in bacterial chemotaxis. *The EMBO Journal*, 29(16), 2724–2733. doi: 10.1038/emboj.2010.178

- Sperandio, V., Torres, A. G., Jarvis, B., Nataro, J. P., & Kaper, J. B. (2003). Bacteria–host communication: The language of hormones. *Proceedings of the National Academy of Sciences*, 100(15), 8951–8956. doi: 10.1073/pnas.1537100100
- Stewart, R. C. (2010). Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. *Current Opinion in Microbiology*, 13(2), 133–141. doi: 10.1016/j.mib.2009.12.013
- Stock, A. M., Robinson, V. L., & Goudreau, P. N. (2000). TWO-COMPONENT SIGNAL TRANSDUCTION. *Biochemistry*, 69(1), 183–215. doi: 10.1146/annurev.biochem.69.1.183
- Stock, J. B., Stock, A. M., & Mottonen, J. M. (1990). Signal transduction in bacteria. *Nature*, 344(6265), 395–400. doi: 10.1038/344395a0
- Szurmant, H., & Hoch, J. A. (2010). Interaction fidelity in two-component signaling. *Current Opinion in Microbiology*, 13(2), 190–197. doi: 10.1016/j.mib.2010.01.007
- Tremblay, Y. D., Deslandes, V., & Jacques, M. (2013). *Actinobacillus pleuropneumoniae* genes expression in biofilms cultured under static conditions and in a drip-flow apparatus. *BMC Genomics*, 14(1), 364–364. doi: 10.1186/1471-2164-14-364
- Typas, A., & Sourjik, V. (2015). Bacterial protein networks: properties and functions. *Nature Reviews Microbiology*, 13(9), 559–572. doi: 10.1038/nrmicro3508
- Ventura, B. D., & Sourjik, V. (2011). Self-organized partitioning of dynamically localized proteins in bacterial cell division. *Molecular Systems Biology*, 7(1), 457–457. doi: 10.1038/msb.2010.111
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. doi: 10.1093/bioinformatics/btp033
- Weigt, M., White, R. A., Surmant, H., Hoch, J. A., & Hwa, T. (2009). Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1), 67–72. doi: 10.1073/pnas.0805923106
- Wolfe, A. J. (2010). Physiologically relevant small phosphodonors link metabolism to signal transduction. *Current Opinion in Microbiology*, 13(2), 204–209. doi: 10.1016/j.mib.2010.01.002
- Wolfe, A. J., Parikh, N., Lima, B. P., & Zemaitaitis, B. (2008). Signal Integration by the Two-Component Signal Transduction Response Regulator CpxR. *Journal of Bacteriology*, 190(7), 2314–2322. doi: 10.1128/jb.01906-07
- Xu, Z., Chen, X., Li, L., Li, T., Wang, S., Chen, H., & Zhou, R. (2010). Comparative Genomic Characterization of *Actinobacillus pleuropneumoniae*. *Journal of Bacteriology*, 192(21), 5625–5636. doi: 10.1128/jb.00535-10

- Xu, Z., Zhou, Y., Li, L., Zhou, R., Xiao, S., Wan, Y., ... Chen, H. (2008). Genome Biology of *Actinobacillus pleuropneumoniae* JL03, an Isolate of Serotype 3 Prevalent in China. *PLoS ONE*, 3(1), e1450. doi: 10.1371/journal.pone.0001450
- Yan, K., Liu, T., Duan, B., Liu, F., Cao, M., Peng, W., ... Bei, W. (2020). The CpxAR Two-Component System Contributes to Growth, Stress Resistance, and Virulence of *Actinobacillus pleuropneumoniae* by Upregulating *wecA* Transcription. *Frontiers in Microbiology*, 11, 1026. doi: 10.3389/fmicb.2020.01026
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3), 1496–1503. doi: 10.1073/pnas.1914677117
- Yang, M., Derbyshire, M. K., Yamashita, R. A., & Marchler-Bauer, A. (2020). NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. *Current Protocols in Bioinformatics*, 69(1), e90. doi: 10.1002/cpbi.90
- Yoshida, T., Cai, S. jian, & Inouye, M. (2002). Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator. *Molecular Microbiology*, 46(5), 1283–1294. doi: 10.1046/j.1365-2958.2002.03240.x
- Zhan, B., Angen, Ø., Hedegaard, J., Bendixen, C., & Panitz, F. (2010). Draft genome sequences of *Actinobacillus pleuropneumoniae* serotypes 2 and 6. *Journal of Bacteriology*, 192(21), 5846–5847. doi: 10.1128/jb.00867-10