LinkedIn™: How Pathogenic Effectors Network with Host Proteins for Survival

By

Seemay Chou

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Tom Alber, Chair
Professor Qiang Zhou
Professor Michael Levine
Professor Zinmay Renee Sung

Fall 2011

LinkedIn™: How Pathogenic Effectors Network with Host Proteins for Survival

By

Seemay Chou

Abstract

LinkedIn™: How Pathogenic Effectors Network with Host Proteins for Survival

by

Seemay Chou

Doctor of Philosophy in Molecular and Cell Biology

University of California at Berkeley

Professor Tom Alber, Chair

Intracellular pathogens depend on their hosts for survival, and pathogen effectors that function at the host interface are critical to this process. Pathogens regulate their hosts through a combination of offensive and defensive strategies - capturing host machinery to carry out virulence functions while evading detection and actively inhibiting the host immune response. Here I report important advances resulting from studies of two relationships, including the oomycete *Hyaloperonospora arabidopsidis* (*Hpa*), which infects the model plant *Arabidopsis thaliana,* and the retrovirus HIV-1, which infects humans.

Hpa secretes a family of virulence effectors into *Arabidopsis* cells where they interact with host targets. Activation of the plant immune response against *Hpa* requires physical recognition of effectors by host Resistance proteins (R-proteins), and *Hpa* escapes host detection through genetic variation of effector recognition surfaces. To understand the basis for RPP1 recognition, I determined the 2.3-Å resolution crystal structure of one *Hpa* effector, *Arabidopsis thaliana* Recognized 1 (ATR1), which is recognized by the *Arabidopsis* R-protein Recognition of Peronospora Parasitica 1 (RPP1). We found that RPP1 recognizes distributed surfaces of ATR1, and different alleles of RPP1 recognize distinct surfaces of the effector. ATR1 belongs to an ancient family of conserved oomycete effectors that evolves rapidly through surface polymorphisms to escape host recognition while maintaining a conserved structural core.

To explore the mechanisms Human immunodeficiency virus (HIV) pathogenesis, I focused on the HIV accessory protein, Tat, which hijacks host machinery to stimulate viral transcription. HIV Tat recruits human positive transcription elongation factor b (P-TEFb) to the HIV LTR, and this recruitment is required for

elongation of viral transcripts. Using a proteomics-based approach, we discovered that P-TEFb associates with several new components, including AFF4, ELL2, and homologs ENL and AF9, that form a larger Super Elongation Complex (SEC). Through *in vitro* reconstitution using purified recombinant proteins, I showed that AFF4 is the major scaffold that directly binds other subunits via short interaction domains. Using limited proteolysis and binding assays to structurally characterize the SEC, I found that AFF4 is an extensively disordered scaffold that remains flexible even upon binding to SEC components. Short, hydrophobic peptides on an otherwise hydrophilic, serine-rich landscape form the binding domains on AFF4. Moreover, ELL2 and ENL function as bridging components that link the SEC to another transcriptional regulator, the PAF complex (PAFc). This work establishes the overall architecture of the SEC and provides insight into how Tat, by binding P-TEFb, can recruit a large ensemble of scaffolded transcriptional regulators to stimulate viral transcription.

Together, these studies provide new insights into pathogenic mechanisms. By characterizing the structural underpinnings of interactions at two different host-pathogen interfaces, we better understand how one effector, *Hpa* ATR1, escapes host detection and how another effector, HIV-1 Tat, finds and recruits multiple host factors to serve a virulence function.

Acknowledgements

I would first like to thank my family, who has always supported my academic and intellectual pursuits. From a young age, my parents have instilled in me a natural curiosity that has fueled my love for research. They have always shown so much pride in my work – and for that, I will always be grateful. I owe a special thanks to my brother See-Pong, who has always been my cheerleader and caretaker – my continual safety net. He has provided a sense of stability to my life since I was a kid, and this thesis is, in part, a culmination of many sacrifices he made to allow me this opportunity. The gummy bear and BBQ care packages were, of course, essential to my success as well. I also want to thank my good friends from Texas, who are practically family – Brenda, Maaya, Jenn, Josh.

I thank Drs. Greg Clark and Stan Roux at UT Austin for allowing me to volunteer in their lab early on in college. Greg was the first to teach me how to pipette, and these initial positive experiences were what drew me into science. Also, Amro Hamdoun, who allowed me to work with him for two summers at Hopkins Marine Station, was the first person to teach me how to ask a question and answer it at the bench. I will always remember our struggle and success in sequencing a sea urchin gene, and his inexplicable faith in me since has kept my path straight during difficult times.

My graduate school experience has involved countless characters who have made it an enjoyable experience, even in light of high drama at times. First, I thank my Meyer Lab homies. Finding an awesome group of smart people with a unique sense of humor to complain with over beers was the key to subsisting in the face of insanity. I will always admire Becky for reminding me the importance of life outside of work, and I will cherish my memories of Protein Snacktime with Bobby and his biceps. Aaron – you always provided a sense of comfort and laughter, despite (or perhaps because of) your sarcasm. Your genuine love of science is a true inspiration, and you deserve credit for making happy hour extra happy. I also thank Dave, the first person that made me cry in grad school, for toughening my skin and sharing all his knowledge about Chinese culture with me. Without his boot camp training, I could not have garnered the courage to later slay the dragon.

I am also grateful that Tom Alber gave me the opportunity to change course in grad school and join his group. His general positive spirit and the confidence he had in me helped restore the joy of science for me. I feel fortunate to have an advisor that was also my friend, and I am proud of the work we have accomplished together. I will dearly miss all the other good friends in the Alber group that have kept me laughing and looking forward along the way. James Fraser, Noelle Lombana, Ho Leung Ng were my original neighbors and helped me tremendously in the beginning. James was like a big brother to me in the lab (and a little brother to me outside of lab), and I am so glad that the twists and

turns of grad school allowed me that friendship. I also was fortunate to have several very talented undergraduates and rotation students, including Nishant Bhat, Anna Nowak, Avi Samelson, Heather Upton, and John Huizar. Avi was such a refreshing presence in the group, and brought so much life to the project. I will never forget our superglue projects! I look forward to visiting and pantsing him. I also thank friends in the lab – Daniela Mavrici, HoJun Lee, and others, who have suffered my pranks and made coming into lab more fun. I also could not have made it through without Katherine Bao, who was our lab manager extraordinaire. Over the last few years, we worked side-by-side, swimming through endless constructs and protein preps, to push our project forward. This work truly could not have moved without our partnership. She was also an incredibly loyal friend and good listener – I would not have survived without her.

Of course, I also thank all my good friends throughout grad school – Galo, Becky, SRod, Panda, Sean, Jose, Alberto, Malik, Jefferson, David, Jaime, Ewald, Emily – I'm not sure when I'll meet another group quite like this. I will miss beer hour shenanigans, ChickenTown nights, water balloon fights, margarita baseball, and pestering you all in general. I especially want to thank Galo, who was an amazing roommate and friend and patiently listened to my daily ramblings about lab life. Our scientific discussions were essential to my progress, but also made the whole experience more fun. I feel so lucky to know Galo and am always inspired by his pure love of science. Thank you for the Battlestar Gallactica/Wendy's nights… hopefully we will have more Wendy's nights in the future!

Finally, I owe so much to Christoph Grundner, who has been a wonderfully patient and supportive partner, scientific advisor, and best friend. It is difficult to put into words what he has meant to me over the last few years. He has made every aspect of my life more rich and I am so lucky to have someone like Christoph, who I myself admire so much, stand behind me. I look forward to climbing more mountains together.

Chapter 1

Lessons Mined from Interactions at the Host-Pathogen Interface

1.1 Introduction

In this modern day, we are obsessed with understanding the complex networks we are a part of. As evidenced by the emergence of various networking tools, such as LinkedIn™ or Facebook™, we believe that understanding the connections we share with individuals as well as with groups of individuals offer a window into who we are in a larger sense. Defining links within our networks also allows us to better ourselves – to identify opportunities for both social and professional expansion. In this way, we have borrowed a page from the storybook of intracellular pathogens, which have long been masters of networking with their hosts for survival. Pathogens have intimate symbiotic relationships with the host systems they live in. Pathogens not only co-opt host machinery for growth, but they also fend off host attack by evading detection or actively shutting down host immune responses. Thus, characterizing how microbes plug into host networks have greatly advanced our understanding of both pathogens and ourselves - elucidating pathways essential for microbial growth as well as fundamental mechanisms of host biology.

One major survival strategy for pathogens involves the deployment of virulence effectors into the host to actively suppress host immune defenses against invasion. For example, bacteria have numerous specialized secretion systems, including type III, type IV, and type V, that deliver specific bacterial effector proteins with host targets (1-4). Interestingly, many bacterial effectors structurally and functionally mimic host proteins in order to exert their virulence effects. The plant pathogen, *Pseudomonas syringae,* translocates an effector protein, AvrPtoB, into plant cells where it inhibits host programmed cell death associated with immunity. Biochemical and structural analyses reveal that AvrPtoB is an E3 ubiquitin ligase with host targets (5,6), further underscoring the important theme of host mimicry.

Clearly, learning how to cope with the host defense responses is critical to pathogen survival, but pathogens have also developed strategies that take advantage of host resources to enhance their own growth. Viruses are, perhaps, the greatest examples in this regard as they lack their own metabolism and depend completely on their hosts. Thus, the fate of the virus is closely intertwined with the host, and viral-host complexes play central roles throughout the virus life cycle by not only remodeling host cell physiology and but also by directing host machinery towards viral cellular processes (7,8).

The human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS), is unique from other retroviruses in its ability to co-opt host proteins for viral growth, latency and replication. In addition to its major structural components, HIV has numerous regulatory and accessory proteins that physically recruit host cellular machinery for viral functions (7,8). For example, early in its life cycle, the HIV genome integrates into the human genome, and this step requires the virally encoded integrase enzyme (IN). IN

forms a large pre-integration complex (PIC) with cellular proteins, such as the IN interactor protein (INI1), the P300 acetyltransferase, lens epithelium-derived growth factor (LEDGF/p75) (9,10). However, the precise organization of IN-containing complexes is unclear, resulting in a poor understanding as to how IN recruits host proteins to integrate DNA (11).

After integration of the HIV genome, another viral regulatory protein, Tat, stimulates transcription from the viral promoter. Tat increases processivity of human RNA polymerase II (Pol II) through its ternary interactions with a nascent viral mRNA structure (TAR) and human elongation factor, P-TEFb (12). In this way, Tat recruits of P-TEFb specifically to the HIV LTR, and disruption of these interactions results in abortive HIV transcripts. HIV Tat also releases cellular P-TEFb from its inactive form, the 7SK snRNA- and HEXIM1-containing complex known as 7SKsnRNP (12). However, our incomplete characterization of the structures of P-TEFb complexes limits our understanding of the mechanisms by which HIV Tat manipulates its host network.

Altough pathogens effectivety manipulate host networks via effector proteins, hosts have, in response, developed surveillance strategies that detect invasion. The eukaryotic immune system senses the presence of pathogens and subsequently initiates a cascade of defense responses. For example, the innate immune system detects general molecular signatures of microbes and viruses by taking advantage of conserved microbial features, such as pathogen-associated molecular patterns (PAMPs) (13). PAMPs are recognized by a diverse family of Pattern Recognition Receptors (PRRs), which can be cytoplasmic, membrane-bound, or extracellular. PRRs include transmembrane toll-like receptors (TLRs), which recognize PAMPs through conserved leucine-rich repeat (LRR) domains and trigger pro-inflammatory cytokine secretion (13). Mammals also have a class of nucleotide-binding, oligomerization domain (NOD)-like receptors (NLRs) that also recognize pathogenic products, such as bacterial peptidoglycan or viral ssRNA, via an LRR domain (14). Recognition of these microbial signatures by NLRs leads to the activation of a diverse set of immune responses, including inflammation and induction of autophagy (14).

Plants also have sophisticated detection mechanisms, which share many components with the mammalian innate immune system. Plant immunity consists of two types of innate resistance, including PAMP-triggered immunity (PTI) and Effector-Triggered immunity (ETI). PTI recognizes PAMPs via PRRs located primarily on the cell surface, while ETI targets specific effectors released by adapted pathogens through direct and indirect recognition of effectors by host Resistance proteins (R-proteins) (15). R-proteins also have a nucleotide-binding leucine-rich region (NB-LRR), which is a conserved domain in mammalian NLRs. Moreover, some plant R-proteins also have a Toll—Interleukin-1 Receptor (TIR) domain (15). Recognition of effectors by their cognate R-proteins leads to activation of the plant immune response, which often includes a local cell death response known as the Hypersensitive Response (HR) (16). Although there has

been considerable progress with regard to identifying virulence effectors recognized by R-proteins, the molecular underpinnings of these interactions remain elusive. Future studies that shed light on exactly how physical recognition or evasion are achieved will provide insight into the mechanisms of the evolutionary struggle between host and pathogen.

The molecular surfaces at the host-pathogen interface offer therapeutic opportunities to inhibit specific pathways without disrupting host functions. To this end, high-resolution structures of host-pathogen complexes are key templates for designing novel drugs. One major obstacle for structurally characterizing HIV-human complexes is the inherent structural plasticity of many HIV proteins, which may be a built-in trait of viral proteins adapting to a diverse array of cellular partners. For example, HIV Tat is a natively unfolded protein, which has long been biochemically intractable. However, recently the three-dimensional crystal structure of Tat-bound P-TEFb reveals a strikingly intimate folding of Tat onto its host partner (17) that is truly reminiscent of a glove fitting onto a hand (Figure 1.1). Tat alone has few intramolecular contacts compared to the numerous intermolecular interactions between Tat and P-TEFb, highlighting the central role P-TEFb plays in facilitating the structural ordering of Tat.

Another roadblock to elucidating the molecular details of viral-host networking is the incomplete identification of all cellular factors recruited to perform viral functions. Although there has been significant progress with finding major binding partners of HIV proteins, the physical complexes these binding partners function as a part of have remained largely ignored until more recent years. HIV Tat was discovered in 1987 as a key regulator of viral transcription (18), and P-TEFb was identified in 1997 as a key Tat co-factor (19,20). However, it was not until more than a decade later that we and others identified important additional components of the Tat/P-TEFb complex (21,22), now collectively known as the Super Elongation Complex (SEC). Although numerous genome-wide genetic screens have been employed to find novel host factors involved in HIV regulation in recent years, the poor overlap between these data sets has limited their usefulness (8,23-25). Thus, there has been increasing interest in proteomics-based approaches as an unbiased strategy for defining the physical make-up of the pathogen-host interface for HIV and other pathogens (26,27).

Here, I report a body of work aimed at expanding our biochemical and structural understanding of host-pathogen interactions in two different examples from biology. I explore the structural basis of host recognition of a virulence effector, ATR1, secreted by *Hyaloperonospora arabidopsidis* (*Hpa*) into *Arabidopsis* cells (Chapter 2). Host immune response against *Hpa* hinges on detection of pathogenic effectors by host Resistance proteins (R-proteins), and our work shows that this detection is based on physical recognition of a wide surface of ATR1 by host R-protein RPP1. I also discuss findings that reveal how *Hpa* escapes host detection by modulating recognition surfaces while maintaining a hydrophobic core conserved in other effectors. In a second study, I investigate the mechanisms by which an HIV accessory protein, Tat, recruits host proteins to

enhance viral transcription (Chapters 3 and 4). We find that, in addition to the known Tat co-factor, P-TEFb, several additional host elongation factors are recruited to the HIV LTR as part of a larger Tat:P-TEFb complex to stimulate elongation (Chapter 3). I discuss results that expand our structural understanding of this complex, the Super Elongation Complex (SEC), and present an emerging model that SEC assembles on a highly disordered scaffold with distributed binding sites that connect SEC to a larger network of transcriptional regulators (Chapter 4). Tat exploits this organization by binding CycT1 of P-TEFb and, like pulling on a handle, subsequently recruits several classes of elongation factors connected along a scaffold. From these *Hpa* and HIV studies emerge two different stories about pathogen survival and exemplify how a precise understanding of the structural and molecular bases for interactions at the host-pathogen interface provide a window into fundamental mechanisms of host biology.

1.2 Figures

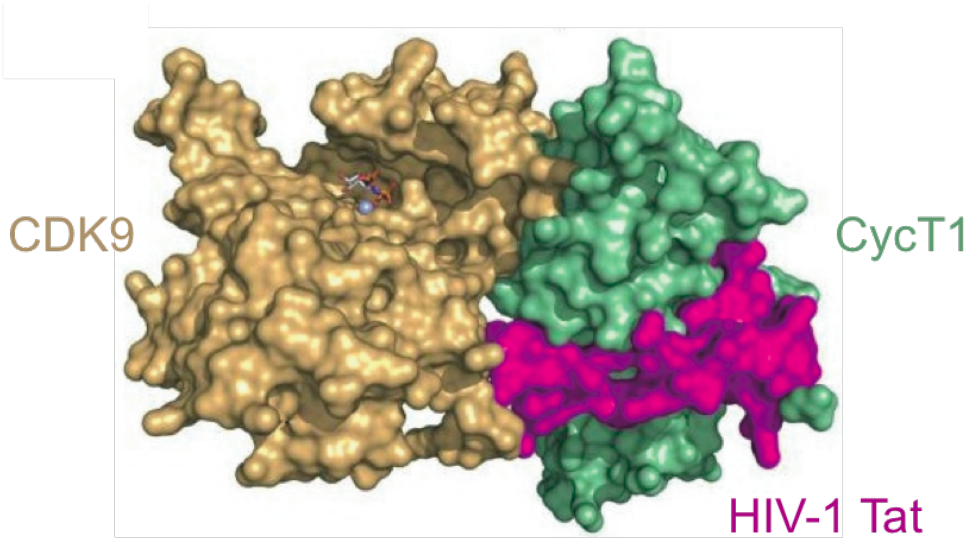


Figure 1.1. HIV-1 Tat folds onto host P-TEFb. Crystal structure of HIV-1 Tat-bound P-TEFb reveals intimate interactions between viral Tat and its host co-factor (17). Cys261 of host CycT1 forms a Zn-mediated bridge to Tat.

1.3 References

1. Galan, J. E. (2009) *Cell Host Microbe* **5**, 571-579
2. Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S., and Cascales, E. (2005) *Annu Rev Microbiol* **59**, 451-485
3. Galan, J. E., and Wolf-Watz, H. (2006) *Nature* **444**, 567-573
4. Filloux, A., Hachani, A., and Bleves, S. (2008) *Microbiology* **154**, 1570-1583
5. Abramovitch, R. B., Janjusevic, R., Stebbins, C. E., and Martin, G. B. (2006) *Proc Natl Acad Sci U S A* **103**, 2851-2856
6. Janjusevic, R., Abramovitch, R. B., Martin, G. B., and Stebbins, C. E. (2006) *Science* **311**, 222-226
7. Swanson, C. M., and Malim, M. H. (2008) *Cell* **133**, 742, 742 e741
8. Brass, A. L., Dykxhoorn, D. M., Benita, Y., Yan, N., Engelman, A., Xavier, R. J., Lieberman, J., and Elledge, S. J. (2008) *Science* **319**, 921-926
9. Poeschla, E. M. (2008) *Cell Mol Life Sci* **65**, 1403-1424
10. Sherman, M. P., and Greene, W. C. (2002) *Microbes Infect* **4**, 67-73
11. Cherepanov, P., Maertens, G. N., and Hare, S. (2011) *Curr Opin Struct Biol* **21**, 249-256
12. Zhou, Q., and Yik, J. H. (2006) *Microbiol Mol Biol Rev* **70**, 646-659
13. Kumar, H., Kawai, T., and Akira, S. (2011) *Int Rev Immunol* **30**, 16-34
14. Elinav, E., Strowig, T., Henao-Mejia, J., and Flavell, R. A. (2011) *Immunity* **34**, 665-679
15. Bernoux, M., Ellis, J. G., and Dodds, P. N. (2011) *Curr Opin Plant Biol*
16. Coll, N. S., Epple, P., and Dangl, J. L. (2011) *Cell Death Differ* **18**, 1247-1256
17. Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C., and Price, D. H. (2010) *Nature* **465**, 747-751
18. Kao, S. Y., Calman, A. F., Luciw, P. A., and Peterlin, B. M. (1987) *Nature* **330**, 489-493
19. Mancebo, H. S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D., and Flores, O. (1997) *Genes Dev* **11**, 2633-2644
20. Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., and Price, D. H. (1997) *Genes Dev* **11**, 2622-2632
21. Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010) *Mol Cell* **38**, 439-451
22. He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., and Zhou, Q. (2010) *Mol Cell* **38**, 428-438
23. Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., Wyniger, J., Descombes, P., Antonarakis, S. E., Letvin, N. L., McMichael, A. J., Haynes, B. F., Telenti, A., and Goldstein, D. B. (2007) *Science* **317**, 944-947

24. Zhou, H., Xu, M., Huang, Q., Gates, A. T., Zhang, X. D., Castle, J. C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D. J., and Espeseth, A. S. (2008) *Cell Host Microbe* **4**, 495-504
25. Bushman, F. D., Malani, N., Fernandes, J., D'Orso, I., Cagney, G., Diamond, T. L., Zhou, H., Hazuda, D. J., Espeseth, A. S., Konig, R., Bandyopadhyay, S., Ideker, T., Goff, S. P., Krogan, N. J., Frankel, A. D., Young, J. A., and Chanda, S. K. (2009) *PLoS Pathog* **5**, e1000437
26. Jager, S., Gulbahce, N., Cimermancic, P., Kane, J., He, N., Chou, S., D'Orso, I., Fernandes, J., Jang, G., Frankel, A. D., Alber, T., Zhou, Q., and Krogan, N. J. (2011) *Methods* **53**, 13-19
27. Dyer, M. D., Murali, T. M., and Sobral, B. W. (2008) *PLoS Pathog* **4**, e32

,

Chapter 2

The *Hyaloperonospora arabidopsidis* ATR1 Effector Escapes Host Detection Through Surface Variations

2.1 Abstract

The *in planta* association of the *Hyaloperonospora arabidopsidis* effector ATR1 with the cognate *Arabidopsis thaliana* RPP1 immune receptor activates a disease-resistance signaling pathway that inhibits pathogen growth. To define the molecular events specifying effector recognition by RPP1, we determined the crystal structure of ATR1 and assayed *in planta* the effects of surface polymorphisms that are critical to activating plant immunity. ATR1 adopts an elongated, all-helical, two-domain seahorse-like structure with an overall architecture unlike any previously described fold. Structural analyses identify a tandemly duplicated, five-helix motif in the C-terminal domain that creates a structural framework for rapid diversification. Moreover, comparison with another oomycete effector, Avr3a11, reveals a common hydrophobic core that creates a structural signature in a subset of oomycete effectors. Identification and mapping of critical recognition sites suggests that ATR1 detection by the RPP1 resistance protein is mediated by several distinct protein surfaces that allow the effectors to escape recognition through diverse surface polymorphisms. ATR1 gain-of-recognition mutants demonstrate that multiple amino acid substitutions are necessary for recognition and that surface polymorphisms exert additive effects. These results suggest that ATR1 is a modular repeat protein belonging to an ancient family of oomycete effectors that rapidly evolves to escape host detection and adopt diverse virulence functions.

2.1 Introduction

Oomycetes form a monophyletic group of organisms that morphologically resemble fungi but are evolutionarily more closely related to brown algae and *Alveolates* (1). Oomycetes include a variety of commercially important plant pathogens with a diverse range of hosts, such as *Phytophthora infestans,* which causes tomato and potato late blight, *P. sojae* - soybean stem and root rot, *P. ramorum* - sudden oak death, and *Plasmopara viicola* – grapevine downy mildew, as well as *Hyaloperonospora arabidopsidis (Hpa*, previously known as *Peronospora parasitica)*, a pathogen of the model plant, *Arabidopsis thaliana*. Similar to many other plant and animal pathogens that deliver virulence effectors into the host to establish infection, oomycetes physically interact with their hosts through specialized haustorial feeding structures that facilitate the delivery of effector proteins into host cells where they have intracellular targets and play critical roles in oomycete survival and growth (2). Despite substantial progress toward characterizing the roles of effectors, the unifying mechanisms by which oomycete effectors promote virulence remain largely unknown.

Oomycete effector genes have a number of conserved features. Although the mechanisms of effector translocation are not well understood, a typical eukaryotic signal sequence found in all effectors is thought to mediate secretion out of the pathogen. In most effectors, oomycete-specific RxLR and dEER motifs promote further translocation of effectors into the host cell (3-6). Recently, genome sequences of *Hpa* and *Phytophthora* species allowed identification of numerous genes containing the conserved N-terminal signal peptide and RxLR-dEER motifs (7,8). A large superfamily of *Phytophthora* effectors contains conserved C-terminal W-, Y- and/or L-motifs that are often repeated in the protein (9).

The *Hpa/Arabidopsis* system has been adopted as a model system for studying plant-oomycete interactions. The *Hpa* genome encodes 135-150 putative RxLR-containing effectors (7,8), including *Arabidopsis thaliana* Recognized 1 (ATR1) (10). During the course of co-evolution, plants have developed surveillance systems dependent on highly polymorphic Resistance proteins (R-proteins). R-proteins directly or indirectly detect pathogen-derived effector molecules (11-13) to induce a cascade of immune responses that are collectively known as effector-triggered immunity (ETI) (14,15). The hallmark of ETI responses is localized cell death, called the hypersensitive response (HR). Thus, pathogen effectors such as ATR1 have dual effects in promoting pathogen growth yet mediating recognition by the plant immune system through R-proteins.

Several dozen *Arabidopsis* R-genes*,* including Recognition of Peronospora Parasitica 1 (RPP1), confer resistance to *Hpa* (10,16,17). RPP1, through its polymorphic Leucine Rich Repeat (LRR) domain, associates with ATR1 variants, leading to activation of plant disease resistance (16), This model of direct, physical interactions between ATR1 and RPP1 is supported by the apparent

positive selection for sequence polymorphisms in both ATR1 and RPP1 (17,18). Moreover, only specific pairs of ATR1 and RPP1 alleles mediate recognition and immunity (19), suggesting a form of immune response in plants that evolves rapidly to adapt to the cognate pathogen.

To explore the basis for ATR1 recognition by RPP1 and the mechanisms by which ATR1 mutations mediate escape from the host HR, we determined the crystal structure of ATR1 at 2.3-Å resolution. ATR1 is a monomeric, modular protein with two structural domains comprised of α-helices. Two structurally similar five–helical repeats that display no sequence homology to each other form the C-terminal domain. Deletion analysis shows that segments of both structural domains are required for recognition of ATR1 by RPP1. Naturally occurring polymorphisms of ATR1 allowed identification of surface residues critical for recognition. Multiple polymorphisms are needed to switch the specificity of ATR1 alleles, and the differential effects of mutations show that different alleles of RPP1 may recognize distinct surfaces of ATR1. Our results show that ATR1 belongs to an ancient family of conserved oomycete effectors that evolves rapidly through surface polymorphisms to escape host recognition while maintaining a conserved structural core.

2.3 ATR1 structure

To determine the ATR1 structure, several protein variants were expressed and purified from *E. coli.* ATR1Δ15, which included the RXLR/dEER translocation motif, expressed but was not amenable to crystallization. Focusing on the effector domain sufficient for recognition by RPP1, we determined the crystal structure of ATR1Δ51 (Figure 2.1A) from *Hpa* Emoy2 at 2.3-Å resolution. Initial phases were generated by multiwavelength anomalous diffraction, and the model was refined to R/R_{free} values of 0.2231/0.2598 (Table 2.1). Three copies of ATR1Δ51 crystallized in the asymmetric unit (AU). The three molecules in the AU are similar, with a Cα root-mean-square deviation (RMSD) of 0.43 Å. No clear electron density was obtained for the N-terminus (residues 51-62 of chain A, 51-62 of chain B, and 51-66 of chain C) and a loop connecting α12 and α13 (residues 278-290 of chain A, 280-289 of chain B, and 282-289 of chain C).

ATR1 adopts a two-domain, extended, seahorse-like structure comprised of 13 α-helices (Figure 2.1B). The N-terminal head (α1-α3) forms a three-helix bundle separated from the larger C-terminal body (α4-α13) by a loop, or neck region (amino acid residues 117-126). The neck contains two β-turns, as well as several hydrogen bonds between Leu122 and Tyr126, Gly120 and His123, His 123 and Thr125, and Asp124 and Asp127 (Figure 2.2). Analysis of the electrostatic surface potential of ATR1 reveals numerous distributed positively and negatively charged patches, including a major region of positive potential on the head and two major negatively charged regions on the body (Figure 2.3). Mapping the sequence conservation among ATR1 alleles shows that polymorphic residues are distributed across the surface of the head, neck and body (Figure 2.4). Hydrophobic and aromatic patches occur on the exposed surfaces of helices α2 and α3 in the head domain, as well as a C-terminal pocket in the groove between helices α11, α12 and α13 containing six Phe or Tyr residues. The first ordered residues in the structure, Trp-Pro-Phe 63-65, are unusually exposed for such hydrophobic amino acids.

Comparison of the entire ATR1 effector domain with available structures using the DALI server did not reveal any significant structural homologs. Separate analysis of the N-terminal head and C-terminal body identified several potential distant homologs, with the circadian regulator KaiA (PDB ID 1R5Q) giving the best match (Z score = 5.8) to an X-type, four-helix-bundle segment (Figure 2.5A). ATR1 α5-α8 aligns with four helices of KaiA (residues 13-91) with an RMSD of 2.7 Å. However, the electrostatic surfaces of the aligned structures are distinct. ATR1 also has a more extended loop between α5 and α6, and the KaiA helices analogous to ATR1 α7 and α8 are longer (Figure 2.5A). To test the putative role of ATR1 in clock regulation, we measured *Arabidopsis* circadian rhythms via the TOC1:LUC reporter (Materials and Methods) in the absence and presence of ATR1. ATR1 had no effect on the transcriptional control of the circadian clock (Figure 2.5B-D).

The C-terminal body domain of ATR1 contains a structural repeat that is not evident in the sequence (Figure 2.6A). Helices α4-α8 (residues 126-208) form a five-helix subdomain that resembles the next five-helix segment (α9-α13, residues 212-311; Cα RMSD=5.1 Å) (Figure 2.6B-C). This ATR1 structural repeat is comprised of a capping helix crossing the first two helices of an X-type, antiparallel four-helix bundle. The arrangement of helices α4-α6 is particularly similar to α9-α11 (Ca RMSD=3.2 Å). The fourth and fifth helices of the two repeats are not only more structurally variable, but they also are connected by the longest loops (11 and 19 residues, respectively) in the repeats. Because the first helix in the first ATR1 repeat is nearly parallel to the first helix in the second repeat, the variable loops fall on the same side of the elongated structure. A structure-based sequence alignment shows only 4% identical residues (Figure 2.6D), making this homology undetectable by amino acid sequence comparisons.

ATR1 packed in the crystals as two equivalent dimers with one formed by a crystallographic two-fold rotation axis and the other formed by a noncrystallographic two-fold. To determine if this dimer reflects solution properties of ATR1, we analyzed the oligomerization state of recombinant protein in solution using size exclusion chromatography (SEC). At 1 mg/mL (28 μM) in neutral pH buffer, ATR1 eluted at a volume corresponding to 30.2 kDa, similar to the molecular weight of a monomer, 29.5 kDa (Figure 2.7A). To test ATR1 stoichiometry *in vivo*, we performed co-immunprecipitations using FLAG- and HA-tagged ATR1 transiently expressed in *Nicotiana tabacum*. We used HA-tagged LRR of RPP1 as a positive control, as it has been shown to interact with ATR1 (10). FLAG-ATR1 was co-expressed with either HA-ATR1 or HA-LRR and immunoprecipitated using anti-FLAG M2 Sepharose. HA-ATR1 failed to co-immunoprecipitate with FLAG-ATR1, showing that ATR1 does not form homo-oligomers *in vivo* (Figure 2.7B).

2.4 ATR1 shares a sub-domain with the *Phytophthora* effector Avr3a11

Recently, the three-dimensional structure of the effector Avr3a11 from *Phytophthora capsici* was determined (unpublished data), enabling comparisons with ATR1. Unexpectedly, Avr3a11 residues 68-132 structurally resemble a four-helix subdomain (residues 138-210) within the C-terminal body of ATR1 (Figure 2.8A). This homology is undetectable by primary sequence comparisons (sequence identity=3.6%), but structural alignment between ATR1 and Avr3a11 reveals good agreement between the two regions (Cα RMSD = 2.8 Å) as well as conservation of several buried hydrophobic residues that make contacts with each other in the respective structures (Figure 2.8B). The electrostatic surfaces in this region of ATR1 and Avr3a11 show considerable variation, despite similarity of overall fold (Figure 2.8C). In contrast to Avr3a, this region of ATR1 occurs within the C-terminal body domain, and the conserved hydrophobic core is extended at each end through extensive contacts with the adjacent helices α4 and α9.

The conserved ATR1 core residues (Ile147, Val170, Trp174, Gly178, Tyr179, and Thr180) align structurally with the W-motif found in a large family of *Phytophthora* effectors (20). To determine whether this structural homology reflects a shared virulence function, we tested if ATR1 inhibits PAMP-induced cell death, as previously shown for Avr3a. However, unlike Avr3a effector from *P. infestans*, ATR1 does not suppress INF-dependent cell death, suggesting a unique role for this subdomain. Identification of residues in ATR1 structurally analogous to the W-motif allowed us to refine a search for the W-motif containing subdomain among other *Hpa* effectors. Using structure-based amino-acid alignment between ATR1 and Avr3a11 (Figure 2.8D) and iterative Hidden Markov Model (HMM)-based searches, we identified 24 additional proteins in the *Hpa* genome that contain a putative W-motif containing subdomain (e-value <1) (Table 2.2). In several proteins, this domain is present in 2 or 3 copies. A subset of the identified proteins also contained predicted secretion sequence (SignalP, HMM score > 0.75) and the effector specific dEER motif. Interestingly, one of the candidates (gene id 800198) had 3 tandem copies of the W-motif domain plus an N-terminal RxLR motif, but did not contain any detectible signal sequence (Table 2.2). Overall, the consensus signature of the W-motif in *Hpa* (Figure 2.8D) is substantially different from the consensus among the *Phytophthora* proteins (20).

2.5 ATR1 recognition by host RPP1

To elucidate the structural basis of ATR1 recognition by the host, we used deletion analysis of ATR1-Emoy2 to define the minimal region recognized by RPP1. We introduced deletion endpoints based on ATR1 secondary structure (Figure 2.9A-B) and assayed their activity by transient co-expression with RPP1-WsB in *Nicotiana tabacum* (Figure 2.9C). The localized cell death due to the HR was used as a marker for activation of RPP1-mediated defense responses. Deletions in ATR1Δ51 preserving or removing helix α1 (Δ67 and Δ87, respectively) did not affect activation of RPP1 but reduced ATR1 protein stability (Figure 2.9D). Further N-terminal deletions failed to induce RPP1-dependent HR. Deletion of the C-terminal 90 amino acids compromised protein stability but did not affect recognition, suggesting that residues 87-222 are sufficient for RPP1 recognition. Further C-terminal deletions resulted in loss of HR (Figure 2.9C), which may be due to lower protein stability or removal of critical amino acids. The minimal recognition region comprising amino acids 87-222 includes helices α2-α3 in the ATR1 N-terminal head and the first five-helix repeat (residues 127-210) in the C-terminal body (Figure 2.9B).

We have previously employed the natural polymorphisms between ATR1-Emoy2 and ATR1-Maks9 to identify two key residues that specify ATR1-dependent activation of RPP1-NdA (16). Here we used a similar approach employing the natural polymorphisms to define key amino acids that specify differential recognition of ATR1-Emoy2 and ATR1-Cala2 by RPP1-WsB. ATR1-Emoy2 and ATR1-Cala2 differ in 69 amino acid sites (Table 2.3), many of which are located in the C-terminal region of the protein. Since our deletion analysis identified the

minimum region of ATR1 sufficient for recognition, we focused on polymorphisms located within this region (residues 87-222). This restricted our analysis to 26 out of the 69 total polymorphisms. We further refined our analysis by looking for polymorphic sites that co-segregated between three ATR1 alleles recognized by RPP1-WsB (Emoy2, Maks9, Emco5) and two unrecognized alleles (Cala2 and Emwa1), yielding 14 sites that fulfilled these criteria (Figure 2.10A).

We individually mutated these sites in ATR1-Cala2 and assessed their relative contributions to activation of RPP1-NdA and RPP1-WsB in *N. tabacum* (Figure 2.10B). Substitutions at four sites produced gain-of-recognition phenotypes with RPP1-WsB that ranged from very mild (Asn158Lys) to intermediate (Val122Leu, Ser125Thr) to strong (Tyr140Asp) (Figure 2.10B). Combining the mutations had additive effects, and the quadruple ATR1-Cala2 mutant (Val122Leu/Ser125Thr/Tyr140Asp/Asn158Lys) induced HR with timing and intensity similar to wild-type ATR1-Emoy2 (Figure 11A-B,12). Interestingly, activation of RPP1-NdA was not affected by any of these mutations. The reciprocal quadruple substitution in ATR1-Emoy2 significantly delayed activation of RPP1-WsB (Figure 2.11B, 2.12), suggesting that although these four residues are sufficient to switch specificity, there are likely to be additional interaction sites. These ATR1 variants expressed to the same levels (Figure 2.11C), indicating that the changes in recognition specificity were not due to differences in protein stability.

Mapping these polymorphisms onto the structure of ATR1 shows that they all are surface-exposed, except Asp140, which is partially buried (Figure 2.10C). Most interestingly, the key ATR1 residues that are important for activation of RPP1-WsB versus RPP1-NdA are located on distinct protein surfaces. This indicates that variable RPP1 alleles are capable of recognizing different unrelated surface “epitopes” of ATR1.

2.6 Discussion

Although considerable progress has been made towards dissecting the molecular mechanisms underlying effector recognition by R-proteins and the structural basis for HR activation, virulence functions for many effectors remain elusive. Oomycete and fungal effectors have been shown to evolve under strong positive selection that drives rapid divergence, making it difficult to detect effector homologs outside their genus using amino-acid sequence comparisons or secondary-structure prediction tools. The three-dimensional structure of the *Hpa* effector ATR1 differs from that of other effector proteins, including AvrL567-A and AvrL567-D, two effectors from *Melampsora lini* (flax rust) differentially recognized by a cognate R-protein L (21), and *Pseudomonas syringae* effectors AvrPto and AvrB (22,23).

ATR1 folds into an elongated structure composed of two major helical domains connected by a linker containing several hydrogen bonds. The C-terminal domain is formed by an extended, right-handed solenoid of 10 helices that form two 5-helix repeats. This ATR1 structural repeat differs from previously defined helical repeat motifs. The ankyrin and tetratrichopeptide (TPR) repeats, for example, form pairs of nearly anti-parallel helices packed together in curved arcs (24). Similarly, the helices in HEAT and 14-3-3 repeats are longer and more nearly antiparallel, and these repeat arrays form curved architectures, rather than the straighter arrangement of the tandem ATR1 repeats. The clathrin repeat (25) forms a straight elongated structure like ATR1, but the clathrin helices are shorter and they cross each other in a more regular, nearly anti-parallel pattern.

The lack of significant sequence identity between the ATR1 repeats raises the question of whether these are genuine repeat sequences that arose by duplication and divergence. We note that many well-established helical repeat sequences show low pairwise sequence identity, and many motifs are recognizable only because they occur in tandem arrays. Similarly, most members of the superfamily of *Phytophthora* effectors containing W–, Y– or L–motifs show low pairwise sequence identity. Consequently, the tandem arrangement and structural similarity of the ATR1 repeats support the conclusion that this motif represents a new helical repeat that is likely to be found not only in ATR1, but also in other proteins.

Although unexpected from sequence comparisons, the structural homology of ATR1 α5-α8 to *P. capsici,* Avr3a11 defines and expands the role of the W-motif found in diverse *Phytophthora* effector proteins. The structural alignment of ATR1 and Avr3a11 shows that the W-motif comprises core hydrophobic residues that stabilize a shared four helical subdomain. This conserved core supports numerous surface variations that mediate escape from host recognition. This fold in ATR1 is not sufficient for suppressing INF1-induced cell death as seen with *P. infestans* Avr3a, which is consistent with the absence critical functional residues in ATR1 (26).

Interestingly, the W-motif sequence found in a family of *Phytophthora* effectors (9) differs from the structurally homologous W-motif in ATR1 and 24 additional proteins in *Hpa* (Figure 8C). This suggests that the W-motif is a structural signature that duplicated and elaborated in oomycete species after their divergence from the last common ancestor. The W-motif subdomain is appended to distinct domains in different proteins, suggesting that it functions as a modular, structural scaffold to enable functional diversification amongst effectors or allow rapid surface modifications to avoid host detection. Alternatively, this fold might have a specific biochemical function that is yet to be uncovered. Several W-motif proteins in the *Hpa* genome lack typical signatures of oomycete effectors, raising the possibility that this motif may have a more general function in oomycete biology.

Several putative distant homologs of ATR1 were found by searching for three-dimensional structures similar to individual ATR1 segments. The highest scoring structural homolog was KaiA, a cyanobacteria *Anabaena* sp PCC7120 circadian clock protein (27) that aligned with ATR1 helices α4-α7. In light of this putative homology and the previously characterized link between circadian regulation and plant immune defense (28), we tested the role of ATR1 in circadian regulation. The presence of ATR1, however, does not significantly alter transcriptional regulation of *Arabidopsis* circadian rhythms. Given that this domain of KaiA is involved in promoting the kinase activity of its interacting partner KaiC, however, it is possible that the analogous region in ATR1 performs a similar biochemical function but a different physiological function. Other lower scoring homologs include an RNA-binding protein RBP8 (aligns with α4-α9) and Skp1 (aligns with α4-α10), an adaptor protein in the human SCF E3 ubiquitin-ligase complex. These structural similarities may offer initial insights into biochemical functions and potential partners of ATR1.

Although dimerization of R-proteins appears to be required in order to activate immune responses (16,29) and several other fungal and bacterial effectors function as dimers (30), ATR1 behaves as a monomer *in vivo* and *in vitro*. These results suggest that ATR1 does not serve as a dimerization platform for RPP1. However, our results do not exclude the possibility that ATR1 may oligomerize in complex with RPP1 or other host partners.

ATR1-Emoy2 recognition by RPP1-NdA relies on two polymorphic sites in *Hpa* strains--Lys92 and Gly191 (16). These sites lie on opposite sides of the ATR1 structure in two different domains, suggesting that both positions can function in recognition. Unexpectedly, three of four natural ATR1 variations (amino acids 122, 125 and 158) associated with recognition by another allele, RPP1-WsB, occur on a different surface of the protein. In contrast, a fourth RPP1-WsB selective residue we identified (Asp140) occurs on the same surface as Gly191, which promotes activation through RPP1-NdA (16). This difference in the basis for activation specificity of Rpp1-WSB and RPP1-Nda supports the idea that

RPP1 directly recognizes ATR1 and suggests that these RPP1 alleles recognize distinct surface “epitopes” of ATR1. The existence of RPP1 alleles capable of recognizing different protein surfaces of ATR1 suggests that LRRs play a versatile role in plant immunity.

The additive effects of the ATR1 polymorphisms suggest that R-protein activation is more complicated than a simple on/off switch. Most likely, the response kinetics are controlled by different binding affinity of ATR1 variants. In turn, this can affect the release of negative regulation through either inter- or intramolecular interactions and/or nucleotide binding activitity of the NBS domain of RPP1. Defining the mechanisms that control kinetics of R-protein activation is the next critical step in understanding initiation of plant defense responses and plant cell death.

Our mutational analysis of ATR1 reveals that RPP1 exhibits rapidly evolving recognition of different ATR1 protein surfaces. This suggests that recognition of each effector could have evolved independently in closely related plant lineages. Changes in plant immune receptors happen in the germline, thus, plant immunity is innate. Thus, R-proteins, such as RPP1 that rapidly gain new recognition specificities provide an adaptive advantage to plant species on the population level. Interpreting ATR1 polymorphisms in the context of the ATR1 structure provides a framework for understanding how pathogens may escape detection and how plant hosts evolve in order to maintain effector recognition. Further understanding of the molecular mechanisms that allow R-proteins to respond to effectors could lead to engineering optimized plant pathogen receptors - potentially powerful new tools to contain some of the most important plant pathogens.

2.7 Materials and Methods

Strains and Growth Conditions

Bacterial DNA transformation was conducted using chemically competent *E. coli* DH5α (Invitrogen), electroporation of *E. coli* Rosetta (DE3) or through freeze/thaw transformation of $CaCl_2$ competent *A. tumefaciens* (31). *Nicotiana tabacum* (variety Turk), *Nicotiana benthamiana* and *Arabidopsis thaliana* plants were grown in a controlled growth chamber at 24 °C at 16-hour-light/8-hour-dark photoperiod before infiltrations and switched to 24 hour light after infiltrations.

ATR1 cloning, protein expression, and purification

The ATR1D51 Emoy2 deletion variant (16) was cloned into the pDUET vector (Addgene) after adding BamHI/NotI sites and a cleavage site for Tobacco Etch Virus (TEV) protease to the 5' and 3' ends of ATR1 through PCR amplification. For $(His)_6$-ATR1D51 protein expression, the pDUET construct was transformed into *E. coli* Rosetta (DE3) (Novagen), cells were grown at 37 °C to OD_{600}=0.7-0.8 and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 6-8 hours at 18 °C. The cells were harvested by centrifugation for 15 minutes at 5000 x g at 4 °C and frozen in liquid nitrogen for storage at -20 °C prior to purification.

Cells were resuspended in Ni-A Buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM TCEP, 25 mM imidazole, 10% glycerol, 0.2 mM AEBSF) and lysed by sonication. The lysate was centrifuged for 1 hr at 20000 x g (4 ^{o}C), and the supernatant was passed over a Ni affinity column (GE). $(His)_6$-ATR1D51 was recovered by gradient elution with Ni-B Buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM TCEP, 350 mM imidazole, 10% glycerol, 0.2 mM AEBSF) using the AKTA Explorer FPLC system (GE). Fractions containing $(His)_6$-ATR1D51 were verified by SDS-PAGE and pooled for tag cleavage for 22 hrs at 4 °C with 1:50 TEV protease. Untagged ATR1D51 was loaded on a Superdex S75 gel filtration column (GE) and eluted as a monomer in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP. The protein was concentrated to 20 mg/mL.

Structure determination and analysis

Preliminary ATR1Δ51 crystals were obtained by hanging drop vapor diffusion trials at 18 °C from a 2:1 mixture of 20 mg/mL protein with 0.1 M MES pH 6.5, 1.6 M $MgSO_4$. Diffraction quality crystals were obtained by hanging drop vapor diffusion at 18 °C by seeding at 1:20,000 into a 4:1 mixture of ATR1Δ51 at 15 mg/mL with 0.1 M MES pH 5.0, 1.2 M $MgSO_4$, 0.01% acetonitrile for 3-4 days. Crystals dehydrated by transfer to a 0.1 M MES pH 5.0, 1.5 M $MgSO_4$ solution for 2 hours at 18 ºC were immersed in mother liquor containing 14% ethylene glycol, mounted, and flash frozen in liquid N_2. Diffraction data were collected at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1 (32). Data reduction and initial maps were obtained using the automated ELVES

program (33). Phases were obtained experimentally with data obtained from selenomethionine-substituted ATR1Δ51. The PHENIX software suite was used for initial model building. The final model was built by iterative manual model building using Coot (34) and maximum likelihood refinement with PHENIX (35). Structure was validated using MOLProbity (36). Images and structural alignments were generated using PyMol (37) and Chimera (38). Multiple sequence alignment of the five ATR1 alleles was done using the MUSCLE algorithm (39) and visualized in CLC Genomics Workbench (www.clcbio.com). Structure comparisons were done using the DALI server (40) and Chimera (38) Coordinates and structure factors were deposited in the Protein Data Bank (PDB ID: 3RMR).

Hidden Markov Model-based sequence searches

Structure-based amino acid alignment between the overlapping region in ATR1 and Avr3a was derived using PyMol and used as an initial seed in HMM building. The HMM building, calibration, searches and subsequent alignments were performed using HMMER software package, using *hmmbuild*, *hmmcalibrate*, *hmmsearch* and *hmmalign* respectively (hmmer.janelia.org). The HMMs were iteratively scanned for three rounds against the Hpa proteome (e-value cutoff of 1). Sequences were analyzed for eukaryotic signal sequences using SignalP 3.0 (41). The W-motif domain sequence logo was made using WebLogo (42).

Functional analysis of ATR1 in *Nicotiana tabacum*

Deletions in ATR1Δ51 pENTRY/TOPO (16) were made through PCR amplification. The resulting products were introduced in pENTRY/TOPO (Invitrogen) and subsequently into pEG202 (35S promoter, N-terminal FLAG tag fusion) (43) using LR clonase (Invitrogen). Site-directed mutants were made in ATR1 pENTRY/TOPO using the Quick-Change SDM Kit (Stratagene) and subsequently introduced into pEG202. *Agrobacterium*-mediated transient expression in *N. tabacum* was performed as previously described (16). Protein expression was sampled at 24-48 hours post induction and assayed as described previously (16).

Measuring the effect of ATR1 on circadian clock regulation in *Arabidopsis*

Transgenic Arabidopsis containing P(TOC1):LUC reporter (kindly provided by Frank Harmon, PGEC) were germinated and grown at 8/16 light dark cycle for the first three weeks, and then transferred into 12/12 light dark cycle conditions to train the circadian clock for additional 2 weeks. The night before the experiment these plants were sprayed with 5 mM luciferin, 0.01% Triton X-100. Next day, around noon, the leaves were infiltrated with *P. fluorescens* containing Type III delivery system and ATR1D49-Emoy2 in pEDV3 or pEDV3 empty vector. An additional $MgCl_2$ (no bacteria) control was included. Three hours later the inoculated leaves were detached, placed on Murashige and Skoog agar plates supplemented with carbenicillin (50 mg/mL) and sprayed with 5 mM luciferin,

0.01% Triton X-100 to boost luciferase activity. Bioluminescence rhythms were measured starting at 4 pm (ZT6) under constant light conditions. The measurements were taken every 2 hours. The data was analyzed using the Night Owl imaging system and BRASS software. Experiments were repeated at least three times.

Collecting and analyzing the time-lapse data

The *Agrobacterium* inoculation time course data was collected using the Sofortbild program (http://www.sofortbildapp.com/) with 5 minute interval between the pictures. The frames were processed in iMovie version 9.0 (http://www.apple.com/ilife/imovie/). Hourly frames were analyzed using ImageJ v1.43 (Abramoff *et al.* 2004). Total leaf area showing visible hypersensitive response (HR) was outlined and measured for each ATR1 mutant co-expressed with RPP1-WsB. The HR-positive area relative to total inoculated space was measured and plotted as a function of time (hours post inoculation). *N. tabacum* leafs varied in timing of the first onset of HR. Slightly older leaves showed considerably slower response and more temporal separation between single, double and quadruple mutants, but the order of HR appearance of different inoculations within the same leaf was preserved.

2.7 Tables

Table 2.1. ATR1 X-ray data collection and refinement statistics.

Parentheses denote the highest resolution shell.

Crystal	**Native**	
Space group	$P6_122$	
Unit cell dimensions (Å)	a=119.421, b=119.421, c=312.796	
Resolution (Å)	50.0-2.30 (2.42-2.30)	
No. of measurements	57,939	
No. of unique reflections	7,685	
% Completeness	99.8 (97.6)	
I/σ	35.6 (4.4)	
R_{merge} (%) [a]	8.3 (7.3)	
Protein molecules in asymmetric unit	3	
Phasing (MAD):	**Fpeak**	**Fhigh**
Wavelength (Å)	1.0722	1.0631
Resolution (Å)	50.0-2.84	50.0-2.84
Rsym	0.209 (2.12)	0.205 (2.00)
Completeness (%)	100	100
Multiplicity	30.6 (30.3)	30.7 (3.1)
Mean I/σ	16.4 (2.1)	17.7 (2.3)
Mean Figure of merit [b] (102.7-2.8-Å resolution)	0.099 (0.761 after solvent flattening)	
Refinement statistics:		
Resolution	62.3-2.3 Å	
No. of reflections	55,573	
Rcryst/Rfree [c]	0.2231/0.2598	
No. of atoms:		
Protein	5,482	
Water	365	
Stereochemistry, RMS deviations [d]:		
Bond lengths	1.02 Å	
Bond angles	0.015 °	
Average B factor	47.2 $Å^2$	
Ramachandran plot:		
Favored (%)	95.3	
Allowed (%)	4.7	
Disallowed (%)	0	

a. $R_{merge} = \sum|I - \langle I\rangle| / \sum I$; I, intensity. b. Mean figure of merit (after density modification) = $\left\langle \left| \sum_{\alpha} P(\alpha) e^{ia} / \sum_{\alpha} P(\alpha) \right| \right\rangle$; a, phase; P(a), phase probability distribution. c. $R_{cryst} = \sum |F_o - F_{calc}| / \sum F_o$; F_o, observed structure-factor amplitude; F_{calc}, calculated structure-factor amplitude. d. Root-mean-square deviations from ideal values.

Table 2.2. Hidden-Markov-Model-based search for *Hpa* genes that contain a modified W-motif.

Gene ID	Score	E-value	Number of domains	Features of effectors
801846	131.7	3.40E-36	1	SP (prob = 0.993), dEER
813261	118.7	2.70E-32	2	
809859	117.4	7.00E-32	1	dEER
801867<-ATR1	112.1	2.60E-30	1	SP (prob = 0.988), RXLR, dEER
802236	111.3	4.70E-30	3	SP (prob = 0.802), dEER
811478	95.3	3.10E-25	1	SP (prob = 0.949), dEER
812377	92.2	2.70E-24	1	
814615	75.2	3.40E-19	1	
808367	48.9	2.80E-11	2	
814280	45	4.30E-10	1	
800198	17.7	0.008	3	RXLR
808581	15.5	0.014	2	
813431	8.9	0.089	1	
811521	7	0.15	1	
806967	5.7	0.21	1	
807782	5	0.26	1	
807781	5	0.26	1	
810794	4.1	0.33	1	
808368	3.5	0.39	1	
802347	3.1	0.44	1	SP (prob = 0.830)
812044	2.7	0.49	1	SP (prob = 0.716), dEER
811880	2	0.59	1	SP (prob = 1.000)
811884	1.8	0.63	1	
810698	0.6	0.86	1	SP (prob = 1.000), dEER
808349	0.2	0.97	1	
801471	0.2	0.97	1	SP (prob = 0.760)
Total 25				

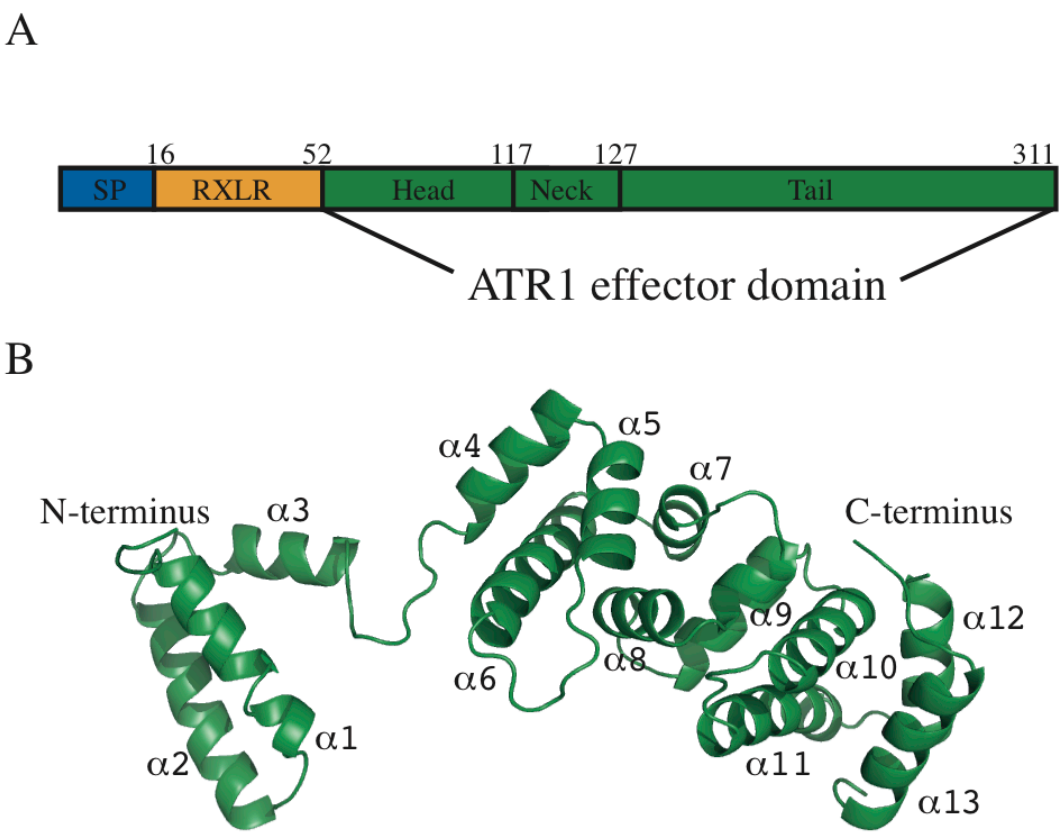
Table 2.3. Comparison of ATR1 alleles links amino acid identities to recognition by RPP1 alleles.

[a] Percent sequence identity to ATR1-Emoy2. [b] Number of polymorphism relative to ATR1-Emoy2, including gaps. [c] Recognition by RPP1 alleles NdA and WsB. + recognized, - not recognized.

	% identity [a]	Differences [b]	NdA [c]	WsB [c]
Emoy2	100	0	+	+
Maks9	98.39	5	-	+
Emco5	85.03	47	-	+
Cala2	78.70	69	-	-
Emwa1	78.15	71	-	-

2.9 Figures

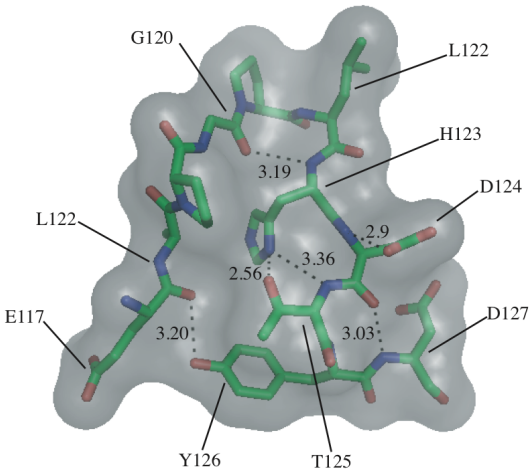
Figure 2.1. The structure of the effector segment of ATR1.



A) A schematic representation of the domain architecture of ATR1.

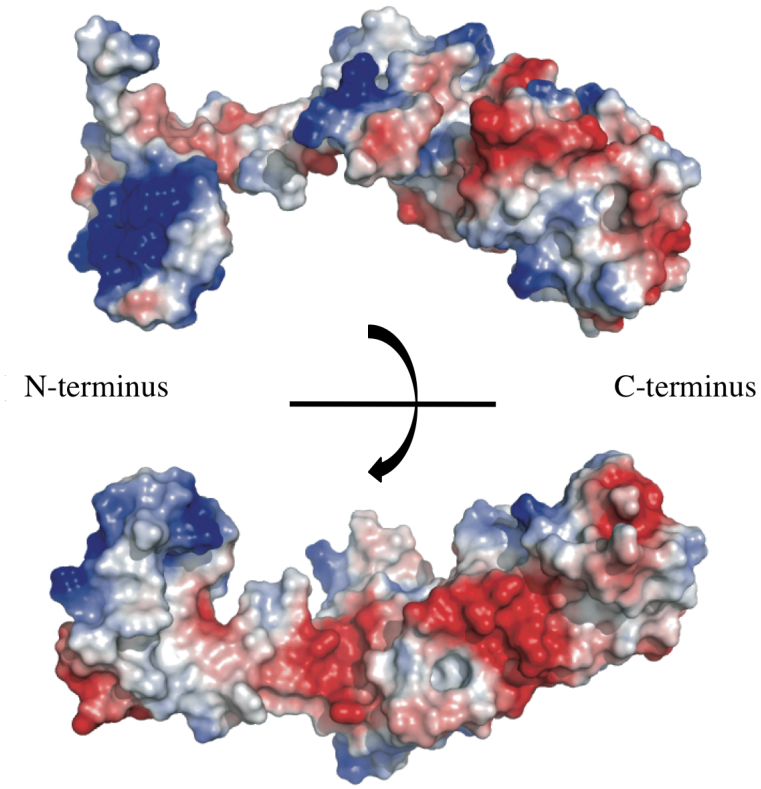
B) Ribbon diagram showing the overall structure of the ATR1D51. Thirteen a-helices form N-terminal (α1-α3) and C-terminal (α4-α13) domains. The elongated structure is over 80 Å across, significantly more extended than a globular protein of the same molecular mass.

Figure 2.2. ATR1 neck region is stabilized by hydrogen bonds.



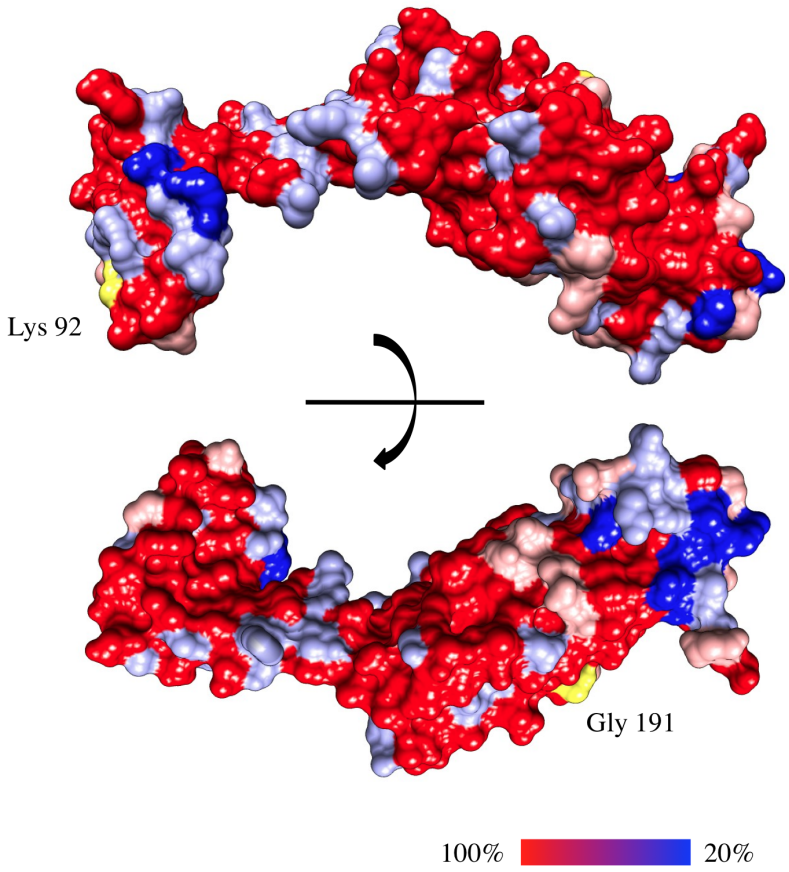
Stick representation superimposed on the transparent solvent-accessible surface.

Figure 2.3. ATR1 surface has numerous electropositive and electronegative patches.



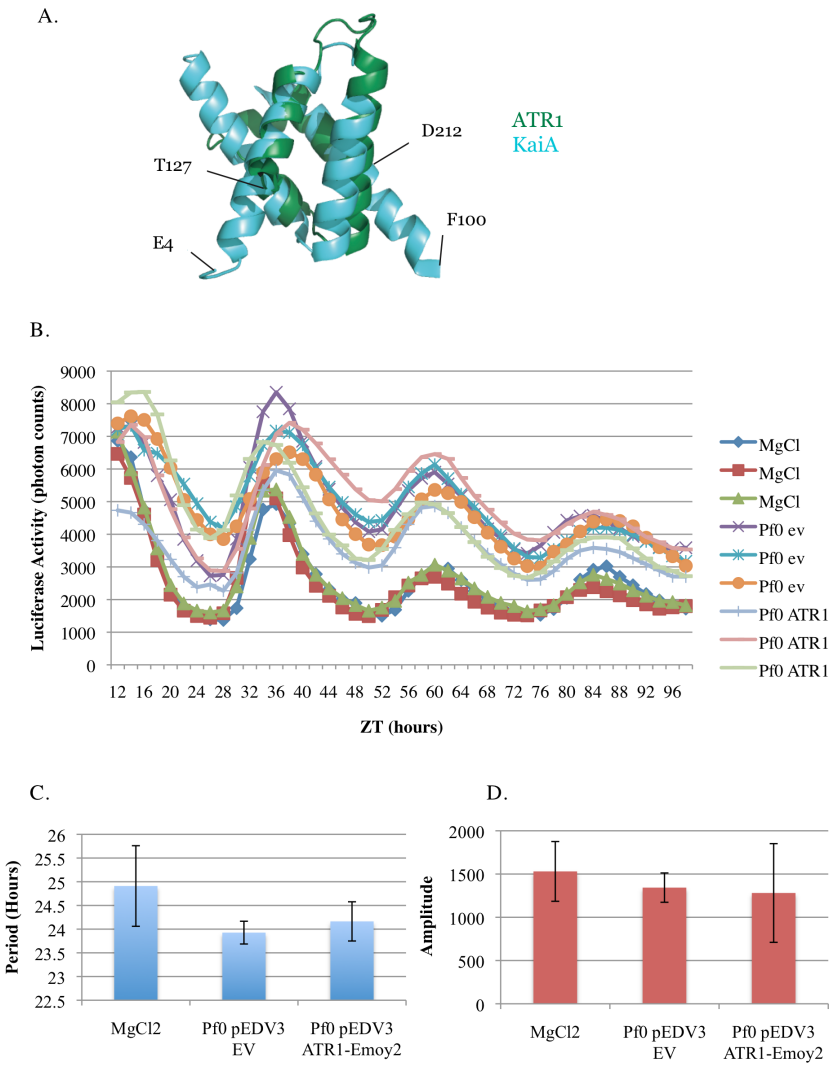
A surface representation of ATR1 showing regions of calculated positive (blue) and negative (red) electrostatic potential.

Figure 2.4. ATR1 polymorphic residues are distributed across the surface of both domains.



Amino acid conservation between five ATR1 alleles (Emoy2, Maks9, Emco5, Cala2, and Emwa1) ranging from 100% conserved (red) to 20% conserved (blue) mapped on the structure of ATR1-Emoy2 using Chimera. In yellow, two sites polymorphic between ATR1-Emoy2 and ATR1-Maks9 that were previously implicated in recognition of ATR1 by RPP1-NdA.

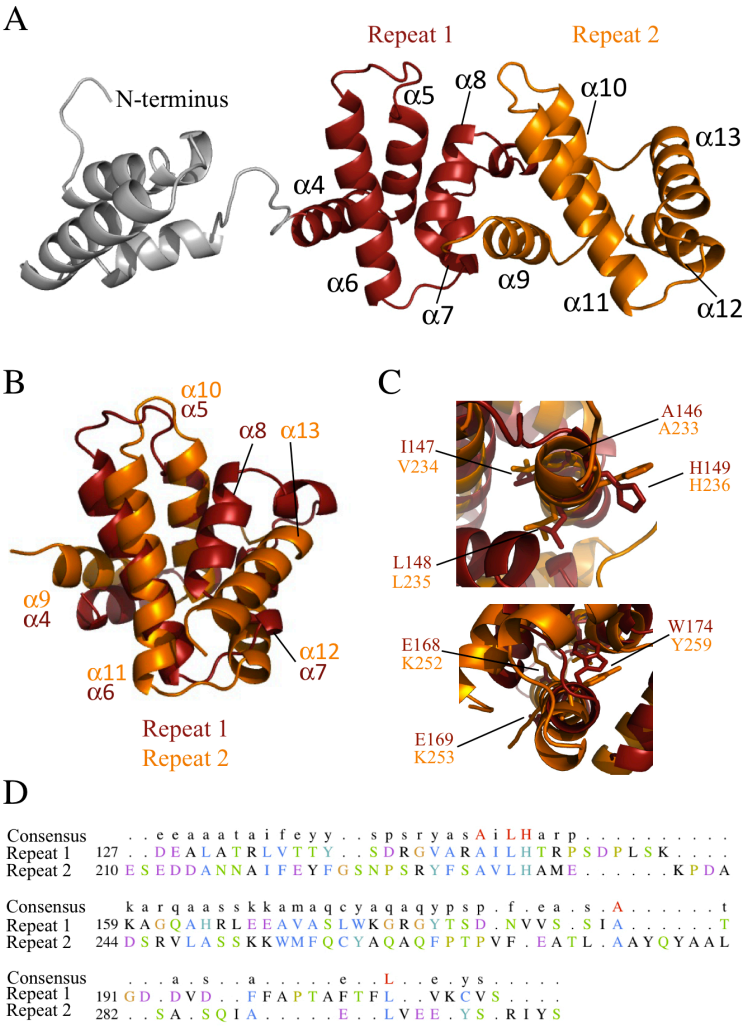
Figure 2.5. ATR1 has a structural but not functional similarity to cyanobacterial circadian regulator, KaiA.



A) ATR1 α5-α9 subdomain has structural similarity to the circadian regulator, KaiA. Ribbon diagram of a structural alignment of ATR1 (green) and KaiA (cyan) generated using Pymol.

B) ATR1 does not affect the transcriptional regulation of the circadian clock in *Arabidopsis*. The transcriptional output was measured using Arabidopsis Col-0 containing a P_{TOC1}:LUC transgene. ATR1-Emoy2 was delivered using the *Pseudomonas fluorescens* Type III delivery system and compared directly to an empty-vector control and a $MgCl_2$ control (no bacteria). *(B)* Relative levels of luciferase activity measured over the period of 96 hours post infiltration. *(C)* Period and *(D)* Amplitude derived from these data.

Figure 2.6. A structural repeat in ATR1.



A) The C-terminal body domain of ATR1 contains two 5-helix repeats with similar folds (orange and red). The ATR1 repeats are related primarily by a translation and a small rotation.

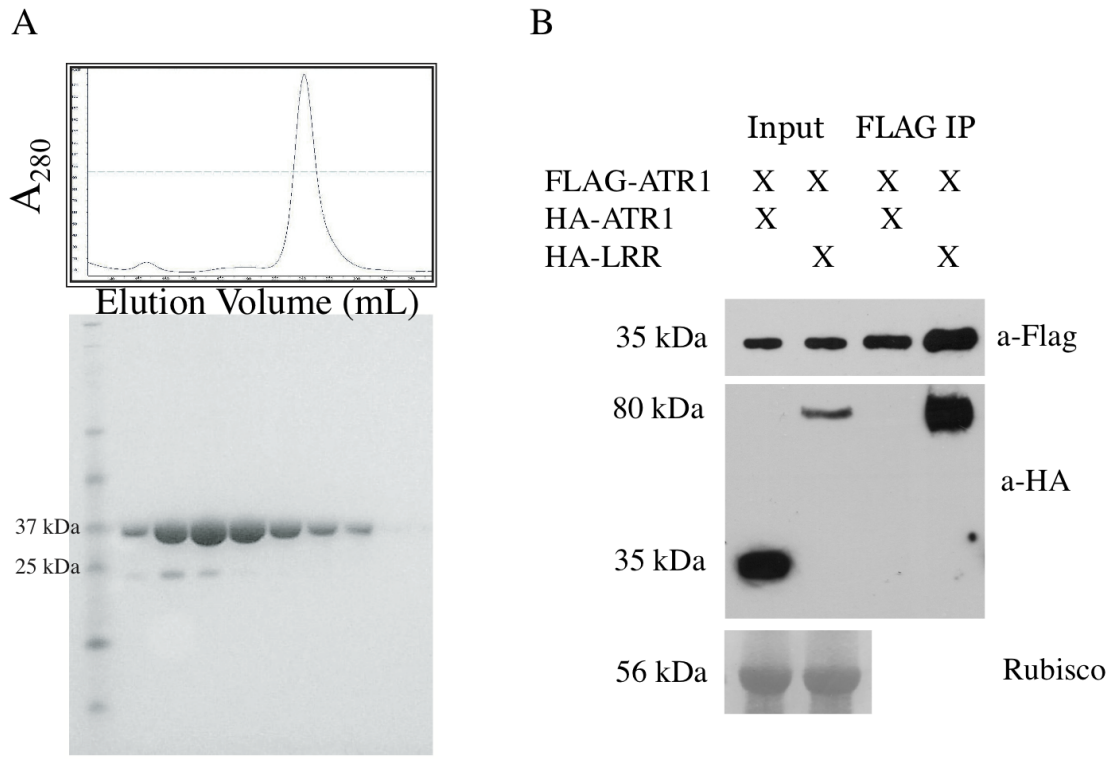
B) Structural alignment based on the second helix in each repeat (α5 and α10 aligned using Pymol) shows the similar helix arrangement with variable connecting loops.

C) The structural alignment reveals conserved residues in helix 2 (top) and similar residues in helix 3 (bottom) of the repeats.

‘

D) Structure-based sequence alignment between the ATR1 repeats shows low sequence identity.

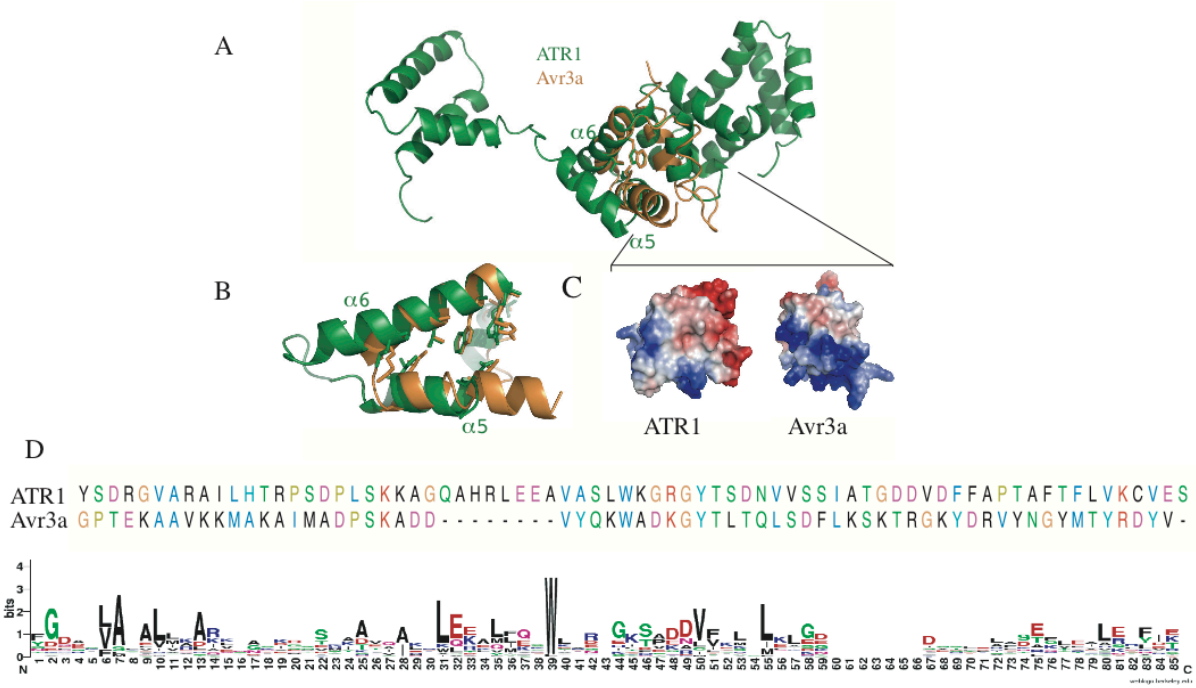
Figure 2.7. ATR1 is a monomer *in vitro* and *in vivo*.



A) Recombinant ATR1 elutes from a Superdex S75 gel filtration column at volume corresponding to a monomer.

B) Co-immunoprecipitation experiment showing that FLAG-ATR1 associates with HA-LRR of RPP1 (lane 4) but not HA-ATR1 (lane 3) in *N. tabacum*. These results suggest that ATR1 is a monomer *in vivo.*

Figure 2.8. Structural alignment of ATR1 and Avr3a reveals conserved core.



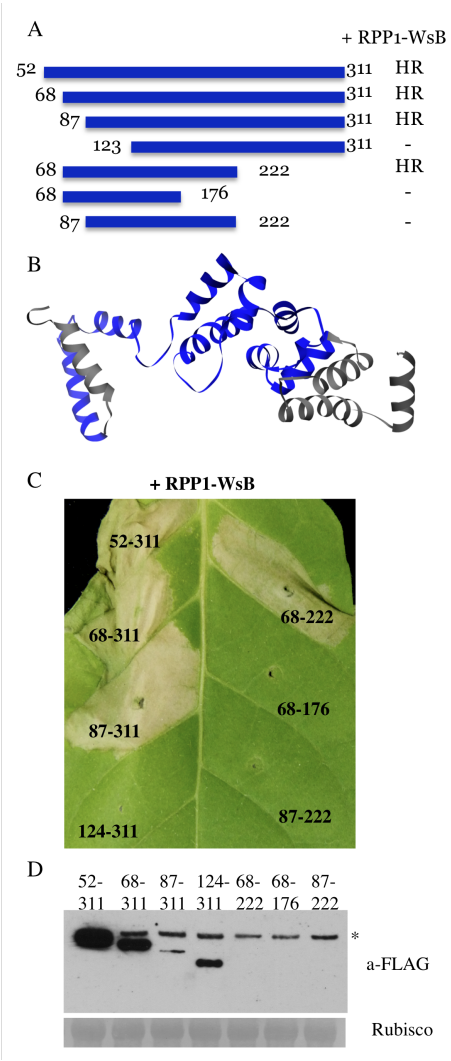
A) ATR1 (green) and Avr3a (orange) share a common 4-helix domain in the body of ATR1. Sequence alignment of this region highlights conserved amino acids.

B) Close-up view of the conserved region. Structural-based alignment identifies several conserved buried residues in the core.

C) Electrostatic view of ATR1 and Avr3a shows diverged surface features.

D) *Top* - Amino acid alignment between ATR1 and Avr3a11 derived from the structural comparison. Alignment is colored based on Clustal X coloring scheme. *Bottom* - Consensus sequence logo from the amino acid alignment of the W-motif in 25 genes identified from *Hpa.*

Figure 2.9. *In planta* deletion analysis of ATR1-Emoy2.



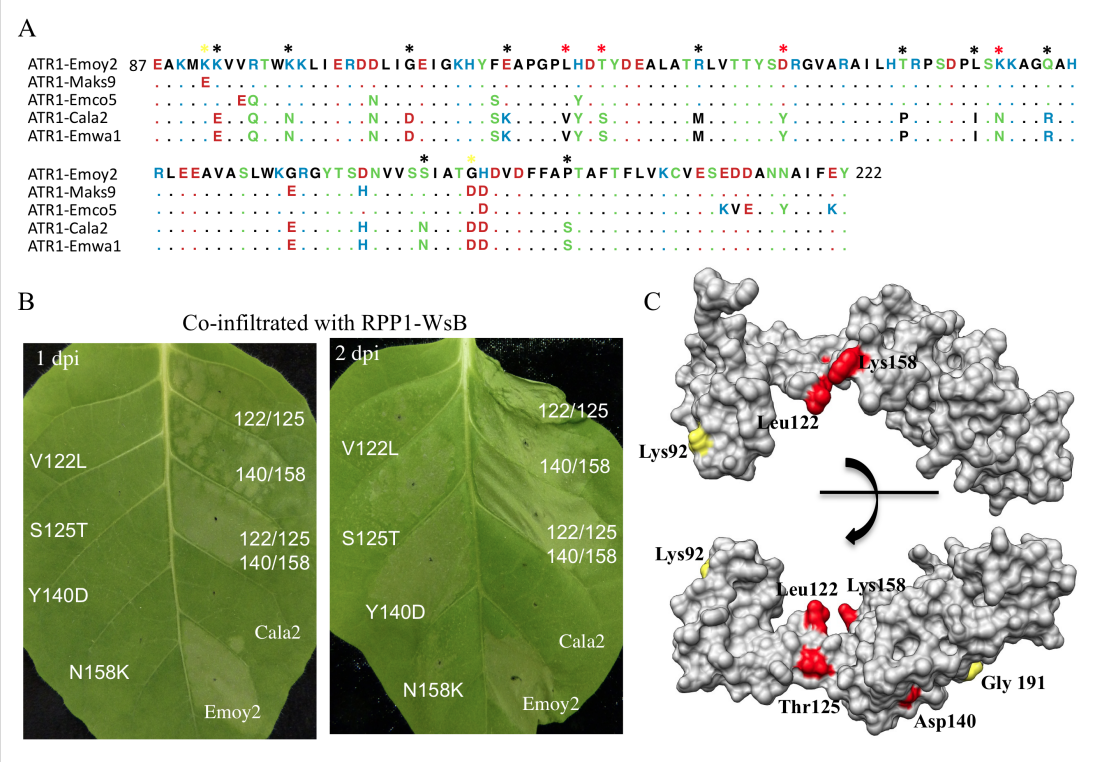
A) Schematic diagram of deletion constructs and a summary of the hypersensitive response (HR) phenotype.

B) Ribbon diagram showing the minimum recognition region (blue) of ATR1 containing the groove surrounding the neck and the first half of the C-terminal body domain.

C) Co-inoculation of ATR1 truncations together with RPP1-WsB in *N. tabacum* showing the induction of RPP1-dependent HR. Fragments starting at residue 87 or ending at residue 222 are sufficient to induce the HR.

D) Western blot showing relative protein levels of the truncated ATR1 variants. Asterisk marks a non-specific ~33-kDa band cross-hybridizing with anti-FLAG antibody. The C-terminal domain (residues 124-311) expresses and accumulates, but it does not trigger the HR. The 87-222 fragment encompassing the minimal recognition region is unstable.

Figure 2.10. Distributed surfaces contribute to RPP1 activation by ATR1.

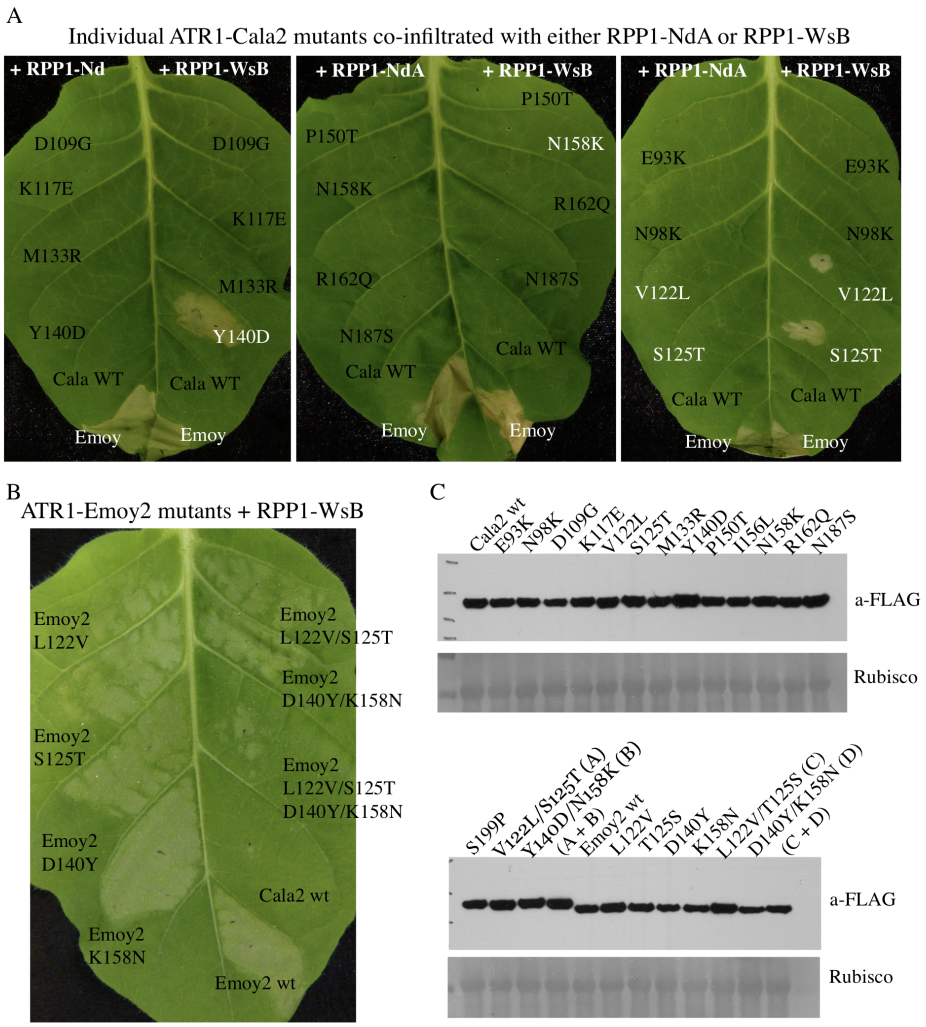


A) Multiple sequence alignment of amino acids 87-222 of five ATR1 alleles. Previously identified amino acids critical for ATR1-Emoy2 recognition by RPP1-NdA (16) are marked with orange asterisks. Residues that co-segregate with recognition of ATR1 by RPP1-WsB are shown with black asterisks. The four positions identified as individually critical for switching the specificity of HR activation are marked with red asterisks.

B) Additive effect of the four amino acids that contribute to recognition of ATR1 by RPP1-WsB as assayed by timing and intensity of HR in *N. tabacum*. The variants include: 122, ATR1-Cala2 Val122Leu; 125, ATR1-Cala2 Ser125Thr, 140, ATR1-Cala2 Tyr140Asp; 158, ATR1-Cala2 Asn158Lys, and combinations of these substitutions, as well as a wild-type ATR1-Cala2 (negative control) and wild-type ATR1 Emoy2 (positive control).

C) Surface representation of ATR1 with mapped polymorphisms, highlighting previously identified residues that were critical for recognition of ATR1 by RPP1-NdA (yellow) and our data revealing that recognition by RPP1-WsB relies on different critical sites (red).

Figure 2.11. Site-directed mutagenesis of ATR1.

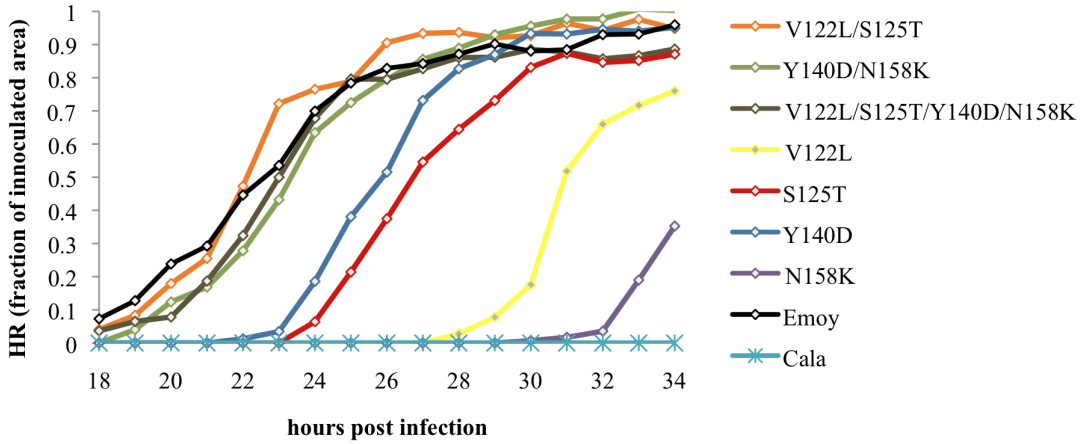


A) Phenotypes of the individual mutations introduced to ATR1 Cala2 and their relative contributions to recognition by RPP1-NdA (left side of the leaf) and RPP1-WsB (right side of the leaf). Picture taken at 3 days post infiltration.

B) Single and combined mutations in ATR1-Emoy2 showing their delayed effect on RPP1-WsB. Picture taken at 24 hours post infiltration. At 2-3 days post infiltration, double and quadruple mutants start to promote HR, indicating functionality of additional amino acid sites.

C) Relative protein expression levels of the ATR1 mutants used in this study. Rubisco – loading control.

Figure 2.12. Combining mutations in ATR1 accelerate RPP1 activation.



The graph shows the timing and intensity of RPP1-dependent HR in *Nicotiana tabacum* induced by different ATR1-Cala2 mutations. The data was collected as time-lapse images. Hourly frames were analyzed in ImageJ to derive leaf areas. The fraction of area undergoing HR relative to total infiltrated area is plotted against the hours post infiltration.

2.10 References

1. Goker, M., Voglmayr, H., Riethmuller, A., and Oberwinkler, F. (2007) *Fungal Genetics and Biology* **44**, 105-122
2. Dodds, P. N., Rafiqi, M., Gan, P. H. P., Hardham, A. R., Jones, D. A., and Ellis, J. G. (2009) *New Phytol* **183**, 993-1000
3. Bhattacharjee, S., Hiller, N. L., Liolios, K., Win, J., Kanneganti, T. D., Young, C., Kamoun, S., and Haldar, K. (2006) *PLoS Pathog* **2**, e50
4. Dou, D., Kale, S. D., Wang, X., Chen, Y., Wang, Q., Wang, X., Jiang, R. H. Y., Arredondo, F. D., Anderson, R. G., Thakur, P. B., McDowell, J. M., Wang, Y., and Tyler, B. M. (2008) *Plant Cell* **20**, 1118-1133
5. Grouffaud, S., van West, P., Avrova, A. O., Birch, P. R. J., and Whisson, S. C. (2008) Plasmodium falciparum and Hyaloperonospora parasitica effector translocation motifs are functional in Phytophthora infestans. in *Microbiology (Reading, Engl)*
6. Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., Van West, P., Chapman, S., Hein, I., Toth, I. K., Pritchard, L., and Birch, P. R. J. (2007) *Nature* **450**, 115-118
7. Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., and Kamoun, S. (2007) *Plant Cell* **19**, 2349-2369
8. Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J. L., Chibucos, M. C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S. I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R. H. Y., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S. W., Rogers, J., Van den Ackerveken, G., Jones, J. D. G., McDowell, J. M., Beynon, J., and Tyler, B. M. (2010) *Science* **330**, 1549-1551
9. Jiang, R. H. Y., Tripathy, S., Govers, F., and Tyler, B. M. (2008) RXLR effector reservoir in two Phytophthora species is dominated by a single rapidly evolving superfamily with more than 700 members. in *Proc Natl Acad Sci USA*
10. Rehmany, A. P. (2005) *THE PLANT CELL ONLINE* **17**, 1839-1850
11. Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P., and Valent, B. (2000) *EMBO (European Molecular Biology Organization) Journal* **19**, 4004-4014
12. Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J. E., and Innes, R. W. (2003) *Science* **301**, 1230-1233
13. Andersson, M. X., Kourtchenko, O., Dangl, J. L., Mackey, D., and Ellerstrom, M. (2006) *Plant J* **47**, 947-959

14. Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006) *Cell* **124**, 803-814
15. Jones, J. D. G., and Dangl, J. L. (2006) *Nature* **444**, 323-329
16. Krasileva, K. V., Dahlbeck, D., and Staskawicz, B. J. (2010) *Plant Cell*
17. Botella, M. A., Parker, J. E., Frost, L. N., Bittner-Eddy, P. D., Beynon, J. L., Daniels, M. J., Holub, E. B., and Jones, J. D. (1998) *Plant Cell* **10**, 1847-1860
18. Rehmany, A. P., Gordon, A., Rose, L. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun, S., Tyler, B. M., Birch, P. R., and Beynon, J. L. (2005) *Plant Cell* **17**, 1839-1850
19. Rehmany, A. P., Gordon, A., Rose, L. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun, S., Tyler, B. M., Birch, P. R. J., and Beynon, J. L. (2005) *Plant Cell* **17**, 1839-1850
20. Jiang, R. H. Y., Tripathy, S., Govers, F., and Tyler, B. M. (2008) *Proc Natl Acad Sci USA* **105**, 4874-4879
21. Wang, C.-I. A., Guncar, G., Forwood, J. K., Teh, T., Catanzariti, A.-M., Lawrence, G. J., Loughlin, F. E., Mackay, J. P., Schirra, H. J., Anderson, P. A., Ellis, J. G., Dodds, P. N., and Kobe, B. (2007) *Plant Cell* **19**, 2898-2912
22. Dong, J., Xiao, F., Fan, F., Gu, L., Cang, H., Martin, G. B., and Chai, J. (2009) *Plant Cell* **21**, 1846-1859
23. Lee, C. C., Wood, M. D., Ng, K., Andersen, C. B., Liu, Y., Luginbuhl, P., Spraggon, G., and Katagiri, F. (2004) *Structure* **12**, 487-494
24. Gardino, A. K., Smerdon, S. J., and Yaffe, M. B. (2006) *Seminars in cancer biology* **16**, 173-182
25. Ybe, J. A., Brodsky, F. M., Hofmann, K., Lin, K., Liu, S. H., Chen, L., Earnest, T. N., Fletterick, R. J., and Hwang, P. K. (1999) *Nature* **399**, 371-375
26. Bos, J. I. B., Chaparro-Garcia, A., Quesada-Ocampo, L. M., McSpadden Gardener, B. B., and Kamoun, S. (2009) *Mol Plant Microbe Interact* **22**, 269-281
27. Williams, S. B., Vakonakis, I., Golden, S. S., and LiWang, A. C. (2002) Structure and function from the circadian clock protein KaiA of Synechococcus elongatus: a potential clock input mechanism. in *Proc Natl Acad Sci USA*
28. Wang, W., Barnaby, J. Y., Tada, Y., Li, H., Tor, M., Caldelari, D., Lee, D. U., Fu, X. D., and Dong, X. (2011) *Nature* **470**, 110-114
29. Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J. G., Kobe, B., and Dodds, P. N. (2011) *Cell Host Microbe* **9**, 200-211
30. Catanzariti, A.-M., Dodds, P. N., Ve, T., Kobe, B., Ellis, J. G., and Staskawicz, B. J. (2010) *Molecular Plant-Microbe Interactions* **23**, 49-57
31. Wise, A. A., Liu, Z., and Binns, A. N. (2006) *Methods Mol Biol* **343**, 43-53
32. MacDowell, A. A., Celestre, R. S., Howells, M., McKinney, W., Krupnick, J., Cambie, D., Domning, E. E., Duarte, R. M., Kelez, N., Plate, D. W., Cork, C. W., Earnest, T. N., Dickert, J., Meigs, G., Ralston, C., Holton, J.

M., Alber, T., Berger, J. M., Agard, D. A., and Padmore, H. A. (2004) *Journal of synchrotron radiation* **11**, 447-455
33. Holton, J., and Alber, T. (2004) *Proc Natl Acad Sci U S A* **101**, 1537-1542
34. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
35. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
36. Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) *Nucleic Acids Res* **32**, W615-619
37. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng* **11**, 739-747
38. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J Comput Chem* **25**, 1605-1612
39. Edgar, R. (2004) *Nucleic Acids Research*
40. Holm, L., and Rosenstrom, P. (2010) *Nucleic Acids Res* **38**, W545-549
41. Bendtsen, J. D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004) *J Mol Biol* **340**, 783-795
42. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) *Genome Res* **14**, 1188-1190
43. Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C. S. (2006) *The Plant Journal* **45**, 616-629

Chapter 3

HIV-1 Accessory Protein Tat Recruits the Super Elongation Complex to Stimulate Viral Transcription

3.1 Abstract

HIV Tat stimulates viral transcription elongation by recruiting the human P-TEFb kinase to the HIV promoter to phosphorylate RNA polymerase II and the negative elongation factors NELF and DSIF. To define the composition of elongation factor assemblies, we purified both active and inactive P-TEFb from Hela cells and analyzed these complexes by mass spectrometry. Importantly, we showed the stoichiometric presence of AFF4, ELL2, and homologs ENL and AF9 in the active Tat/P-TEFb complex that are important for viral transcription.. Here, we biochemically define these new interactions *in vivo* and reconstitute the complex *in vitro* using purified recombinant subunits. Our results show that these proteins form a larger Tat/P-TEFb complex, known as the Super Elongation Complex (SEC). Although HIV Tat enhances SEC recruitment and activity at the HIV promoter, Tat/P-TEFb requires AFF4 for assembly of the complex. AFF4 is the central SEC scaffold and contains several discrete interaction domains that independently and directly bind P-TEFb, ELL2, and ENL/AF9.

3.2 Introduction

A virus knows better than anyone that life is better with friends. Throughout its life cycle, HIV relies on the host for production of virions. Viral-host complexes play essential roles by not only remodeling host cell physiology, but also by co-opting host machinery for HIV growth and replication. For example, the HIV accessory protein, Transactivator of Transcription (Tat), stimulates viral transcription by recruiting the general human elongation factor, P-TEFb, specifically to the HIV promoter. Tat makes direct interactions with the cyclin T1 (CycT1) subunit of P-TEFb and with the viral RNA hairpin, TAR (1,2). Disruption of these interactions results in abortive HIV transcripts, highlighting the importance of Tat and its host co-factor, P-TEFb, for HIV survival (1,2). Clearly, the virus and host are intimately linked at the molecular level. Understanding the relationships at this interface offers insights into essential processes for both organisms.

A major substrate of P-TEFb is the carboxy-terminal domain (CTD) of the Pol II subunit, RPB1. During transcription, the Pol II CTD acts as a scaffold for the assembly of transcriptional machinery and regulators, and the dynamic phosphorylation state of the CTD heptapeptide repeat, YSPTSPS, is important for facilitating proper interactions at different stages of the transcription cycle (3,4). Following initiation, the CTD is phosphorylated at the Ser5 position of the heptapeptide repeat, and this mark is required for Pol II to clear the promoter and move into early elongation. Pol II is soon arrested by two negative factors, 5,6-dichloro-1-B-D-ribofuranosylbenzimidazole sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) (5,6), and elongation aborts absent any further stimulation. P-TEFb can phosphorylate these two factors to antagonize their inhibitory effects and further stimulate Pol II elongation by phosphorylating the Ser2 position of the Pol II CTD repeats (1,2,7). These concerted actions allow the continuation of elongation and are essential for transcription of the long HIV genome, which is hypersensitive to elongation defects.

In addition to facilitating P-TEFb recruitment to the HIV promoter, Tat also regulates the cellular levels of active P-TEFb. In the cell, more than half of the P-TEFb population is sequestered into a large, inactive complex consisting of the 7SK snRNA and several additional regulatory proteins, collectively known as the 7SK snRNP(8). The remaining P-TEFb is bound by a ubiquitously expressed, bromodomain-containing protein known as Brd4 that can recruit P-TEFb to paused Pol II molecules at endogenous promoters through its own interactions with acetylated histones – a hallmark of active transcription (1,9-11). The equilibrium of these two opposing populations of P-TEFb, active and inactive, is carefully regulated in the cell and rapidly adjusts according to the needs of the cell. Upon HIV-infection or HIV-independent Tat-transfection, an increase in active P-TEFb relative to inactive P-TEFb is observed. This shift is due to both the Tat-stimulated release of P-TEFb from the 7SK snRNP and the competition between Tat and Brd4 for P-TEFb recruitment (12).

Tat-dependent release of P-TEFb from both its positive and negative regulators is a poorly understood aspect of Tat function. It is unknown whether the observed decrease in Brd4- and HEXIM1-bound complexes and the increase in Tat-bound complexes are born out of a direct or indirect competition between the regulators. Tat and Brd4 bind the same region of CycT1, and in vitro competition assays suggest that Tat can directly compete with Brd4 for P-TEFb-binding (1,2). Moreover, the recent high-resolution crystal structure of Tat-bound P-TEFb shows Tat folded tightly around the CycT1 surface (13), strongly suggesting that its presence may exclude binding of other factors. Sequence and structural similarities between the RNA-binding domains of Tat and HEXIM1, as well as the 7SK and TAR RNA stem-loop regions, suggest that Tat-TAR may also directly compete with HEXIM1-7SK for binding to CycT1 (1). Tat can inhibit assembly of the 7SK snRNP and release P-TEFb from the cellular 7SK snRNP (14). Although these data raise the intriguing possibility of direct P-TEFb capture from endogenous regulators by Tat, further structural comparisons of Brd4/P-TEFb and the 7SKsnRNP with Tat/P-TEFb are needed to address this hypothesis.

Adding to the mystery of P-TEFb function and regulation is the incomplete biochemical characterization of inactive and active P-TEFb complexes. Although significant progress in the field had been made at the start of this project with respect to identifying key subunits of these complexes, it was unclear whether additional factors exist in P-TEFb complexes. Moreover, it was unknown whether there could be overlapping components of the inactive and active complexes, which could play a role in regulating the switch between these states. Also, studies showing crosstalk between transcriptional elongation, splicing and termination strongly suggested the possibility of physical interactions between regulators of these different processes (15,16). Thus, characterizing the interaction landscape that P-TEFb complexes function in could not only expand our understanding of viral transcriptional requirements but also mechanisms of eukaryotic transcriptional regulation, as well.

In this chapter, I describe our search for novel interactors of the P-TEFb complex using a proteomics-based approach. Following double immunoprecipitations to purify both inactive and Tat-bound P-TEFb from Hela cells, we used mass spectrometry to identify candidate interactors of both 7SKsnRNP and Tat/P-TEFb. 7SKsnRNP associated with several RNA splicing factors and serine-threonine kinases. Tat/P-TEFb associated with several host elongation factors, including AFF4, ELL2, ENL, and AF9. We validated candidate interactors *in vivo* and reconstituted the interactions *in vitro* to demonstrate that these exist as part of a single Tat/P-TEfb complex, now known as the Super Elongation Complex (SEC). Although Tat stimulates SEC recruitment to the HIV LTR, our data reveal that Tat also regulates SEC indirectly by modulating stability of individual components. AFF4, but not HIV Tat, scaffolds the assembly of this new Tat/P-TEFb complex by directly binding other components. These data establish a new framework for understanding the mechanisms by which Tat stimulates viral transcription.

3.3 HIV-1 Tat / Human P-TEFb complex interacts with elongation factors

Despite significant advances in understanding the role P-TEFb plays in Tat-dependent transactivation of HIV transcription, many questions remain about the biochemical and functional differences between active and inactive P-TEFb (7SKsnRNP). To identify additional regulators that function specifically with active Tat/P-TEFb, we used a proteomics-based approach to identify host proteins that physically interact with the complex. We isolated P-TEFb complexes from HeLa cells and analyzed the composition of both 7SKsnRNP and active Tat/P-TEFb assemblies using mass spectrometry (MS). To isolate the complexes, we used a sequential double immunoprecipitation (IP) purification scheme to enrich for specific P-TEFb complexes (Figure 3.1). To purify native 7SKsnRNP, we tagged the P-TEFb subunit, CDK9, with an HA-tag at the C-terminus and the 7SKsnRNP-specific component, HEXIM1, with a Flag-tag at the N-terminus (HEXIM-F). Nuclear extract was prepared from cells and an anti-FLAG IP was followed by an anti-HA IP to purify HEXIM1-bound P-TEFb complexes, enriching for 7SKsnRNP (Figure 3.2A). We similarly isolated Tat-bound P-TEFb complexes with sequential affinity purifications from an engineered human cell line inducibly expressing Tat-HA and stably expressing CDK9-Flag (CDK9-F).

MS analysis of purified complexes detected all known components (Figures 3.2A- B, Tables 3.1-3.2). The 7SKsnRNP complexes contained CDK9, CycT1, HEXIM1, PIP7S, and BCDIN3 as well as several additional candidate components, including several host kinases and RNA splicing factors, that may have direct or indirect physical interactions with the complex (Figure 3.2A, Table 3.1). Analysis of Tat/P-TEFb complexes by MS verified the presence of HIV Tat, CDK9, and CycT1 (Figure 3.2B, Table 3.2). Interestingly, visualization of purified P-TEFb by silver-stained SDS-PAGE showed several additional proteins that co-purified with P-TEFb (Figure 3.2B). MS revealed that these candidate interactors, AFF4, ELL2, ENL, and AF9, uniquely associated with Tat/P-TEFb and not 7SKsnRNP (Tables 3.1-3.2). Together, these results raise the possibility that Tat/P-TEFb may specifically interact with four additional host factors.

AFF4, ELL2, ENL, and AF9 belong to a family of host elongation factors that are fusion partners of mixed lineage leukemia (MLL) proteins in acute myeloid leukemia (Figure 3.2D) (17-20). ELL2, a member of the ELL family, increases the catalytic rate of Pol II and suppresses transient Pol II pausing during transcriptional elongation (21-23). AFF4 is an elongation factor in the AF4 family, and ENL and AF9 are homologs containing the chromatin-associated YEATS domain. AFF4, ENL, and AF9 have previously been shown to associate with P-TEFb and regulate elongation (20,24), although their roles in HIV transcription remain unclear. Moreover, during the time of this study, it was unknown whether these factors exist together in a single complex or in several, distinct complexes.

3.4 Four host elongation factors are part of the Tat/P-TEFb complex.

Since the initial identification of AFF4, ELL2, ENL, and AF9 as potential partners of Tat/P-TEFb, studies from our group and others have expanded our biochemical and functional understanding of the larger complex HIV Tat recruits to stimulate viral transcription (25-28). HIV Tat binds P- TEFb in a larger complex with AFF4, ELL2, and homologs ENL and AF9, now known as the Super Elongation Complex (SEC). Importantly, these new components synergize with Tat to activate HIV transcription by stimulating transcriptional elongation (25,27). Interestingly, although Tat is not required for P-TEFb to interact with the additional factors, the presence of Tat enhances association their association with P-TEFb (25,27), which may explain why these components have eluded identification in the past. Moreover, He et al. further defined the mechanism by which Tat stimulates recruitment of one factor, ELL2, to the P-TEFb complex by showing that HIV Tat significantly elevates overall ELL2 protein levels in the cell by increasing its stability (25). AFF4 similarly enhances ELL2 protein stability in the cell to promote ELL2 association with the complex (25) and is required for SEC assembly (25,26). Although many questions remain regarding the precise mechanisms by which these factors affect both viral and endogenous transcription, these studies provide initial insights into composition and assembly of the SEC.

Recently, mapping studies showed that AFF4 associates with P-TEFb, ELL2, ENL, and AF9 *in vivo* through discrete, modular interaction domains. Specifically, we tested a series of Flag-tagged AFF4 truncations for their ability to co-precipitate (co-IP) with SEC components (Figure 3.3). The deletion mutant lacking the first 300 residues (Δ300) no longer associated with P-TEFb, although all other components were bound (lane 3). Moreover, a C-terminal truncation containing only the first 300 residues ($AFF4_{1-300}$) was sufficient to co-IP only P-TEFb (lane 6), suggesting that the N-terminal 300 amino acids independently mediates the interaction between AFF4 and P-TEFb. The region of AFF4 between amino acids 301 to 600 seemed to be an ELL2 interaction domain. AFF4 missing the first 600 residues (Δ600) no longer co-precipitated with ELL2 (lane 4), while the AFF4 truncation containing residues 301 to 600 ($AFF4_{301-600}$) was sufficient for this interaction (lane 7). Similarly, residues 601-900 seemed to constitute an interaction domain for ENL and AF9 (lanes 5 and 8). We conclude that AFF4 assembles the SEC *in vivo* by interacting with other subunits via discrete, independent domains (Figure 3.4).

3.5 AFF4 assembles the SEC by directly binding complex members

To determine whether AFF4 interaction domains nucleate SEC assembly by directly binding P-TEFb, ELL2, and ENL/AF9, we expressed and purified recombinant His-tagged AFF4 truncations in *E. coli* and tested their ability to bind the other factors. Due to poor expression and solubility of full-length SEC components in *E. coli*, we focused our efforts on deletion mutants encompassing

regions previously shown to be both necessary and sufficient for interactions in vivo (25-27). Specifically, *in vivo* IPs show that the N-terminus of CycT1 mediates the interaction between P-TEFb and the first 300 residues of AFF4, so we asked whether His-CycT1$_{1-303}$ could form a complex with His-AFF4$_{1-300}$. The two proteins were incubated at equimolar concentrations for 15 minutes at 4 °C in Buffer 0.4 (0.4 M NaCl, 20 mM HEPES, pH 8.0, 0.5 mM TCEP, 10% glycerol) and analyzed the reaction by gel exclusion chromatography. The proteins eluted together at a volume corresponding to 81.0 kDa, while His-CycT1$_{1-303}$ alone eluted at a volume corresponding to 40.6 kDa (Table 3, Figure 5A). Coomassie stained SDS-PAGE of the eluted binding reaction shows the presence of both His-AFF4$_{1-300}$ and His-CycT1$_{1-303}$ (Figure 3.5B), confirming that the two proteins form a larger complex. Thus, we conclude that AFF4$_{1-300}$ directly binds the N-terminus of CycT1.

We also asked whether the next 300 amino acids of AFF4 (residues 300-600) could bind the C-terminus of ELL2 (residues 518-640). His-AFF4$_{300-600}$ and ELL2$_{518-640}$ purified from *E. coli* were incubated at equimolar concentrations using the same conditions described above and analyzed by gel exclusion chromatography. ELL2$_{518-640}$ alone eluted at a volume corresponding to 26.8 kDa, but ELL2$_{518-640}$ incubated with His-AFF4$_{300-600}$ eluted at a volume corresponding to 61.9 kDa (Table 3.3, Figure 3.6A). Both proteins were present in the complex elution (Figure 3.6B), showing that AFF4 also directly binds the C-terminus of ELL2.

Finally, we tested whether AFF4 can bind the homologs ENL and AF9. Recent studies in the Zhou group demonstrate that the C-termini of ENL and AF9 interact with residues 600-900 of AFF4 (data not shown) and that ENL and AF9 compete for this interaction with AFF4. Moreover, the C-terminal 100 residues of ENL and AF9 share 68.8% sequence identity, strongly suggesting that the homologs interact with AFF4 in a similar way. Thus, we focused our analysis on the C-terminus of ENL (residues 433-559) to gain insight into how both ENL and AF9 interact with AFF4. ENL$_{433-559}$ alone eluted from the size exclusion column at a volume matching 14.2kD, while ENL$_{433-559}$ incubated with His-AFF4$_{600-900}$ in Buffer 0.4 eluted as a much larger 67.7kD species (Table 3.1, Figure 3.7A). Although Coomassie-stained SDS-PAGE of the elution reveals minor degradation of His-AFF4$_{600-900}$, we clearly observe the presence of both His-AFF4$_{600-900}$ and ENL$_{433-559}$ (Figure 3.7B) and conclude that AFF4 directly recruits ENL and AF9 through their C-termini. This *in vitro* reconstitution of binary complexes within the SEC demonstrate that *in vivo* interactions between AFF4 and other components are both direct and independent of each other. We conclude that AFF4 functions as a scaffold for the SEC via modular binding domains.

3.6 AFF4 can bind RNA through its N-terminus

Given that AFF4 is a positively charged scaffold in a complex that regulates mRNA elongation, we hypothesized that AFF4 may interact with RNA. Moreover, a recent study suggested that the C-terminus of members of the AF4 family exhibits RNA-binding (29), although the functional implications of this finding remain unclear. To expand our insight into the physical relationship between AFF4 and RNA, we tested whether the N-terminal half of AFF4 can bind RNA in an *in vitro* binding assay. We incubated His-AFF4$_{1-600}$ with a library of RNAs transcribed from a randomized synthetic 15-mer DNA template and observed that approximately 2.1% of the RNA pool bound to 100 μM His-AFF4$_{1-600}$ (Figure 3.7A). We repeated the assay with a His-AFF4$_{1-600}$ dilution series and generated a binding curve that reached saturation (Figure 3.7A). Next, we mapped which region within His-AFF4$_{1-600}$ is responsible for RNA-binding by also incubating His-AFF4$_{1-300}$ and His-AFF4$_{300-600}$ with the random RNA library and compared resulting binding curves with the curve derived from His-AFF4$_{1-600}$ (Figure 3.7B). Despite the widespread prevalence of charged amino acids in both regions, we found that only AFF4$_{1-300}$ produced a binding curve that closely matched that of His-AFF4$_{1-600}$, suggesting that RNA-binding is restricted to the first 300 residues.

To determine whether the N-terminus of AFF4 preferentially binds a specific RNA sequence, we extracted and amplified RNAs bound by His-AFF4$_{1-600}$ and performed iterative rounds of binding. We reasoned that if AFF4 selectively bound a specific RNA motif(s), we would observe enrichment of bound RNAs with each round. Indeed, there was an increase in the fraction of RNA bound after each round at the same His-AFF4$_{1-600}$ concentration (100 μM) (Figure 3.7C). After two rounds of enrichment, 31% of the RNA library was bound by 100 μM AFF4 (1-600) compared to only 2.1% from the first round (Figure 3.7C). Next, to identify motifs enriched in RNA bound by AFF4, we sequenced the library resulting from three rounds of enrichment and derived a position-specific scoring matrix (PSSM), which yielded 3 putative motifs with e-values less than 0.1 (Figure 3.7D). Although further validation both *in vitro* and *in vivo* will be necessary to determine the significance of these findings, these data provide preliminary insights the ability of AFF4 to physically associate with RNA. AFF4 is capable of interacting with RNA through its first 300 amino acids, and this region exhibits some selectivity for specific RNA motifs.

3.7 Discussion

After the discovery of P-TEFb as a human co-factor for Tat transactivation of viral transcription (30-32), much focus has been placed on trying to understand the physical and functional basis for stimulation of HIV elongation by P-TEFb. Although these studies laid the groundwork for our understanding of both HIV transcription and transcriptional elongation in general, efforts to expand our mechanistic understanding were hampered by the biochemical intractability of many transcriptional regulators and the seeming interconnectedness of numerous pathways. Recently, we and other groups have identified several new components of the Tat/P-TEFb complex, including AFF4, ELL2, ENL, and AF9 (25-27). These factors are inherently unstable in the cell and thus have eluded detection in the past. However, our approach of double-tagging the complex in the stabilizing presence of HIV Tat has allowed us to capture these ghost regulators and identify them using mass spectrometry.

Tat functionally synergizes with SEC components to stimulate elongation of HIV transcripts (12,25,27). Interestingly, the presence of Tat dramatically enhances ELL2 recruitment to the HIV LTR, as well as overall ELL2 levels in the cell through stabilization of ELL2 protein (12,27). AFF4, which directly binds the C-terminus of ELL2, similarly stabilizes cellular ELL2. These observations raise the question of whether Tat, like AFF4, recruits SEC components through direct contacts. However, our results show that Tat is not required for AFF4-dependent assembly of the SEC and that Tat has no intrinsic ability to directly bind ELL2 or ENL *in vitro*. Although it remains possible that HIV Tat does share important contacts with SEC subunits that are simply insufficient for full complex assembly, more quantitative analysis of these interactions will necessary to tease this possibility apart. Nevertheless, it seems that although Tat and AFF4 similarly enhance SEC assembly and stability, they do so through different mechanisms. Dissecting precisely how Tat regulates the complex will be critical for the design of drugs that specifically inhibit Tat-dependent SEC recruitment and function.

Our data also show that SEC factors bind AFF4 independently of each other *in vitro*, indicating an inherent compositional flexibility that may be modulated in the cell according to transcriptional requirements and allow greater SEC functional diversity. Although proteomics-based approaches identified many SEC components, dynamic variations in the composition of the complex when binding to different gene targets or different gene regions have not yet been characterized. Adding to this regulatory complexity is the redundancy between SEC components and their homologs. For example, we and other groups have shown that the AFF4 homolog, AFF1, can also support assembly of P-TEFb *in vivo*. Together, these data demonstrate the potential for combinatorial variability in SEC composition. To accurately define the function of these factors and understand why ELL2 and AFF4 are specifically required for HIV transcription, it will be necessary to determine how SEC composition is regulated in the cell both spatially and temporally.

We also discovered using an unbiased *in vitro* approach that the N-terminus of AFF4 exhibits the ability to bind RNA. Although further investigation will be necessary to confirm these preliminary results both *in vitro* and i*n vivo*, it seems likely that AFF4 may physically associate with RNA in light of its close proximity to nascent mRNA transcripts in the cell. It is unclear whether AFF4 preferentially binds one specific RNA motif or several different motifs or structural features. Also, if AFF4 does indeed have RNA binding partners *in vivo*, does this provide clues to AFF4 function or recruitment? Future studies will be important for addressing the significance of these early findings, but it seems likely that interactions between SEC components and RNA play some role in SEC function.

3.8 Materials and Methods

Purification of P-TEFb complexes

Hela cells stably expressing CDK9-HA and HEXIM1-F as well as TTAC-8 cells stably expressing CDK9-F and inducibly expressing Tat-HA were generated as described (33). Nuclear extracts prepared from cells were incubated overnight with anti-Flag M2 agarose beads (Sigma). The beads were washed in Buffer D0.3 (0.3M KCl, 20 mM HEPES-KOH (pH 7.9), 15% glycerol, 0.2 mM EDTA, 0.2% NP-40, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride) and immobilized proteins were eluted with 0.5 mg/ml Flag peptide dissolved in Buffer D0.3. The eluate was then incubated with anti-HA beads (Sigma) for 2 hr. The beads were washed with Buffer D0.3 and bound proteins were eluted by a low pH solution (200 mM glycine, pH 2.5). The eluate was neutralized (with 1/20 volume of 2 M Tris-HCl, pH. 8.8) and analyzed by SDS-PAGE followed by silver staining. Bands of interest were excised for identification by mass spectrometry.

In vivo co-IPs

Nuclear extracts were prepared from HeLa cells transfected with specific cDNA constructs. Anti-Flag or anti-HA agarose beads (purchased from Sigma-Aldrich) were incubated with nuclear extracts for 2hr at 4^{o}C. IPs were washed with D0.3 Buffer (0.3M KCl, 20mM HEPES pH 7.9, 10% glycerol, 0.2mM EDTA, 0.2% NP-40, 1mM DTT, 0.5mM PMSF), and purified materials were eluted off the beads with buffers containing synthetic Flag or HA peptides as described(. Co-IPs were analyzed by Western blotting with anti-ENL (A3202-267A) and anti-AF9 (A300-595A) purchased from Bethyl Laboratories, Inc. (Montgomery, TX) as well as previously described antibodies(25).

Recombinant protein expression

AFF4 cDNA fragments were inserted into the pET28b expression vector (Invitrogen) and CycT1 cDNA fragments were inserted into MCS1 of the pETDuet-1 expression vector (Invitrogen) containing an N-terminal $(His)_6$ tag. ELL2 and ENL fragments were cloned into pGEX-6P-3 vectors (GE Healthcare) containing an N-terminal GST tag. TEV-protease cleavage sites were engineered onto the N-terminus of expression fragments through PCR amplification. Tagged proteins were expressed in *E. coli* BL21 CodonPlus-(DE3)-RIL cells (Stratagene) grown in Terrific Broth (TB) at 37 °C until OD_{600} 0.5, followed by induction with 100 μM IPTG for 16 hours at 16 °C. CycT1 was expressed by auto-induction at 25 °C. Cells were harvested by centrifugation for 15 minutes at 5000 x g (4 °C) and frozen in liquid nitrogen for storage at –80 °C.

Purification of proteins

Full-length Flag-AFF4 was affinity-purified under highly stringent conditions (1.0 M KCl plus 0.5% NP-40) to remove binding partners as described (25). Flag-AFF4 was immobilized on anti-Flag-agarose beads and eluted with Flag peptide. Western blotting confirmed the absence of all known binding partners.

For purification of $(His)_6$-tagged proteins, cells were resuspended in Ni-A Buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM TCEP, 25 mM imidazole, 10% glycerol, 0.2 mM AEBSF) and lysed by sonication. Lysate was centrifuged for 1 hr at 20,000 x g (4°C), and the supernatant was passed over a 5 mL Ni-NTA affinity column (GE). Bound proteins were collected by gradient elution with Ni-B Buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM TCEP, 350 mM imidazole, 10% glycerol, 0.2 mM AEBSF) using an AKTA Explorer FPLC system. Fractions containing desired proteins were verified by Coomassie stain of SDS-PAGE gels. Tag cleavage required incubation for 22 hrs at 4°C with 1:50 $(His)_6$-TEV protease followed by a second Ni-affinity binding and elution step to collect cleaved (unbound) proteins. Affinity-purified proteins were further separated using an S75 gel filtration column in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP. $CycT1_{1-303}$ and $CycT1_{1-268}$ yielded approximately 10-15 mg protein/L culture, and AFF4 truncations yielded approximately 2-5 mg protein/L culture.

GST-tagged proteins were purified using a similar scheme, but instead cells were resuspended in GST-A Buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM TCEP, 5% glycerol, 0.2 mM AEBSF). Supernatant was passed over a 5 mL GST-FF column (GE), and proteins were collected by gradient elution with GST-B Buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM TCEP, 20 mM glutathione, 5% glycerol, 0.2 mM AEBSF). Tag cleavage also required incubation for 22 hrs at 4°C with 1:50 $(His)_6$-TEV protease followed by a second GST-affinity (to bind tag) and Ni-affinity (to bind TEV) column to collect cleaved (unbound) proteins. Affinity-purified proteins were further separated using an S75 gel filtration column in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP. ELL2 expression yielded approximately 1-3 mg protein/L culture, and ENL yielded approximately 1-2 mg protein/L culture. Purified proteins were concentrated to 1 mg/mL, frozen in liquid nitrogen and stored at -80° °C.

I*n vitro* affinity chromatography binding assays

Proteins were incubated at equimolar ratios at 4 °C for 30 minutes in Buffer 0.4 (0.4 M NaCl, 20 mM HEPES pH 8.0, 5% glycerol, 0.5 mM TCEP). GST-Sepharose or Ni-NTA agarose was added to relevant reactions and incubated with nutation for 20 minutes at room temperature. Beads were washed 6x using wash buffer (0.4 M NaCl, 20 mM HEPES pH 8.0, 10% glycerol, 0.5% NP-40, 0.5 mM TCEP), and proteins were eluted by boiling in 20 μL SDS PAGE Buffer (0.045 M SDS, 10% glycerol, 0.01% bromphenol blue, 50 mM DTT). Bound proteins were analyzed Coomassie blue.

Gel exclusion chromatography

Complex formation was carried out by combining purified proteins in 150 mM NaCl, 20 mM HEPES pH 8.0, 0.5 mM TCEP) at 4 °C for 20 minutes. Binding reactions were injected onto a Sephadex S75 column (GE) and run at 4 °C using an AKTA Explorer FPLC system. Eluted complexes were compared with eluted individual proteins as well as molecular weight standards.

SELEX

RNA-binding was performed via nitrocellulose filter binding assays (34). AFF4 protein was incubated in SELEX Buffer (20 mM HEPES, pH 7.6, 100 mM KCl, 0.5 mM DTT, 0.05% NP-40, 0.5 mg/mL BSA) with an RNA library transcribed from a randomized 15-nucelotide DNA library (Integrated DNA Technologies) in 100 μL at room temperature for 30 minutes. Selection was performed by filtration under vacuum through nitrocellulose filters and washing with 10 mL SELEX Buffer. The percent of bound RNA was calculated by Cherenkov counting, and RNA was extracted from the filters by phenol:chloroform extraction. Subsequent enrichment of libraries was carried out as described (35), and sequences were analyzed for motifs using MEME (36).

3.9 Tables

Table 3.1. Proteins identified by MS of purified 7SKsnRNP complex.

	# Peptides	% Coverage
CDK9	12	23.9
CycT1	29	33.9
HEXIM1	11	20.9
BCDIN3	19	26.9
PIP7S	4	5
Serine/threonine-protein kinase, RIO1	7	14
Serine/threonine-protein kinase, STK38	4	7.7
ATP-dependent RNA helicase, DDX17	4	8
ATP-dependent RNA helicase, DDX5	3	5.2
Serine/threonine-protein kinase, CCRK	4	8

Table 3.2. Proteins identified by MS of purified Tat/P-TEFb complex.

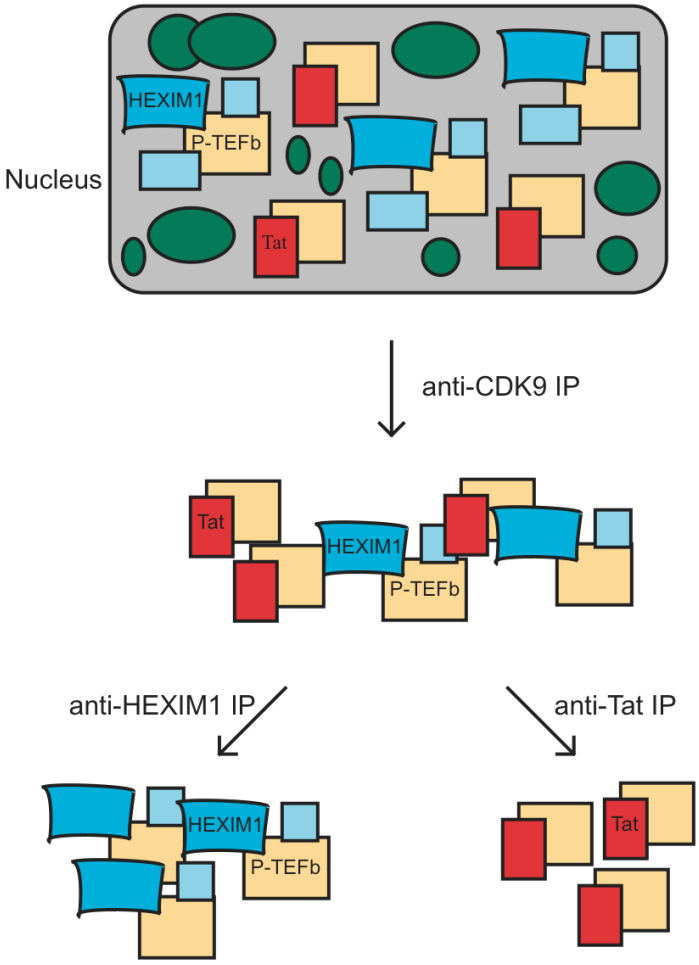
	# Peptides	% Coverage
CDK9	5	11.3
CycT1	5	9.2
AFF4	5	5.8
ELL2	2	3.3
ENL	2	3.8
AF9	1	2

Table 3.3. Gel exclusion chromatography elution volumes and calculated molecular weights.

	Elution Volume (mL)	Calculated MW (kD)
His-CycT1 (1-303)	10.56	40.6
His-CycT1 (1-303) + His-AFF4 (1-300)	9.17	81.0
ELL2 (518-640)	11.39	26.8
ELL2 (518-640) + His-AFF4 (300-600)	9.71	61.9
ENL (433-559)	12.67	14.2
ENL (433-559) + His-AFF4 (600-900)	9.53	67.7

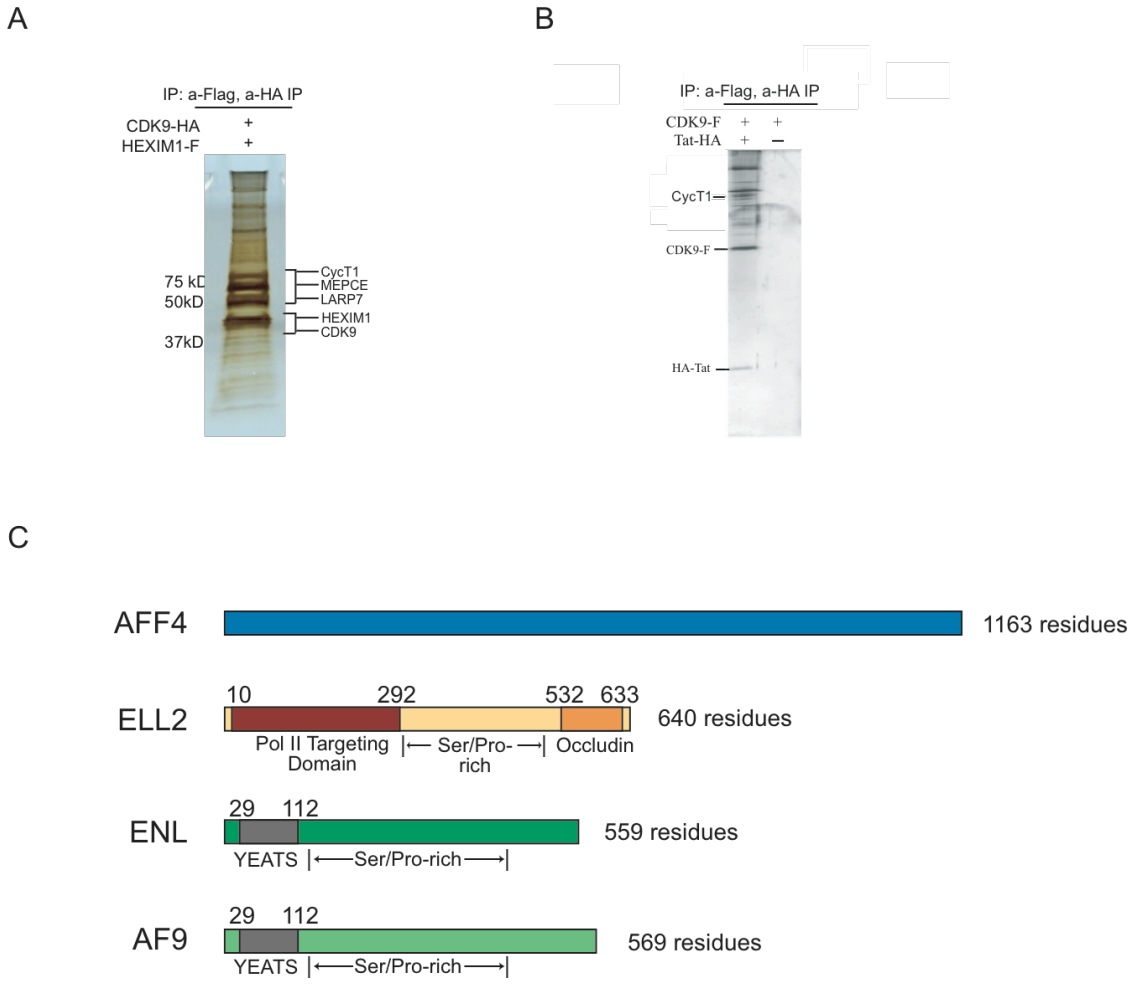
3.10 Figures

Figure 3.1. Isolation of P-TEFb complexes.



A double pull-down strategy was employed to purify native P-TEFb complexes from cells. In this approach, cell express two differentially-tagged proteins – one component of P-TEFb, CDK9, and one component specific to the active or inactive complex. The first purification step enriches for P-TEFb and all the proteins it is associated with. The second purification step specifically targets either Tat-bound or HEXIM1-bound P-TEFb.

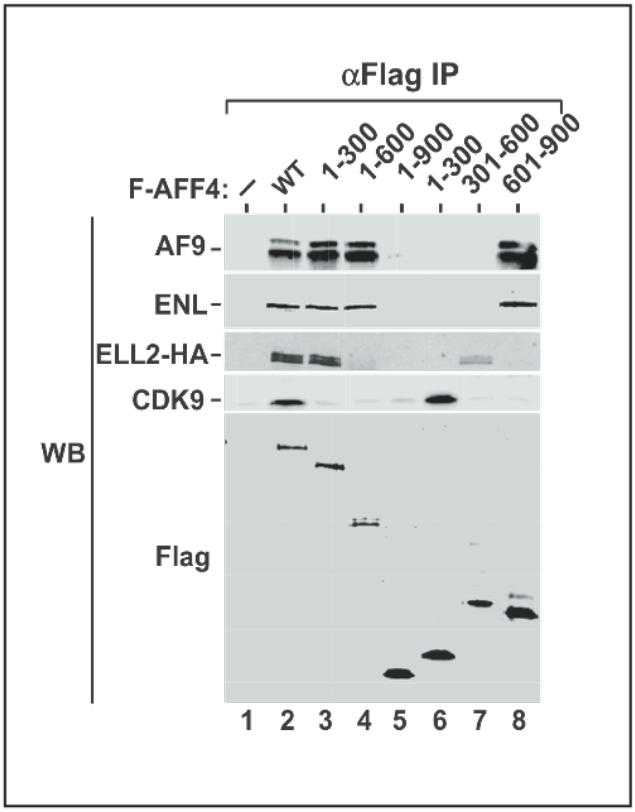
Figure 3.2. Identification of new Tat/P-TEFb partners.



A-B) CDK9-HA, HEXIM1-Flag and their associated factors were isolated through sequential immunoprecipitations (IPs). CDK9-F, Tat-HA, and their associated factors were also isolated through sequential IPs upon induction of Tat-HA expression. IPs from cells without Tat-HA expression were used as a negatve control (right). IPs from both (A) and (B) were analyzed on a silver-stained SDS gel, with their identities indicated.

C) Domain architecture of identified Tat/P-TEFb partners, AFF4, ELL2, ENL, and AF9. AFF4 has no known structural or functional domains.

Figure 3.3. AFF4 interacts with SEC components via independent interaction domains.



Anti-Flag IPs of nuclear extract from cells transfected with cDNA constructs expressing Flag-tagged AFF4 truncations were analyzed by Western blotting.

Figure 3.4. Model for SEC assembly along the scaffold AFF4.

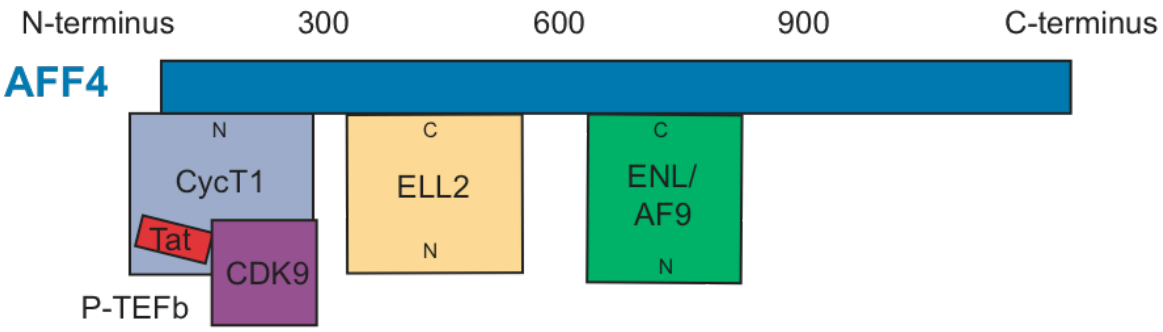
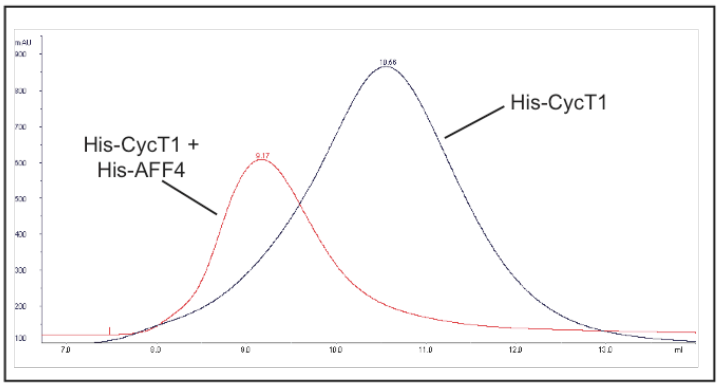
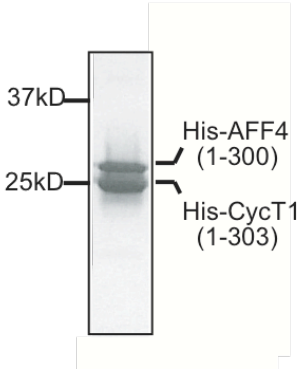


Figure 3.5. The N-terminus of AFF4 directly binds CycT1 of P-TEFb.

A



B

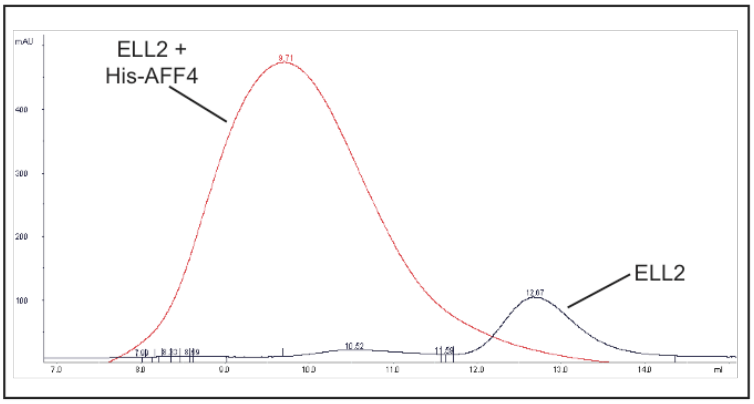


A) His-CycT1$_{1-303}$ elutes from a gel exclusion column at a volume corresponding to 40.6 kDa. His-CycT1$_{1-303}$ incubated with His-AFF4$_{1-300}$ causes a shift in elution volume corresponding to an apparent molecular mass of 81.0 kDa, suggesting the two proteins together form a larger complex.

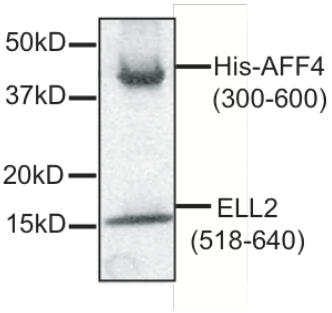
B) Coomassie stained SDS-PAGE analysis of the shifted binding reaction peak shows the presence of both His-CycT1$_{1-303}$ and His-AFF4$_{1-300}$.

Figure 3.6. AFF4 directly binds the C-terminal domain of ELL2.

A



B

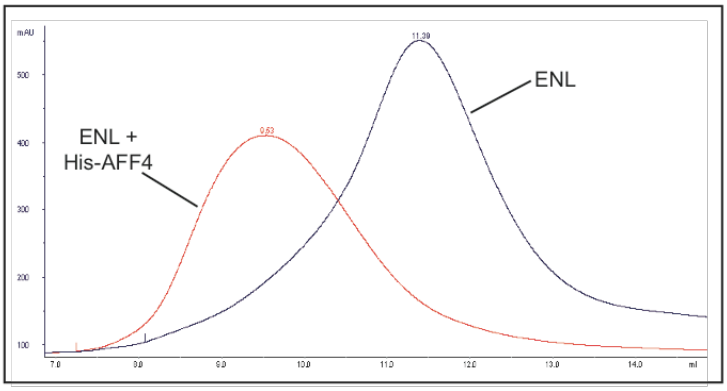


A) $ELL2_{518-640}$ elutes from a gel exclusion column at a volume corresponding to 26.8 kDa. $ELL2_{518-640}$ incubated with $His\text{-}AFF4_{300-600}$ causes a shift in elution volume corresponding to an apparent molecular mass of 61.9 kDa, suggesting the two proteins together form a larger complex.

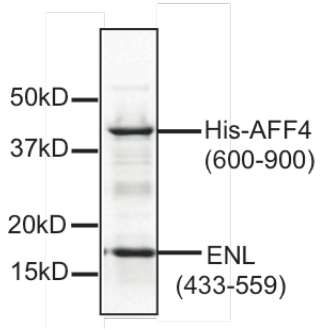
B) Coomassie stained SDS-PAGE analysis of the shifted binding reaction peak shows the presence of both $ELL2_{518-640}$ and $His\text{-}AFF4_{300-600}$.

Figure 3.7. AFF4 directly binds the C-terminus of ENL.

A



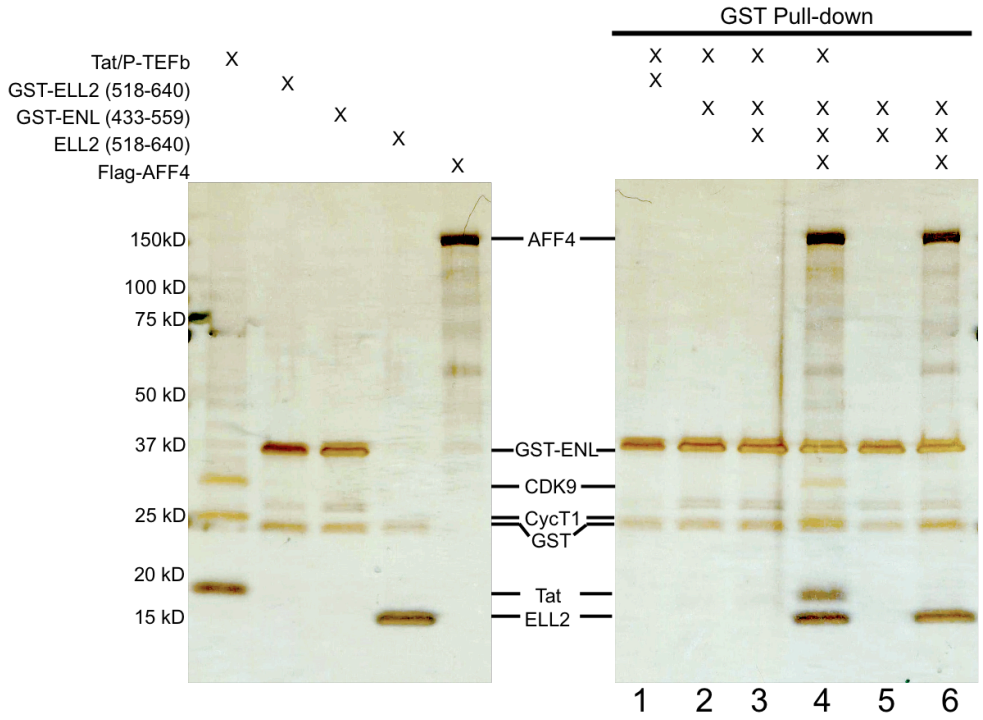
B



A) $ENL_{433-559}$ elutes from a gel exclusion column at a volume corresponding to 14.2 kDa. $ENL_{433-559}$ incubated with $His\text{-}AFF4_{600-900}$ causes a shift in elution volume corresponding to an apparent molecular mass of 67.7 kDa, suggesting the two proteins together form a larger complex.

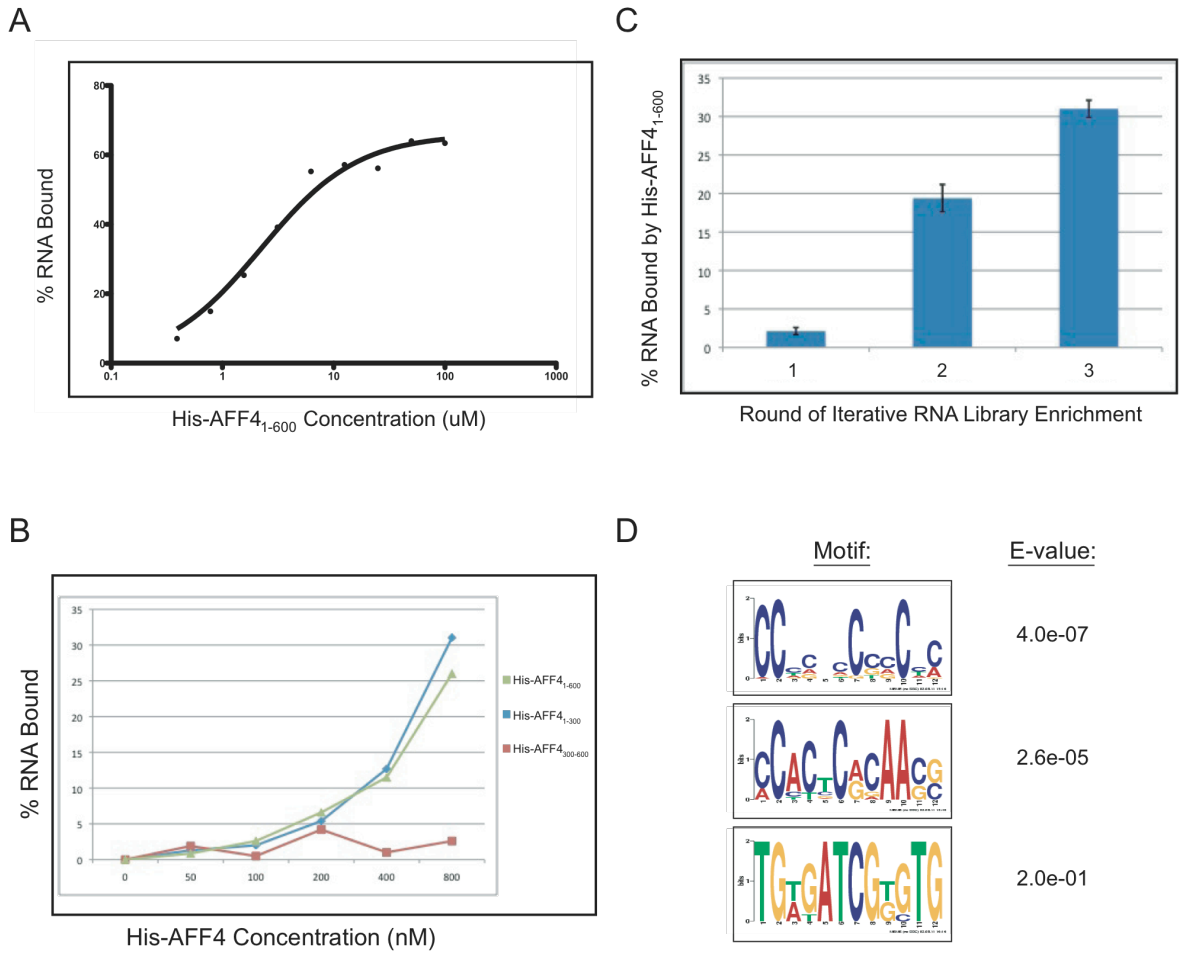
B) Coomassie stained SDS-PAGE analysis of the shifted binding reaction peak shows the presence of both $ENL_{433-559}$ and $His\text{-}AFF4_{600-900}$.

Figure 3.8. HIV Tat does not recruit SEC components through direct binding.



Purified SEC components (left) were assessed for assembly *in vitro* using a GST affinity capture assay. Tat/P-TEFb did not bind GST-ELL2$_{518-640}$ (lane 1) or GST-ENL$_{433-559}$ (lane 2). ELL2$_{518-640}$ and GST-ENL$_{433-559}$ also did not interact with each other (lane 5), even in the presence of Tat/P-TEFb (lane 3). Addition of full-length Flag-AFF4 restored these interactions both in the presence and absence of Tat/P-TEFb (lanes 4 and 6), suggesting that AFF4 is the central complex scaffold.

Figure 3.9. AFF4 can bind RNA through its N-terminal 300 residues.



A) A dilution series of His-AFF4$_{1-600}$ binding reactions with a random ^{32}P-labeled 15-mer RNA library generates a saturated binding curve (y=65.98x/(2.223 + x)).

B) Comparison of binding curves generated from His-AFF4$_{1-300}$ and His-AFF4$_{300-600}$ with His-AFF4$_{1-600}$ suggest that RNA-binding is restricted to the first 300 amino acids of AFF4.

C) Iterative rounds of binding of His-AFF4$_{1-600}$ with 15-mer RNAs extracted from each previous binding reaction shows an increase in the fraction of the RNA library bound, suggesting enrichment of a specific RNA motif(s).

D) The RNA library resulting from 3 rounds of enrichment was deep sequenced. MEME analysis (36) of sequences identified 3 sequence motifs with E-values < 0.1 that AFF4$_{1-600}$ may bind preferentially.

3.11 References

1. Zhou, Q., and Yik, J. H. (2006) *Microbiol Mol Biol Rev* **70**, 646-659
2. Bres, V., Yoh, S. M., and Jones, K. A. (2008) *Curr Opin Cell Biol* **20**, 334-340
3. Ptashne, M. (2005) *Trends Biochem Sci* **30**, 275-279
4. Buratowski, S. (2009) *Mol Cell* **36**, 541-546
5. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998) *Genes Dev* **12**, 343-356
6. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999) *Cell* **97**, 41-51
7. Saunders, A., Core, L. J., and Lis, J. T. (2006) *Nat Rev Mol Cell Biol* **7**, 557-567
8. He, N., Pezda, A. C., and Zhou, Q. (2006) *Mol Cell Biol* **26**, 7068-7076
9. Yang, Z., Yik, J. H., Chen, R., He, N., Jang, M. K., Ozato, K., and Zhou, Q. (2005) *Mol Cell* **19**, 535-545
10. Yang, Z., He, N., and Zhou, Q. (2008) *Mol Cell Biol* **28**, 967-976
11. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007) *Cell* **130**, 77-88
12. He, N., and Zhou, Q. (2011) *J Neuroimmune Pharmacol* **6**, 260-268
13. Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C., and Price, D. H. (2010) *Nature* **465**, 747-751
14. Krueger, B. J., Varzavand, K., Cooper, J. J., and Price, D. H. (2010) *PLoS One* **5**, e12335
15. Oesterreich, F. C., Bieberstein, N., and Neugebauer, K. M. (2011) *Trends Cell Biol* **21**, 328-335
16. Munoz, M. J., de la Mata, M., and Kornblihtt, A. R. (2010) *Trends Biochem Sci* **35**, 497-504
17. Thirman, M. J., Levitan, D. A., Kobayashi, H., Simon, M. C., and Rowley, J. D. (1994) *Proc Natl Acad Sci U S A* **91**, 12110-12114
18. Stock, W., Thirman, M. J., Dodge, R. K., Rowley, J. D., Diaz, M. O., Wurster-Hill, D., Sobol, R. E., Davey, F. R., Larson, R. A., Westbrook, C. A., and et al. (1994) *Leukemia* **8**, 1918-1922
19. Smith, E., Lin, C., and Shilatifard, A. (2011) *Genes Dev* **25**, 661-672
20. Harper, D. P., and Aplan, P. D. (2008) *Cancer Res* **68**, 10024-10027
21. Shilatifard, A., Lane, W. S., Jackson, K. W., Conaway, R. C., and Conaway, J. W. (1996) *Science* **271**, 1873-1876
22. Shilatifard, A., Duan, D. R., Haque, D., Florence, C., Schubach, W. H., Conaway, J. W., and Conaway, R. C. (1997) *Proc Natl Acad Sci U S A* **94**, 3639-3643
23. Shilatifard, A., Haque, D., Conaway, R. C., and Conaway, J. W. (1997) *J Biol Chem* **272**, 22355-22363
24. Estable, M. C., Naghavi, M. H., Kato, H., Xiao, H., Qin, J., Vahlne, A., and Roeder, R. G. (2002) *J Biomed Sci* **9**, 234-245

25. He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., and Zhou, Q. (2010) *Mol Cell* **38**, 428-438
26. Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., and Shilatifard, A. (2010) *Mol Cell* **37**, 429-437
27. Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010) *Mol Cell* **38**, 439-451
28. Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M. L. (2010) *Cancer Cell* **17**, 198-212
29. Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J., and Bardoni, B. (2011) *Hum Mol Genet* **20**, 1873-1885
30. Mancebo, H. S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D., and Flores, O. (1997) *Genes Dev* **11**, 2633-2644
31. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., and Jones, K. A. (1998) *Cell* **92**, 451-462
32. Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., and Price, D. H. (1997) *Genes Dev* **11**, 2622-2632
33. Yik, J. H., Chen, R., Nishimura, R., Jennings, J. L., Link, A. J., and Zhou, Q. (2003) *Mol Cell* **12**, 971-982
34. Tuerk, C., and Gold, L. (1990) *Science* **249**, 505-510
35. Amarasinghe, A. K., MacDiarmid, R., Adams, M. D., and Rio, D. C. (2001) *Rna* **7**, 1239-1253
36. Bailey, T. L., and Elkan, C. (1994) *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36

Chapter 4

The Structural and Molecular Basis for Super Elongation Complex Assembly and Function

4.1 Abstract

The Super Elongation Complex (SEC) is a multi-protein complex comprised of P-TEFb, AFF4/1, ELL2, ENL, and AF9 that stimulates human and HIV transcriptional elongation. HIV Tat recruits the SEC to the HIV promoter, making the SEC a key regulator of HIV transcription. Recent studies also show that the SEC has a widespread role in regulating leukemogenesis and general endogenous transcription. To probe mechanisms of SEC function and assembly we established the overall molecular architecture of the SEC by mapping binding sites that mediate complex assembly *in vitro*. Using limited proteolysis and sequence-based analyses, we also examined the folded state of SEC components. Our results show that AFF4 is the central scaffold that recruits other SEC factors through direct interactions with short regions along its structurally disordered axis. In addition, both ELL2 and ENL directly interact with the scaffold of another transcriptional regulator, the PAF Complex. Thus, AFF4 is a flexible scaffold with binding partners that act as bridging components to link the SEC to P-TEFb to a larger network of transcriptional regulators.

4.2 Introduction

Eukaryotic transcription occurs in discrete steps of initiation promoter escape, elongation and termination, and RNA polymerase II (Pol II) activity is tightly regulated throughout. Transition into each stage is guarded by several factors, which serve as checkpoints (1-4). Additionally, differential phosphorylation of the Pol II C-terminal domain (CTD) during transcription allows preferential binding of stage-specific modulators of Pol II activity (5). Historically, focus has been placed on transcriptional initiation, but mounting evidence suggests that elongation is the rate-limiting step for most actively expressed genes during cell growth and differentiation (2,6-8).

The approximately 9 kB HIV genome integrates into the host DNA upon infection, and the transcription of the complete HIV RNA as a single long unit makes the process especially sensitive to defects in elongation. Identified as a co-factor for the HIV accessory protein Tat, human positive transcription elongation factor (P-TEFb), a heterodimer of the kinase CDK9 and CycT1, phosphorylates Ser2 of the CTD heptad repeat YSPTSPS, and disruption of this activity inhibits elongation (5). P-TEFb also acts as a gatekeeper for paused Pol II escape into elongation by phosphorylating and subsequently alleviating two negative elongation factors that block Pol II: NELF and DSIF(9-11). Thus, HIV provides a powerful system for studying transcription in general, and further probing Tat-dependent transactivation can provide additional clues into the mechanisms by which elongation is controlled.

Recently, proteomics-based studies by our group and others identified additional human proteins involved in regulating HIV transcription, defining a larger HIV Tat-containing complex known as the Super Elongation Complex (SEC) (12-14). The SEC contains known elongation factors, P-TEFb and ELL2, as well as several other factors, including AFF4 and the homologs ENL and AF9. Interestingly, P-TEFb kinase activity is required for Tat-dependent enhancement of ELL2 recruitment to the SEC (12). A recent study in embryonic stem cells suggests that the SEC resides at most actively transcribed human genes and is required for stimulating transcription during differentiation (15). These results suggest that the SEC has a general and widespread role in regulating transcription in addition to a specific role in HIV transcription. Interestingly, HIV specifically requires the action of AFF4 and ELL2 (12,13,16) and not their homologs, although genetic experiments suggest redundancy among members of these protein families.

To better understand the role of the SEC in both endogenous and HIV transcription, characterization of SEC organization and biochemical function is needed. Although biochemical studies have advanced our understanding of how the SEC subunits are organized (12,13,17-20), the contact points between components have not been defined and sequence-based analysis of SEC components fail to identify any clear signatures of structurally defined domains. In chapter 3, we showed that AFF4 is the central scaffold of the SEC and

nucleates assembly of the complex through modular interaction domains that independently and directly bind other subunits. Here we characterize the overall structural architecture of the SEC through a combination of interaction mapping and limited proteolysis. In striking contrast to other scaffolds, AFF4 is largely unstructured and likely to be largely flexible, even in the presence of its binding partners. Binding sites in AFF4 map to short hydrophobic clusters linearly arrayed in the sequence and separated by long flexible segments. ELL2 and ENL/AF9 function as bridging modules by mediating an interaction between SEC and another transcriptional regulator, the PAF Complex (PAFc). These results establish a model for SEC assembly and show that CycT1, ELL2 and ENL/AF9 bridge the complex to a larger network of transcriptional regulators.

4.3 AFF4 scaffolds SEC components along a flexible axis

Biochemical analyses described in Chapter 3 show that AFF4 is a central factor for multiple interactions of the SEC and constitutes the scaffold that coordinates complex assembly. However, AFF4 is a 1163-amino acid long, serine-rich protein with long stretches of hydrophilic residues and no known structural motifs, raising the question of how AFF4 structurally supports SEC complex formation. To further define AFF4 scaffolding properties, we carried out sequence-based secondary structure analysis using DisoPred prediction tools (21). Strikingly, we found based on a 2% false positive rate threshold that the 94% of characterized AFF4 binding regions (residues 1-900) are predicted to be disordered (Figure 4.1A). In contrast, only 13.1% of $CycT1_{1-268}$, the structurally characterized AFF4-binding domain of P-TEFb, is predicted to be disordered (Figure 4.1B).

To test the prediction that the AFF4 binding regions are intrinsically disordered, we assessed the susceptibility of AFF4 to proteolysis in an *in vitro* assay using trace amounts of proteinase K. The susceptibility of a protein to proteolysis depends on the accessibility of cleavage sites and thus provides a measure of protein folding. Protein-protein interactions can mask proteolytic cleavage sites and thus report on the burying of proteolytic sites upon binding. We incubated recombinant $AFF4_{1-300}$, $AFF4_{300-600}$, and $AFF4_{600-900}$ with proteinase K for 10 minutes at 4 ^{o}C and analyzed proteolysis by Coomassie-stained SDS-PAGE (Figure 4.1C). All three fragments of AFF4 were extensively cleaved under these conditions, which is suggestive of high intrinsic disorder of AFF4.

To assess whether the AFF4 binding partners mask proteolytic cleavage sites in AFF4 or promote transitions from disorder to order, we compared fragmentation of AFF4 alone to AFF4 in complex with $CycT1_{1-268}$, $ELL2_{518-640}$, and $ENL_{433-559}$. The AFF4 fragmentation pattern was similar in the absence and presence of the three binding partners (Figure 1C), indicating that all AFF4 protease cleavage sites are still readily accessible in the complexes. Moreover, in contrast to AFF4, CycT1, ELL2, and ENL were resistant to proteolysis under the conditions tested, further highlighting the unusual proteolytic susceptibility of AFF4.

4.4 AFF4 recruits subunits through short, hydrophobic binding sites

Sequence-based structure prediction and limited proteolysis indicate that AFF4 is extensively unfolded, suggesting that AFF4 mediates SEC assembly through unstructured regions. To test this idea, we further mapped the AFF4 binding sites. Because protein-protein interactions are largely driven by the burial of hydrophobic surfaces, we sought to initially define candidate hydrophobic regions of AFF4. A hydropathy plot detected only a few short hydrophobic clusters interspersed between highly hydrophilic stretches along AFF4 (Figure 4.2A).

AFF4 regions 300-600 and 600-900 contained only one major hydrophobic cluster each. To test whether these hydrophobic clusters were sufficient to recruit

ELL2 and ENL, we generated 20-residue peptides corresponding to these sites and tested their ability to bind ELL2 and ENL in a native gel shift assay (Figure 4.2B-C). ELL2 and ENL were separated on native gels individually and after incubation with peptides. A peptide encompassing $AFF4_{318-337}$ but not $AFF4_{303-322}$ caused a change in electrophoretic mobility of $ELL2_{518-640}$. Also, a peptide of $AFF4_{710-729}$ changed the electrophoretic mobility of $ENL_{433-559}$ on the native gel, indicating an interaction between ENL and the peptide. In conjunction with our analysis of AFF4 structural order, these data suggest that AFF4 directly recruits ELL2 and ENL/AF9 via discrete, short, hydrophobic binding modules connected by linker regions that remain flexible upon complex assembly.

In contrast to the other binding regions, AFF4 residues 1-230 contains multiple high-scoring hydropathy peaks that might mediate binding to CycT1. To test whether these hydrophobic regions correspond to one or several CycT1 binding sites within AFF4, we analyzed binding between recombinant $CycT1_{1-268}$ and recombinant AFF4 truncations of $AFF4_{1-230}$. We incubated proteins at equimolar concentrations (10 μM) in 0.3 M NaCl, 20 mM HEPES pH 8.0, 0.5 mM TCEP for 30 minutes at 4 ^{o}C followed by gel exclusion chromatography. When combined with $AFF4_{1-209}$ or $AFF4_{80-300}$, His-$CycT1_{1-268}$ and each AFF4 fragment eluted at volumes corresponding to molecular weights of the individual proteins (data not shown), indicative of no complex formation. These results suggest the interaction between $CycT1_{1-268}$ and $AFF4_{1-230}$ may involve several contact points along AFF4.

The AFF4 homolog, AFF1, also recruits several SEC components *in vivo*, including ENL and P-TEFb (18). Comparison of the AFF1 and AFF4 sequences reveals that the ENL-binding site of AFF4 shares only 45% sequence identity in AFF1 (Figure 4.3A). This low sequence identity suggests that not all residues contribute to binding or that AFF1 and AFF4 bind ENL differently. On the other hand, the AFF4 residues that constitute the ELL2-binding site (residues 318-337) are 100% conserved in AFF1 (Figure 4.3A). These data suggest that AFF1 also binds ELL2, which has yet to be demonstrated. To test these predictions, we co-precipitated Flag-tagged AFF1 and ELL2 *in vivo*. AFF1 bound ELL2 as well as several other SEC components previously shown to interact with AFF1, including ENL, AF9, and CDK9 (Figure 4.3B). Thus, AFF1 physically associates with ELL2, which correlates with our observation that the ELL2 binding domain we define in AFF4 is sufficient to interact with ELL2 in AFF1.

4.5 ELL2, ENL, and AF9 bridge the SEC to PAFc and Pol II

During transcriptional elongation, Pol II activity is controlled by several modulators, including the human Polymerase-Associated Factor complex (PAFc) (24,25). Recently, it was shown that PAFc also associates with the HIV Tat-containing SEC complexes (13), suggesting there may be direct physical contacts between these two complexes important for regulating viral transcription. To explore this hypothesis, we assessed whether the N-termini of

ELL2 and ENL/AF9 are required for the interaction between PAFc and SEC in cells. We first tested the ability of several ENL deletion mutants to physically associate with the PAFc scaffold, PAF1, in IPs from HeLa cells. Interestingly, ENL lacking its N-terminal YEATS domain no longer co–precipitated with PAF1, while ENL lacking its C-terminal AFF4-binding domain retained this interaction (Figure 4.4A). We made similar observations for the ENL homolog, AF9 (Figure 4B), suggesting that the YEATS domain mediates the connection between PAFc and SEC by interacting with PAF1. To determine if the YEATS domain was sufficient for direct binding to PAF1, we purified recombinant His-PAF1 and GST-tagged ENL truncations and tested whether they could interact in a pull-down assay in vitro. Indeed, we found that the YEATS-containing N-terminus of ENL (GST-ENL-N) interacted with His-PAF1, while the C-terminus of ENL (GST-ENL-C) could not (Figure 4.4C). These results show that ENL/AF9 functions as a bridge composed of two binding modules: a C-terminal domain that interacts with AFF4, the scaffold of the SEC, and an N-terminal domain that interacts with PAF1, the scaffold of PAFc.

In our analysis, we also found that the N-terminus ELL2 is important, although not required, for association of the SEC with PAFc. We tested the ability of ELL2 deletions to pull down with PAF1 *in vivo* and found that the C-terminal deletion mutant ELL2 (Δ499-640) co-precipitated with PAF1, while the N-terminal deletion mutant ELL2 (Δ50-194) showed reduced binding to PAF1 (Figure 4.5A). Additionally, when we co-expressed His-PAF1 with either GST-ELL2$_{50\text{-}194}$ or GST-ELL2$_{518\text{-}640}$ in *E. coli*, only GST-ELL2$_{50\text{-}194}$ co-purified with His-PAF1 after sequential elutions from Ni- and GST-affinity columns (Figure 4.5B), confirming a direct interaction between PAF1 and the N-terminus of ELL2. Therefore, the N-terminus of ELL2, which is not required for SEC assembly, binds PAF1 *in vivo* independently of ENL/AF9. Together, these results provide initial insights into how two major regulatory complexes function within a physically-linked network to control transcriptional elongation and establish a new structural framework for SEC assembly (Figure 4.6).

4.6 Discussion

Transcriptional elongation is mediated by large protein assemblies. Understanding the spatial organization of these large complexes is important to understand their specific functions. Although significant progress has been made to identify and characterize the components of transcriptional elongation complexes, the underlying structural organization of these large complexes is often beyond the reach of three-dimensional structure determination. In addition to their large size, intrinsic disorder of SEC components such as the MLL-associated elongation factors add further challenges for structural analysis. Here, we define the overall architecture of the Super Elongation Complex (SEC) using a combination of interaction-mapping and limited proteolysis, and identify novel contact points between SEC and the transcription elongation regulator, PAFc.

Several unique principles of SEC complex organization arise from our data. Scaffolding proteins influence signaling events by physically assembling individual molecules in a pathway to increase efficiency of interactions between them. Scaffolds are often structurally rigid platforms that control spatial organization of their partners through tethering mechanisms and, in some cases, allosterically control components of the signaling cascade (26). In contrast, AFF4 is a strikingly unstructured protein that coordinates its binding partners through short binding regions interspersed by flexible linker regions (Figure 4.6). While further analysis is necessary to determine whether there is local folding of the short binding sites, it seems that SEC components assemble on AFF4 like clothes on a line. The long and unstructured nature of AFF4 implies that flexibility may be the most important organizational principle required for SEC function. This flexibility allows the coordination of many components over long distances and provides mechanisms for dynamic adaptation to new binding partners and spatial requirements. Alternatively, flexibility may be the result of missing binding partners that target the charged regions of AFF4, such as RNA. It remains an open question whether or not RNA-binding by AFF4 is significant for SEC function, and the structural consequences of RNA-binding are currently unknown.

The binding modes of AFF4 binding partners provide insight into how flexibility translates into specific function. ENL/AF9 and ELL2 are predicted to have small, folded N- and C-terminal domains that are separated by linker regions with little predicted secondary structure. Both proteins bind the AFF4 scaffold similarly via their C-termini ends, while recruiting other transcriptional regulators through their N-terminal domains. These properties allow ELL2 and ENL/AF9 to bridge the SEC and PAFc, which may be important for the crosstalk between the complexes during transcription. Currently, it remains unclear whether ELL2 and ENL/AF9 bind similar regions of PAF1 and if these interactions have different functional consequences. Previous reports show that ELL2 stimulates Pol II during elongation (27), but it is unknown whether PAF1 is involved in this activity. Our analyses point to large intrinsic flexibility in the SEC, and further studies will be

important for dissecting the functional importance of spacing and organization of linker regions and protein-protein contact points within the complex.

Another key observation of our studies is that AFF4 protein-protein interactions map to the few hydrophobic segments in AFF4. A major challenge in predicting function of largely unstructured proteins is the identification of interaction domains. The strong correlation between hydrophobicity and protein-protein binding in AFF4 could reflect a general property of disordered proteins. By the same token, these data suggest that we have identified most AFF4 protein-protein interaction domains. One exception warranting further study is the C-terminal end of AFF4, which contains a cluster of hydrophobic regions that may mediate additional, not yet characterized interactions. Other studies have pointed to the possibility that this region mediates homo- or hetero-dimerization between AFF4 and AFF1 (18), although this has yet to be definitively determined. Another recent study showed that the SEC physically contacts MED26 of the Mediator Complex (20), raising the possibility that AFF4/AFF1 also connects to Mediator.

In addition to its role in HIV, the SEC plays an important role in leukemogenesis (16,17,28). Human mixed-lineage leukemia (MLL) is an H3K4 methyltransferase that, upon chromosomal rearrangement, forms chimeric proteins with transcriptional regulators, including members of the ELL family (ELL1-3), the AFF4 family (AFF1-4), and ENL/AF9 (29). Misdirection of the SEC through the C-termini of MLL fusion partners causes inappropriate expression of their altered gene targets, leading to oncogenic transformation in myeloid and lymphoblastic leukemia (28,29). Thus, the assembly sites between AFF4 and its binding partners are central to this leukemogenic transformation. From our studies, we find that AFF4 binding is almost exclusively encoded by short amino acid sequences with only minimal structural contributions, providing initial insights into how these interfaces may provide new targets for therapeutic interference of both HIV and leukemia. However, further characterization of how SEC assembly is regulated by post-translational modifications or expression levels will be important to allow more selective inhibition of SEC function at both leukemia-related gene targets and the HIV LTR.

Our data expand the current model of SEC assembly in several ways. First, although the presence of HIV Tat enhances SEC assembly and recruitment, Tat does not itself directly bind SEC subunits. AFF4 remains the major scaffold for both Tat-containing and endogenous SEC. Second, AFF4 contains modular binding regions that independently and directly bind SEC components CycT1, ELL2, ENL/AF9 through their N-terminal domains. Interestingly, AFF4 binds subunits through short, hydrophobic clusters, which are separated by long, flexible linkers. Finally, the assembled SEC components ELL2 and ENL also contain discrete binding domains on their N- and C-termini separated by a disordered linker that allow them to function as binding modules to connect the SEC to another transcriptional regulator, PAFc.

Our characterization of the SEC further provides a new paradigm for scaffold structure. Specialized cellular functions may require flexible scaffolds to support the full range of function and dynamic interaction. Although there may be post-translational modifications or additional binding partners that further alter the structural landscape, AFF4 clearly has flexible linker regions between binding sites, establishing a novel scaffolding architecture that is likely important for SEC function. Moreover, binding partners of AFF4 exhibit a significant degree of internal flexibility, resulting in distinct binding modules that allow dynamic connections with other transcriptional regulators. Future studies aimed at understanding mechanisms of SEC assembly and transcription stimulation will shed light on how this unique organization underpins its function in HIV, leukemogenic, and endogenous transcription in the cell.

4.7 Materials and Methods

Native peptide gel shifts

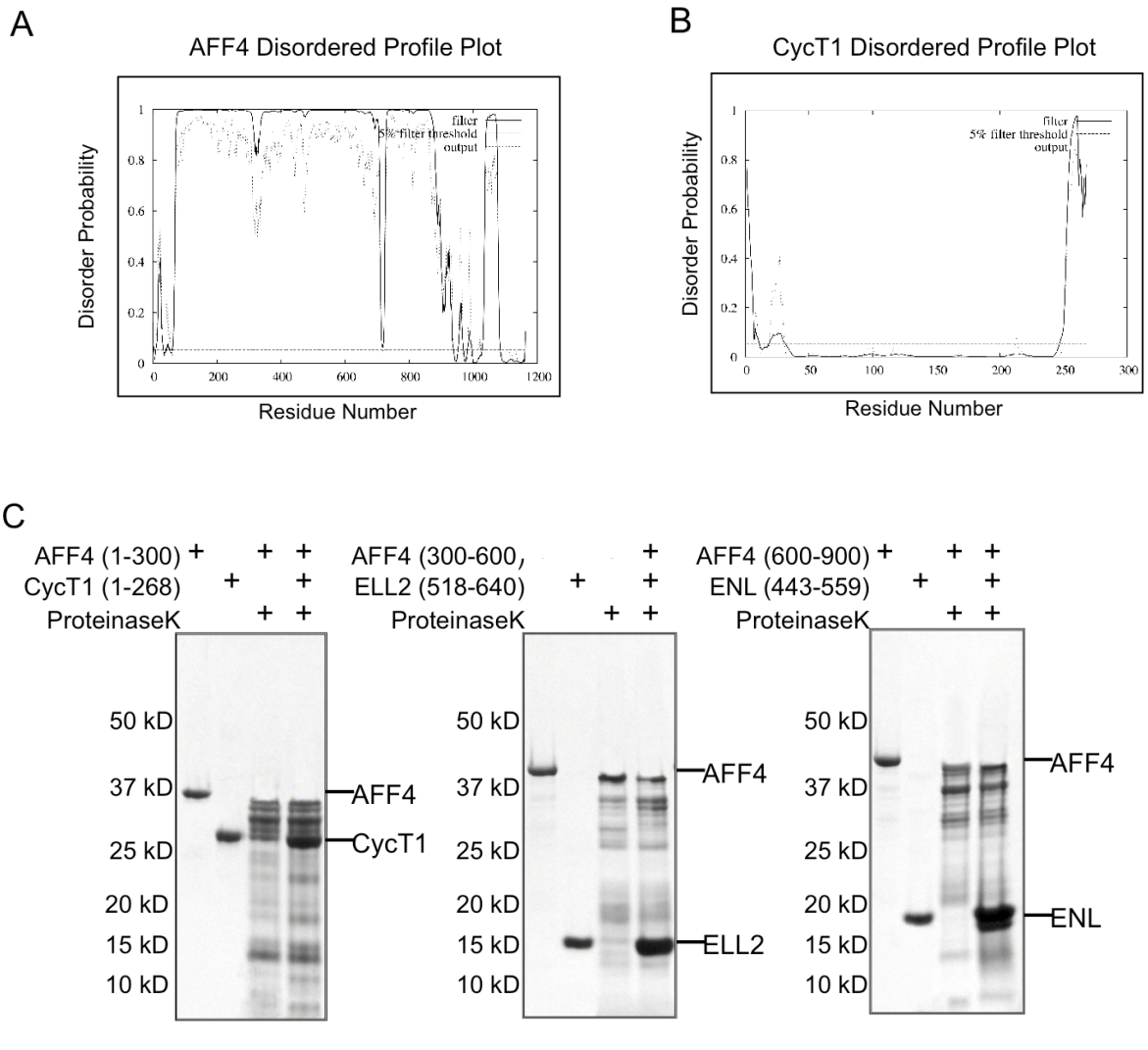
Peptide binding reactions were carried out in 100 mM NaCl, 20 mM HEPES pH 8.0, 0.5 mM TCEP. ELL2 or ENL (2 μg) was incubated in the presence or absence of 2 μg peptide (approximately 10x molar excess) for 15 minutes on ice. Reactions were separated by native gel electrophoresis in native gel buffer (0.025 M Tris, 0.192 M glycine, pH 8.5) at 4 ^{o}C (60V) for 3 hours using the BioRad 4-20% Tris-Glycine gel system. Gel shifts were analyzed by Coomassie-stain.

Limited proteolysis

Proteolysis of AFF4 was carried out in 150 mM NaCl, 20 mM HEPES pH 8.0, 0.5 mM TCEP. For binding reactions, proteins were first mixed at equimolar ratios for 30 minutes at room temperature. Reactions were mixed with proteinase K at a mass ratio of 1:2000 for 15 minutes at room temperature and stopped by addition of SDS-PAGE buffer. Protein fragments were separated by gel electrophoresis using the BioRad 4-20% Tris-Glycine gel system and stained with Coomassie blue.

4.8 Figures

Figure 4.1. AFF4 is a disordered protein in both the absence and presence of its binding partners.

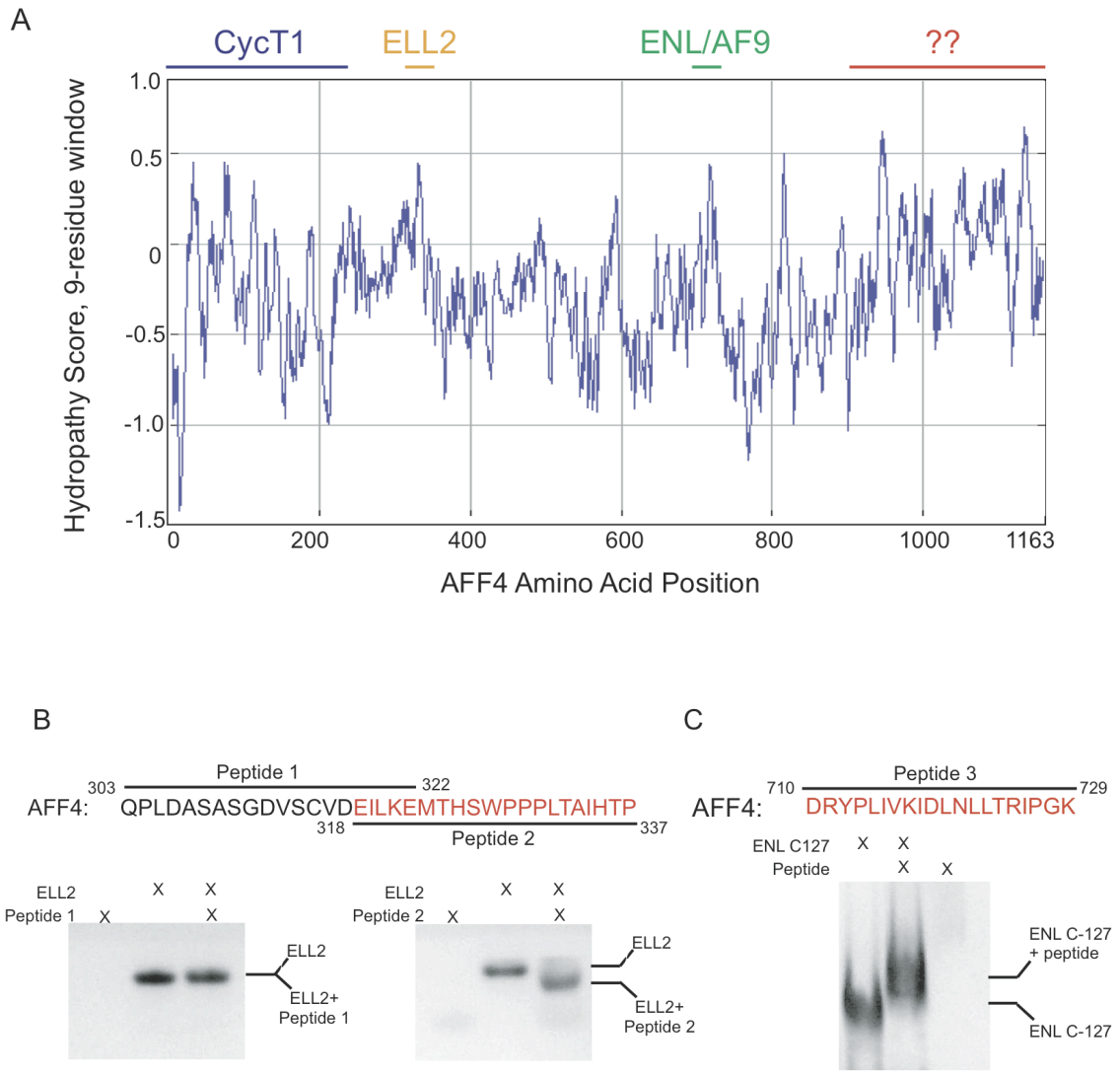


A) Disorder profile plot of AFF4 reveals significant predicted lack of folded structure. Profile was calculated using DisoPred (21).

B) Disorder profile plot of $CycT1_{1\text{-}268}$ reveals contrastingly low predicted disorder typical of a folded protein.

C) AFF4 is a natively unfolded scaffold. Limited proteolysis of AFF4 binding regions shows significant fragmentation in comparison with its binding partners. Complex formation with binding partners does not significantly alter fragmentation degree or patterns.

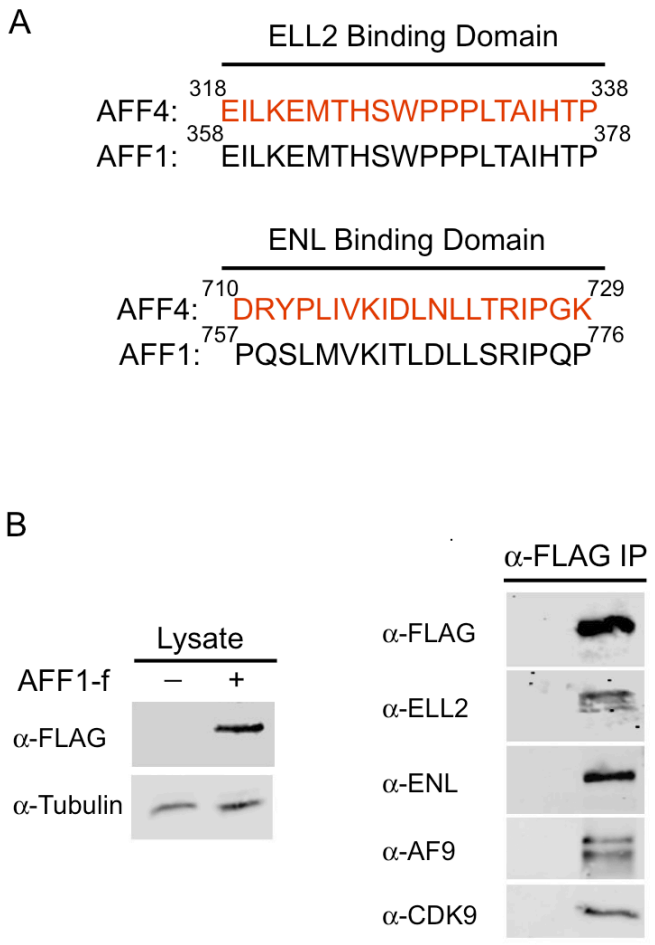
Figure 4.2. AFF4 recruits SEC components via short hydrophobic clusters.



A) Hydropathy plot of AFF4 calculated using a 9-residue window. Highest-scoring regions above 0.25 correspond to ELL2- and ENL-binding domains.

B) Peptides corresponding to hydrophobic clusters identified in AFF4 bind the C-terminal domains of ELL2 and ENL. Binding reactions were analyzed using native gel electrophoresis on 4-20% Tris-Glycine gels stained with Coomassie blue.

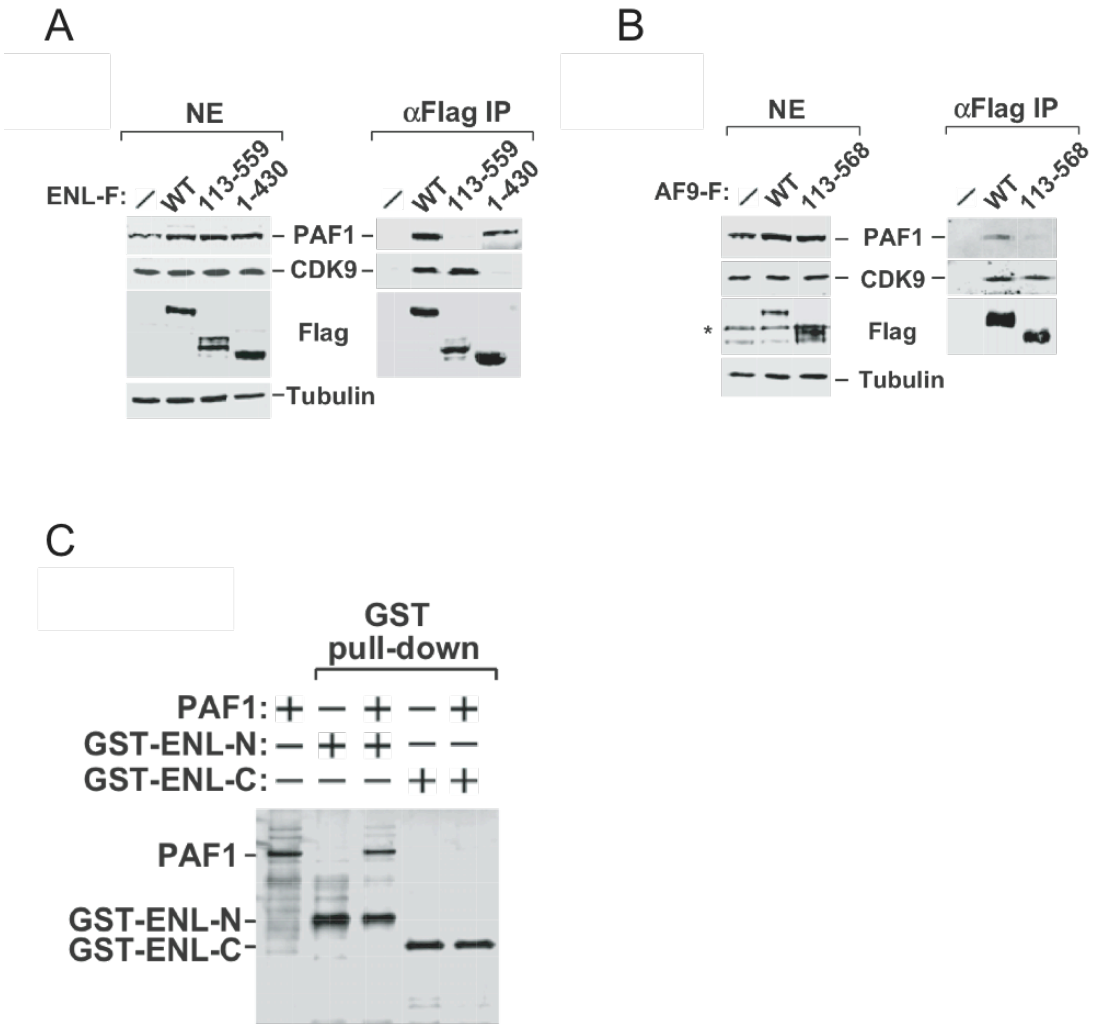
Figure 4.3. ELL2 binding site is conserved in AFF4 homolog, AFF1.



A) Sequence alignment of ELL2- and ENL-binding domains in AFF4 and AFF1.

B) AFF1 co-precipitates with ELL2 in HeLa cells. Western analysis of total cell lysate from HeLa cells transfected with AFF1-Flag (left) shows expression of AFF1. Western analysis of anti-Flag pull-down of these cells shows co-precipitation of ELL2, ENL, AF9, and CDK9.

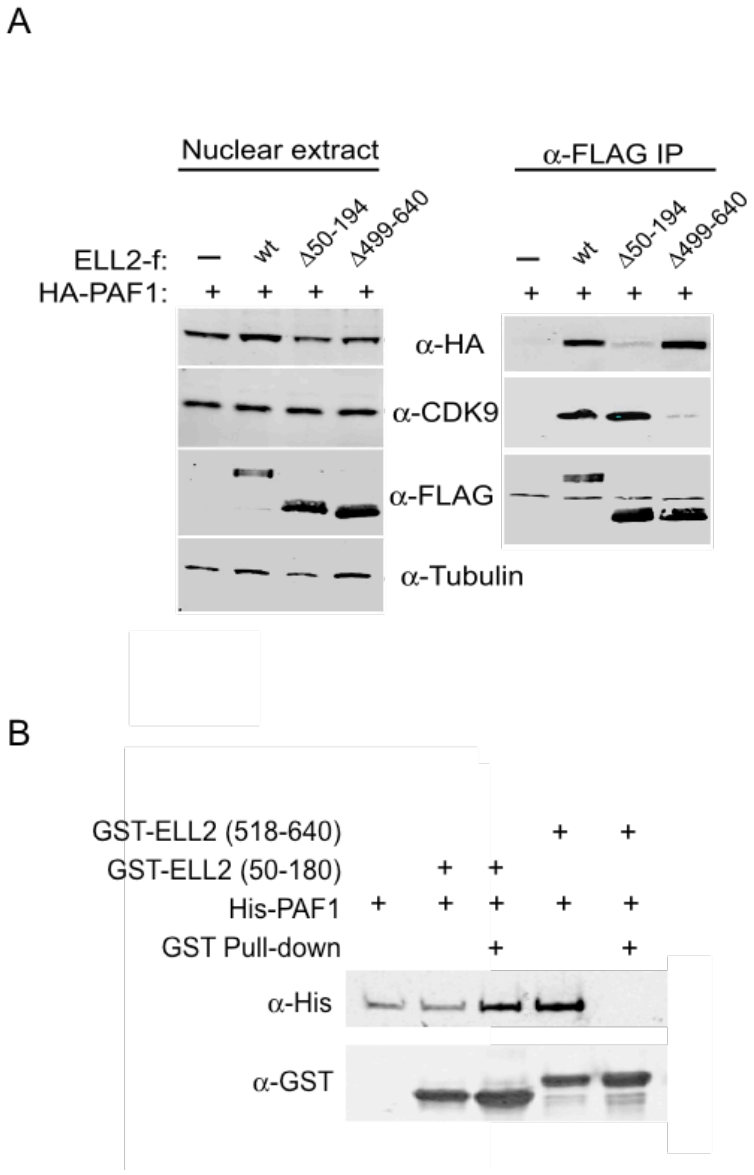
Figure 4.4. ENL physically connects the SEC to PAFc and Pol II



A-B) Immunoprecipitations from nuclear extracts of HeLa cells transfected with constructs expressing (A) ENL-Flag or (B) AF9-Flag were analyzed by Western blotting.

C) GST pull-down assays with purified recombinant proteins show that GST-ENL_{1-154}, but not GST-$ENL_{433-559}$, interacts with PAF1 *in vitro*. SDS-PAGE of pull-downs were analyzed by silver staining.

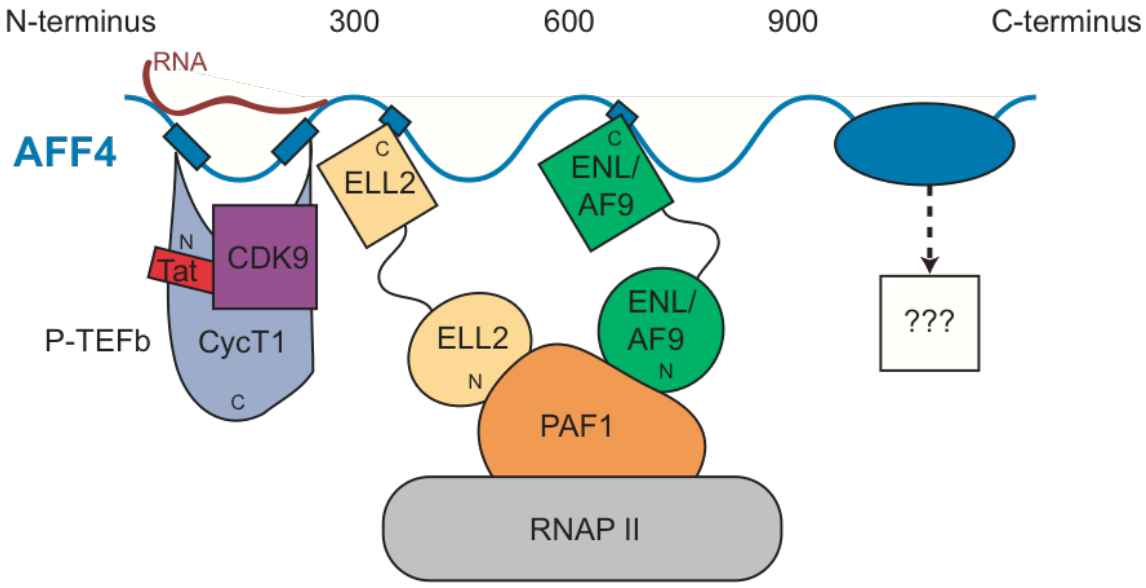
Figure 4.5. ELL2 acts as a physical bridge between the SEC and PAFc.



A) ELL2 co-precipitates with PAF1 in HeLa cells. Nuclear extracts prepared from HeLa cells co-transfected with HA-PAF1 and Flag-ELL2 wild-type, ELL2 (Δ499-640), or ELL2 (Δ50-194) were analyzed by Western blotting (left). Anti-Flag IPs were analyzed by Western blotting (right). HA-PAF1 was substantially diminished in the Flag-ELL2 (Δ50-194) IP (lane 3), suggesting the N-terminal domain of ELL2 is important for mediating an interaction with PAF1.

C) The N-terminal domain of ELL2 directly binds PAF1. The GST-ELL2$_{518-640}$ and GST-ELL2$_{50-194}$ were co-expressed with His-PAF1in *E. coli*. Western analysis of GST pull-downs of cell lysate shows co-purification of His-PAF1 with GST-ELL2$_{50-194}$ (lane 3) but not GST-ELL2$_{518-640}$ (lane 5).

Figure 4.6. AFF4 is a flexible scaffold with interspersed binding sites.



A new model of the SEC structural architecture arises from our data, suggesting that AFF4 is a highly disordered scaffold that contacts SEC subunits through short, hydrophobic binding sites. SEC components are also binding modules that connect the SEC to PAFc and Pol II.

4.9 References

1. Kuras, L., and Struhl, K. (1999) *Nature* **399**, 609-613
2. Sims, R. J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004) *Genes Dev* **18**, 2437-2468
3. Ptashne, M. (2005) *Trends Biochem Sci* **30**, 275-279
4. Saunders, A., Core, L. J., and Lis, J. T. (2006) *Nat Rev Mol Cell Biol* **7**, 557-567
5. Buratowski, S. (2009) *Mol Cell* **36**, 541-546
6. Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J., and Adelman, K. (2007) *Nat Genet* **39**, 1507-1511
7. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007) *Cell* **130**, 77-88
8. Core, L. J., and Lis, J. T. (2008) *Science* **319**, 1791-1792
9. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998) *Genes Dev* **12**, 343-356
10. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999) *Cell* **97**, 41-51
11. Cho, S., Schroeder, S., and Ott, M. (2010) *Cell Cycle* **9**, 1697-1705
12. He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., and Zhou, Q. (2010) *Mol Cell* **38**, 428-438
13. Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010) *Mol Cell* **38**, 439-451
14. Jager, S., Gulbahce, N., Cimermancic, P., Kane, J., He, N., Chou, S., D'Orso, I., Fernandes, J., Jang, G., Frankel, A. D., Alber, T., Zhou, Q., and Krogan, N. J. (2011) *Methods* **53**, 13-19
15. Lin, C., Garrett, A. S., De Kumar, B., Smith, E. R., Gogol, M., Seidel, C., Krumlauf, R., and Shilatifard, A. (2011) *Genes Dev* **25**, 1486-1498
16. He, N., and Zhou, Q. (2011) *J Neuroimmune Pharmacol* **6**, 260-268
17. Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., and Shilatifard, A. (2010) *Mol Cell* **37**, 429-437
18. Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M. L. (2010) *Cancer Cell* **17**, 198-212
19. Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J., and Bardoni, B. (2011) *Hum Mol Genet* **20**, 1873-1885
20. Takahashi, H., Parmely, T. J., Sato, S., Tomomori-Sato, C., Banks, C. A., Kong, S. E., Szutorisz, H., Swanson, S. K., Martin-Brown, S., Washburn, M. P., Florens, L., Seidel, C. W., Lin, C., Smith, E. R., Shilatifard, A., Conaway, R. C., and Conaway, J. W. (2011) *Cell* **146**, 92-104
21. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) *J Mol Biol* **337**, 635-645
22. Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984) *Proc Natl Acad Sci U S A* **81**, 140-144

23. Eisenberg, D., Wilcox, W., and McLachlan, A. D. (1986) *J Cell Biochem* **31**, 11-17
24. Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. (2009) *Genes Dev* **23**, 2765-2777
25. Kim, J., Guermah, M., and Roeder, R. G. (2010) *Cell* **140**, 491-503
26. Good, M. C., Zalatan, J. G., and Lim, W. A. (2011) *Science* **332**, 680-686
27. Shilatifard, A., Haque, D., Conaway, R. C., and Conaway, J. W. (1997) *J Biol Chem* **272**, 22355-22363
28. Smith, E., Lin, C., and Shilatifard, A. (2011) *Genes Dev* **25**, 661-672
29. Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L. M., Ge, K., Krumlauf, R., and Shilatifard, A. (2009) *Mol Cell Biol* **29**, 6074-6085

Appendix

The *Trypanosoma brucei* Life Cycle Switch *Tb*PTP1 Is Structurally Conserved and Dephosphorylates the Nucleolar Protein, NOPP44/46

A1. Abstract

Trypanosoma brucei adapts to changing environments as it cycles through arrested and proliferating stages in the human and tsetse fly hosts. Changes in protein tyrosine phosphorylation of several proteins, including NOPP44/46, accompany *T. brucei* development. Moreover, inactivation of *T. brucei* protein tyrosine phosphatase 1 (*Tb*PTP1) triggers differentiation of bloodstream stumpy forms into tsetse procyclic forms through unknown downstream effects. Here, we link these events by showing that NOPP44/46 is a major substrate of *Tb*PTP1. *Tb*PTP1 substrate-trapping mutants selectively enrich NOPP44/46 from procyclic stage cell lysates, and *Tb*PTP1 efficiently and selectively dephosphorylates NOPP44/46 *in vitro*. To provide insights into the mechanism of NOPP44/46 recognition, we determined the crystal structure of *Tb*PTP1. The *Tb*PTP1 structure, the first of a kinetoplastid PTP, emphasizes the conservation of the protein tyrosine phosphatase (PTP) fold, extending to one of the most diverged eukaryotes. The structure reveals surfaces that may mediate substrate specificity and affords a template for the design of selective inhibitors to interfere with *T. brucei* transmission.

A2. Introduction

T. brucei causes Human African Trypanosomiasis (HAT) or African sleeping sickness, which is marked by debilitating neurologic symptoms ranging from sensory impairment to the characteristic aberrant sleeping patterns that progress to coma. If untreated, HAT is fatal. With 30,000 deaths a year and 60 million people living at risk (1), HAT is a major disease burden in sub-Saharan Africa. Current drugs are ineffective and toxic, and drug resistance is becoming a growing hurdle for treatment (2).

T. brucei alternates between human and tsetse fly hosts, requiring extensive and rapid physiologic adaptations. In humans, the major *T. brucei* population consists of the extracellular, proliferative slender form in the bloodstream, which irreversibly differentiates into the G1-arrested stumpy form poised for transmission to the tsetse fly. Taken up by the tsetse fly, the stumpy form differentiates into the proliferative procyclic form in the insect midgut. Eventually, the tsetse salivary gland becomes populated with metacyclic forms, which infect the human host (3). This differentiation cycle requires survival in a diverse set of environments and forms the basis for infectivity and transmission.

The molecular signals, regulators, and effectors underlying this complex sequence of events are not well understood but could provide novel targets for therapeutic interference. Distinct patterns of protein tyrosine phosphorylation accompany and often precede stage progression (4), suggesting tyrosine phosphorylation is a key mechanism of developmental regulation. Studies of the *T. brucei* dual specificity kinases also provide evidence that tyrosine phosphorylation regulates the trypanosome life cycle (5-7). Moreover, NOPP44/46, a nucleolar RNA-binding protein required for ribosome biogenesis (8), exhibits dramatic changes in tyrosine phosphorylation in concert with the *T. brucei* life cycle transitions (9). NOPP44/46 is tyrosine phosphorylated in both proliferating procyclic and non-proliferating stumpy forms, but not in proliferating slender forms, indicating a complex interplay between life cycle and cell cycle in modulating tyrosine phosphorylation.

Recently, *Tb*PTP1, a PTP with sequence similarity to classical human PTPs, was identified as a central molecular switch for the stumpy-to-procyclic progression (10). *Tb*PTP1 activity arrests stumpy bloodstream forms, suggesting a model in which *Tb*PTP1 inactivation in the fly midgut releases the arrest and triggers development into the procyclic form (10). Thus, *Tb*PTP1 might function downstream of the recently described PAD transporters, which represent the first known step in the pathway that allows the differentiation signals citrate or cis-aconitate to trigger developmental changes (11). However, the substrates and downstream effects of *Tb*PTP1 remain unknown.

By sequence comparison, *Tb*PTP1 is similar to human classical PTPs such as the prototypical PTP1B. *Tb*PTP1 has an ortholog in *T. cruzi* and *L. major*, and

contains six regions that appear to be specific to trypanosomatids. The Kinetoplastida, including *T. brucei*, constitute some of the most diverged eukaryotes. The evolutionary distance and unique sequences raise the question if the overall structural conservation of the PTP fold is maintained in these distant eukaryotes.

Here, we identify NOPP44/46 as a major substrate of the life cycle switch *Tb*PTP1. We also describe the *Tb*PTP1 crystal structure, revealing strong conformational similarity to other eukaryotic PTPs and surface characteristics that help rationalize NOPP44/46 binding. Trypanosome-specific sequence motifs follow the canonical PTP fold, and all major functional elements are structurally conserved. These data establish the structural correlates of kinetoplastid PTPs within the PTP family and provide a new link in the signaling pathway controlling the stumpy-to-procyclic transition.

A3. NOPP44/46 is a major substrate of *Tb*PTP1

To identify cellular substrates of *Tb*PTP1, we generated a substrate-trapping mutant by replacing the general acid Asp199 with Ala. This mutant is catalytically inactive, but retains substrate binding and the ability to form the cysteinyl-phosphate intermediate, allowing for covalent trapping and isolation of substrates (24). Based on previous studies suggesting that *Tb*PTP1 is inactive and tyrosine phosphorylation most pronounced in procyclic forms (4,10), we used cell lysates from *T. brucei* procyclic forms for trapping experiments. Tyrosine phosphorylation was preserved throughout cell lysis by the addition of iodoacetamide and sodium orthovanadate to inhibit endogenous PTPs. The wild-type and Asp199Ala *Tb*PTP1 variants were covalently linked to NHS-Sepharose beads and incubated with *T. brucei* lysate. After high salt, guanidinium hydrochloride, and detergent washes, bound proteins were eluted by boiling in SDS-PAGE loading buffer with 1% DTT to release the covalently bound substrates and sodium orthovanadate to inhibit PTP activity. Western blotting of eluates using anti-pTyr antibody showed the enrichment of several tyrosine phosphorylated proteins by the Asp199Ala mutant relative to the wild-type *Tb*PTP1 (Figure A.1A). Two major bands migrated at a molecular weight of ~45kD and ~70kD, and a minor band at ~40kD. The enrichment of these bands was selective, as other pY proteins apparent in the total lysate were not trapped by *Tb*PTP1.

Because the molecular masses of the 45kD species corresponded to that of NOPP44/46, a protein known to be tyrosine phosphorylated in procyclic forms *in vivo* (4,9), we explored the possibility that NOPP44/46 was a trapped substrate of *Tb*PTP1. Trapping eluates were probed with anti-NOPP44/46, resulting in signal at the position identical to the band detected with anti-pY antibodies (Figure A.1B). Some phospho-independent binding of NOPP44/46 to wild type *Tb*PTP1 was also observed. The specific enrichment of NOPP44/46 with the trapping *Tb*PTP1 identifies NOPP44/46 as a potential *in vivo* substrate of *Tb*PTP1.

To confirm the observed interaction between *Tb*PTP1 and NOPP44/46, we tested dephosphorylation of NOPP44/46 by *Tb*PTP1 *in vitro*. Although the pH optimum of *Tb*PTP1 is 6 (10), this preference is unlikely to reflect physiologic function rather than the generally higher nucleophilicity of cysteine residues at low pH. *In vitro* dephosphorylation assays were therefore performed at pH 7.5, more similar to the pH at which *Tb*PTP1 likely functions. Phosphorylated NOPP44/46 was obtained by overexpressing a C-terminally TAP-tagged version of the full-length protein in procyclic forms. Rapid dephosphorylation of NOPP44/46 was observed at the lowest *Tb*PTP1 concentration tested (3.7 nM), and was complete at *Tb*PTP1 concentrations at or above 33 nM (Figure A.2A). To test if dephosphorylation of NOPP44/46 by *Tb*PTP1 is specific, we tested the activity of a panel of unrelated microbial PTPs on NOPP44/46. Of the five tested PTPs, only *Tb*PTP1 and *Yersinia* YopH dephosphorylated NOPP44/46; no activity was observed using *Mycobacterium tuberculosis* PtpA and PtpB, *Staphylococcus aureus Sa*PtpA, or *Listeria monocytogenes* lmo1935 (Figure

A.2B). With YopH known to have broad substrate specificity (25), these data show complete, efficient, and selective dephosphorylation of NOPP44/46 by *Tb*PTP1 *in vitro*.

To define the phosphorylation site(s) on NOPP44/46 recognized by *Tb*PTP1, we assessed the phosphorylation state of NOPP44/46 variants in which each of the five Tyr residues was replaced individually with Phe. Tyr181Phe completely abrogated tyrosine phosphorylation of NOPP44/46, as shown by anti-pTyr Western analysis (Figure A.3). None of the other mutations reduced the level of NOPP44/46 Tyr phosphorylation, indicating that Tyr 181 is the only phosphorylated Tyr and the target of *Tb*PTP1. The NOPP44/46 phosphorylation site is located in a 40-residue acidic loop encompassing residues 167-207. The sequence of this segment, 167-207, indicates that phosphorylation adds additional negative charges to a nearly uninterrupted acidic sequence. This acidic region containing the target Tyr is found only in *T. brucei* homologs.

A4. *Tb*PTP1 has a classical PTP fold

To explore the basis for recognition of the unusual substrate target sequence and the architecture of this diverged kinetoplastid PTP, we determined the crystal structure of *Tb*PTP1 at 2.4 Å resolution (Table A.1). The asymmetric unit contains two *Tb*PTP1 molecules with a root-mean-square-deviation (rmsd) of all atoms of 0.3 Å. The structure comprises residues Ser6-Ile295 in chain A and residues Met1-Leu298 in chain B (Figure 4), as well as one phosphate per *Tb*PTP1 and 187 water molecules. No clear electron density was visible for residues Val65-Ala74 of chain A and residues Leu66-Ala76 and Phe146-Asp152 of chain B. The backbone is interrupted between residues Arg46 and His47 of chain B, probably due to in-drop trypsin cleavage during crystallization. The protein contains four mutations. The catalytic cysteine is mutated to alanine, possibly through desulfurization by the phosphine TCEP present in high concentrations in the protein drop, and Glu138-144 were changed to alanine to improve the crystallization properties of the protein (17).

The overall fold of *Tb*PTP1 resembles that of other classical PTPs (26), with an extended, twisted β-sheet at the center and 7 α-helices surrounding it (Figure A.4A). The catalytic loop, or P-loop, is situated at the center of the active site and comprises the invariant PTP signature motif Cys-X5-Arg. A phosphate binds in the position similar to that of pTyr substrate phosphate in PTP:peptide substrate structures (27) (Figure A.4B). The active-site cavity is further delineated by the pTyr loop that deepens the cavity to ~9 Å, thus excluding pSer and pThr residues. The WPD loop containing the general acid Asp199 assumes a closed conformation, similar to that seen in other PTP structures with small ligands bound (28). *Tb*PTP1 contains all 10 PTP sequence motifs (29) in the same spatial organization as human PTPs. The six trypanosome-specific sequence motifs of *Tb*PTP1 follow the canonical PTP fold and do not give rise to new structural features. Consistent with a role in substrate recognition or regulation,

the trypanosome specific sequences predominantly map to the surface of *Tb*PTP1 (Figure A.5A). The phosphate engages in the typical interactions with the P-loop, hydrogen bonding with six main chain amides and the invariant Arg 235 side chain.

The closest structural homolog of *Tb*PTP1 found by the DALI server is the human PTPRO (Glepp1, PDB ID: 2G59), with a Cα rmsd of 1.9 Å. The prototypical human PTP1B phosphatase is the second closest structural homologue with a Cα rmsd of 2.3 Å. The superposition of *Tb*PTP1 with PTPRO shows the overall large similarity, with major differences only at the termini and surface loops (Figure A.5B). The *Tb*PTP1 loop from 62-79, although mostly invisible in the structure, has shifted at the base compared to the equivalent PTPRO loop and contains a single-residue insertion compared to PTPRO and up to seven residues compared to other human PTPs. The *Tb*PTP1 loop 138-154 containing a PEST sequence shows the most divergence from the PTPRO structure.

*Tb*PTP1 forms weak interactions between the two molecules in the asymmetric unit in the crystal (data not shown). The interactions comprise two symmetric salt bridges between Lys123 and Glu201, hydrogen bonds between Gly127 and Glu201, as well as 44 non-bonded interactions resulting in an interface of ~500 $Å^2$. PTPs such as PTPγ form dimers in the crystal and in solution (28). However, *Tb*PTP1 migrates as a monomer in size exclusion chromatography (data not shown), suggesting that the crystal contacts are not reflecting a physiologic state but rather are a result of crystal packing.

Although the three-dimensional organization of the PTP active site is highly similar in all PTPs across families, PTP surface properties vary widely and produce large diversity (28). The *Tb*PTP1 electrostatic surface shows distinct and continuous electronegative and positive areas (Figure A.6). The active site shows moderately electropositive potential, with the closed WPD loop burying additional electropositive regions of the phosphate binding pocket. A continuous electropositive stretch runs across the active site and along one side of the molecule. This stretch includes the side chains of Arg15, 23, 30, 50, 125, 175, 276, and Lys 113 and 131 (Figure A.6).

A5. Discussion

T. brucei requires stringent control of developmental programs to successfully infect its human and tsetse fly hosts, and a molecular hallmark of life-cycle transitions in *T. brucei* is the coordinated change in tyrosine phosphorlyation. Recently, the tyrosine phosphatase *Tb*PTP1 was identified as a key regulator of the trypanosome life cycle. *Tb*PTP1 has an ortholog in *T. cruzi* with 61.3 % sequence identity. Although an intracellular parasite, *T. cruzi* shares the bloodstream-to- insect route of transmission controlled by *Tb*PTP1 in *T. brucei*, suggesting that the function of the two PTPs may be conserved.

Despite the evolutionary distance between humans and trypanosomes, our crystal structure of *Tb*PTP1 shows a high degree of structural conservation of the conventional PTP fold. The 24% sequence identity of *Tb*PTP1 to human PTP1B translates into an unusually high structural similarity with a C-α rmsd of 2.3 Å. The conservation of the PTP fold thus extends not only to bacteria but also to ancient eukaryotes and underscores the utility and evolutionary success of this scaffold for tyrosine dephosphorylation. *Tb*PTP1 shares 9 of 10 signature motifs with the human PTPs and has 6 additional trypanosome-specific motifs that may play roles in functional regulation or substrate recognition. The folding of these motifs suggests that although not giving rise to new structural elements, their position mostly on the surface is consistent with a role in substrate recognition and/or regulation.

The *Tb*PTP1 structure also provides the basis for inhibitor design. The strong similarity to human PTPs highlights the need for structural information to guide the design of selective *Tb*PTP1 inhibitors as tools and potential therapeutics. *Tb*PTP1 prevents premature differentiation of stumpy bloodstream forms to procyclic forms, which lack immune evasion mechanisms that allow survival in the mammalian host. Thus, inhibition of *Tb*PTP1 would reduce the pool of tsetse-infective parasites within the mammalian host, potentially attenuating transmission, an approach that has gained acceptance for the reduction of malaria (30). Blocking transmission could be particularly advantageous for controlling animal trypanosomiasis, which affects livestock and remains a major hurdle to economic development in sub-Saharan Africa. Moreover, *T. brucei rhodesiense* infects both humans and animals, providing a parasite reservoir for human infection.

Although tyrosine phosphorylation is emerging as a key regulator of the trypanosome life cycle, little is known about the molecular pathways that lead to downstream developmental changes. Identification of *Tb*PTP1 substrates is essential to understanding the mechanisms by which *Tb*PTP1 regulates *T. brucei* differentiation. Among the phosphoproteins selectively enriched using a *Tb*PTP1 trapping mutant, we identified NOPP44/46 and a yet unidentified ~70kD phosphoprotein as a substrate of this PTP. The functional interaction of *Tb*PTP1 and NOPP44/46 is supported by efficient and selective *in vitro*

dephosphorylation. Furthermore, the phosphorylation pattern of NOPP44/46 during developmental stages, unlike that of other major tyrosine phosphorylated species, matches the proposed activity profile of *Tb*PTP1 in slender and procyclic forms (4,9). The presence of phosphorylated NOPP44/46 in stumpy forms may reflect decreasing *Tb*PTP1 activity or changes in the activity of the cognate kinase(s) in combination with a large increase in NOPP44/46 protein levels observed in stumpy forms (9).

The *Tb*PTP1 crystal structure allows rationalizing NOPP44/46 substrate binding. The distinct, continuous electropositive area running across the *Tb*PTP1 surface and the active site might be a footprint of electronegative regions of its substrate(s), such as the acidic stretch harboring the NOPP44/46 pTyr. Phosphorylation sites are usually found in flexible loop regions, suggesting that linear sequences rather than conformational sites serve as dephosphorylation substrates. This is consistent with the NOPP44/46 dephosphorylation site, which is predicted to be highly disordered. Moreover, the kinetics of substrate peptide turnover by PTPs are often approaching the limits of diffusion, suggesting that the selectivity and binding determinants of peptides are contained within the primary peptide sequence.

A phosphoproteomic study of tyrosine-phosphorylated proteins in *T. brucei* procyclic forms identified 34 phosphoproteins (31). NOPP44/46 Tyr181, however, was not identified, likely due to experimental limitations of that study. By immunofluorescence, phosphoproteins mostly associated with the cytoskeleton and the nucleolus, which is also the site of phospho-NOPP44/46 (32). A physiological interaction between *Tb*PTP1 and NOPP44/46 would require cellular co-localization, and consistent with this tenet, *Tb*PTP1 was found to associate with cytoskeletal and the nuclear fractions (10). However, it remains possible that NOPP44/46 is dephosphorylated outside of the nucleus as a recent study suggests that a pool of NOPP44/46 is exported out of the nucleus via exportin 1 (33).

*Tb*PTP1 is a molecular switch for the stumpy-to-procyclic transition, as both genetic and pharmacological inhibition of the phosphatase lead to spontaneous differentiation of committed stumpy forms to procyclic forms in vitro (10). Since trypanosomatids do not use transcriptional control to regulate expression of protein-coding genes (with a few exceptions), the key substrates of *Tb*PTP1 that mediate this effect are likely to modulate mRNA stability, translation, or protein turnover. The identification of a known ribosome biogenesis protein, NOPP44/46 (8), as a potential *in vivo* substrate of *Tb*PTP1, points towards a possible role in translational control. For example, the tyrosine phosphorylation state of NOPP44/46 may modulate ribosome biogenesis, which would in turn affect translational capacity. This possibility will be examined in future studies.

NOPP44/46 may also have uncharacterized cellular functions in addition to its essential role in ribosome biogenesis that could play specific roles in

differentiation. *In vivo*, NOPP44/46 acts a structural scaffold for several nucleolar proteins (34,35), and its phosphorylation state may modulate these interactions to promote specific cellular changes. Studies in other organisms suggest the existence of functional links between proteins involved in ribosome biogenesis and development, highlighting complex yet conserved modes of developmental regulation. In zebrafish, Pescadillo, an essential gene required for nucleolar assembly and 60S biogenesis, was originally discovered in a screen for regulators of embryonic development (36-39). In yeast, the Pescadillo ortholog, Yph1p, is found in two distinct multiprotein complexes with different functions in ribosome biogenesis and DNA replication (40). Interestingly, depletion of NOPP44/46 (unpublished results) and Yph1p both lead to cell cycle arrest with defective S-phase progression (40,41). Furthermore, the ribosomal biogenesis factors such as nucleolin and nucleophosmin play roles in the cytosol or nucleus distinct from their function in ribosome biogenesis (42,43). As other substrates of *Tb*PTP1 are identified and tools for the study of *Tb*PTP1 *in vivo* are refined, we will be better able to determine how these processes act together or apart to influence trypanosomatid differentiation and potentially provide insight into a novel signaling mechanism conserved in eukaryotic development.

A6. Materials and Methods

Cloning, protein expression, and purification

The full-length *Tb*PTP1 (systematic ID Tb10.70.0070) gene was amplified from genomic *T. brucei* DNA (kindly provided by Dr. Christian Klotz) and cloned into the pET28b expression vector in frame with the N-terminal six histidine tag. Point mutants were generated according to the QuikChange protocol (Stratagene). BL21 (DE3)-CodonPlus cells were transformed, and protein expression was induced at OD_{600} of 0.6 by adding 100 μM IPTG. After 20 hours of induction at 20 °C, cells were harvested, resuspended in 20 mM Tris pH 7.5, 100 mM NaCl, and lysed by sonication. The lysate was centrifuged for 1 hr at 20,000 x g and the supernatant loaded on a metal chelating affinity column. Fractions containing *Tb*PTP1 were identified by measuring the hydrolysis of p-nitrophenylphosphate (pNPP (12)). Fractions were pooled, loaded on a gel filtration column and eluted in 20 mM Tris pH 7.5, 100 mM NaCl. Recombinant *Tb*PTP1 was concentrated to 10 mg/ml.

In vitro dephosphorylation

NOPP44/46 was amplified from *T. brucei* strain 29.13 (13) genomic DNA and cloned into plew-MHTAP (14) for expression in procyclic form *T. brucei* 29.13. Expression of the tagged protein was induced with tetracycline for 24 h, and the protein was purified using a modified two-step affinity purification TAP protocol with 1 mM sodium orthovanadate in the lysis buffer (15,16). The purified preparation was treated with 10 mM DTT for 15 min to inactivate the sodium orthovanadate. Dephosphorylation reactions were carried out at RT for 15 min in 20 mM Tris pH 7.5, 100 mM NaCl buffer, with varying amounts of *Tb*PTP1. For dephosphorylation of NOPP44/46 with other PTPs, PTP input was normalized to the activity of 100 nM *Tb*PTP1 at a saturating concentration of the noncognate substrate pNPP. Reactions were stopped by the addition of SDS-PAGE loading buffer, separated by SDS-PAGE, and detected by Western blot using the 4G10 anti-pY antibody. Blots were stripped and reprobed with monoclonal anti-NOPP44/46 1D2 (9).

Substrate trapping

*Tb*PTP1 resin was prepared by coupling *Tb*PTP1 to NHS-activated SepharoseTM 4 Fast Flow (GE Healthcare) according to the manufacturer's protocol. Wild-type or the Asp199Ala mutant *Tb*PTP1 was coupled at a concentration of 1 mg/ml, followed by an incubation in 0.1 M Tris blocking buffer. Lysates were prepared from procyclic form *T. brucei* grown for 16 hours in medium containing 1.5 μM sodium orthovanadate. Cells were extracted in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EGTA, 1% Triton X-100) containing 1 mM sodium orthovanadate, complete protease inhibitor cocktail (Roche), and 5 mM iodoacetamide to inhibit endogenous PTP activity. Iodoacetmide and orthovanadate were inactivated by addition of 10 mM DTT. To capture substrates of *Tb*PTP1, 10 μl of wild-type or Asp199Ala *Tb*PTP1 resin was incubated for 2 hours at 4°C with 500 μl lysate corresponding to $0.5x10^9$ *T.*

brucei cells. The resin was washed five times in high salt buffer (20mM HEPES pH 8.0, 2M NaCl, 15% glycerol, 0.5% NP-40), followed by 5 washes alternating guanidine-HCl buffer (20 mM HEPES pH 8.0, 200 mM guanidine-HCl, 15% glycerol, 0.5% NP-40), and low salt buffer (20mM HEPES pH 8.0, 300 mM NaCl, 15% glycerol). The resin was boiled in reducing SDS-PAGE buffer for 15 minutes, run on a 12% Tris-glycine gel, and transferred to nitrocellulose membrane for Western analysis with 4G10 anti-pTyr (GE Healthcare) and anti-NOPP44/46 antibody (9).

Crystallization, structure determination, and structure analysis

Initial crystals were obtained by sitting drop vapor diffusion trials at 18 °C from a 1:1 mixture of *Tb*PTP1 at 10 mg/ml and 10% PEG 3000, 100 mM CHES pH 9.5. Diffraction quality crystals were obtained by hanging drop vapor diffusion after introducing mutations E138A, E139A, and E140A, predicted to reduce the surface entropy of *Tb*PTP1 (17), addition of 10 mM TCEP, and in-drop trypsin cleavage of the His6 tag using a trypsin:*Tb*PTP1 ratio of 1:1,000 (w/w). Crystals were immersed in mother liquor containing 10% glycerol, mounted, and flash frozen in liquid nitrogen.

Diffraction data were collected at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1. Data were reduced using the HKL2000 program suite (18). Phases were obtained by molecular replacement using MolRep (19) and the search model PTPN3 (Protein Data Bank Accession 2B49) modified by CHAINSAW (20). After automated model building in PHENIX (21), the final model was built by alternating manual model building using Coot (22) and maximum likelihood refinement using PHENIX. The R_{free} was determined using a random 5% of the data. The structure was validated using MOLProbity (23). Images were generated in Pymol, and structure comparisons were performed using the DALI server and PDBsum.

A7. Tables

Table A.1. Crystallographic data and refinement statistics for *Tb*PTP1.

Data Collection and Refinement Statistics for *Tb*PTP1	
Crystal symmetry	$P2_12_12_1$
Unit cell	
a, b, c (Å)	76.63, 77.38, 117,2
α, β, γ (^{o})	90, 90, 90
Wavelength (Å)	
Resolution (Å)	2.4
Completeness (%)	99.46
Multiplicity	
Mean I/σI	
Refinement	
Resolution (Å)	47-2.4
Reflections	28,040
R_{work}/R_{free} (%)	20/25
Rms Δ bonds (Å)	0.008
Rms Δ angles (^{o})	1.089
Average B-factor (Å^2)	
Main chain dihedral angles	
Most favored (%)	84.6
Allowed (%)	14
Disallowed (%)	1.4

A8. Figures

Figure A.1. Substrate trapping identifies NOPP44/46 as a major *Tb*PTP1 substrate.

A) Anti-pTyr Western blot after *Tb*PTP1 substrate trapping. The trapping mutant (trap) selectively enriches three major tyrosine-phosphorylated proteins.

B) Anti-NOPP44/46 Western blot of proteins bound to wild type (WT) and trapping mutant.

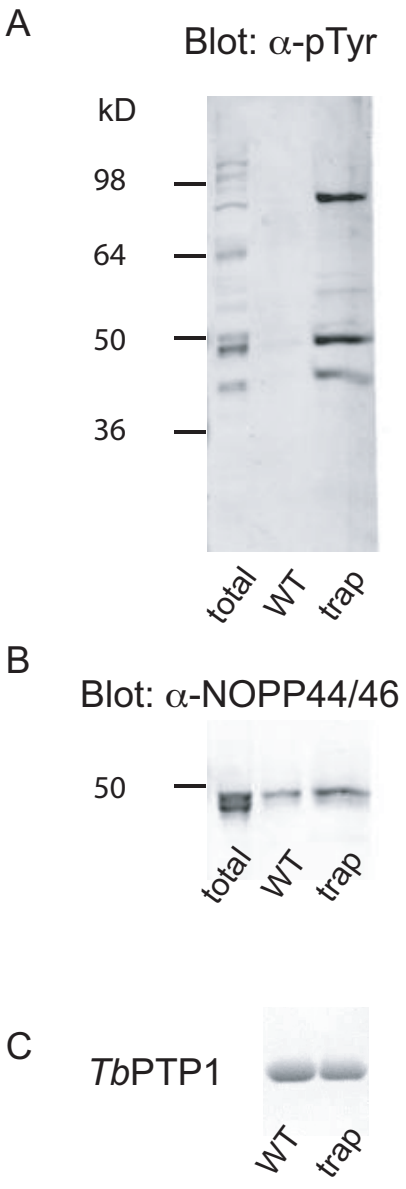


Figure A.2. *Tb*PTP1 efficiently and selectively dephosphorylates NOPP44/46 *in vitro*.

A TAP-tagged allele of NOPP44/46 was expressed in procyclic forms and affinity purified for dephosphorylation reactions.

A) *Tb*PTP1 dephosphorylates NOPP44/46 in a dose-dependent manner.

B) *Tb*PTP1, but not unrelated phosphatases from *Mycobacterium tuberculosis* (PtpA and PtpB), *Staphylococcus aureus* (*Sa*PtpA), and *Listeria monocytogenes* (*lmo*1935) dephosphorylate NOPP44/46.

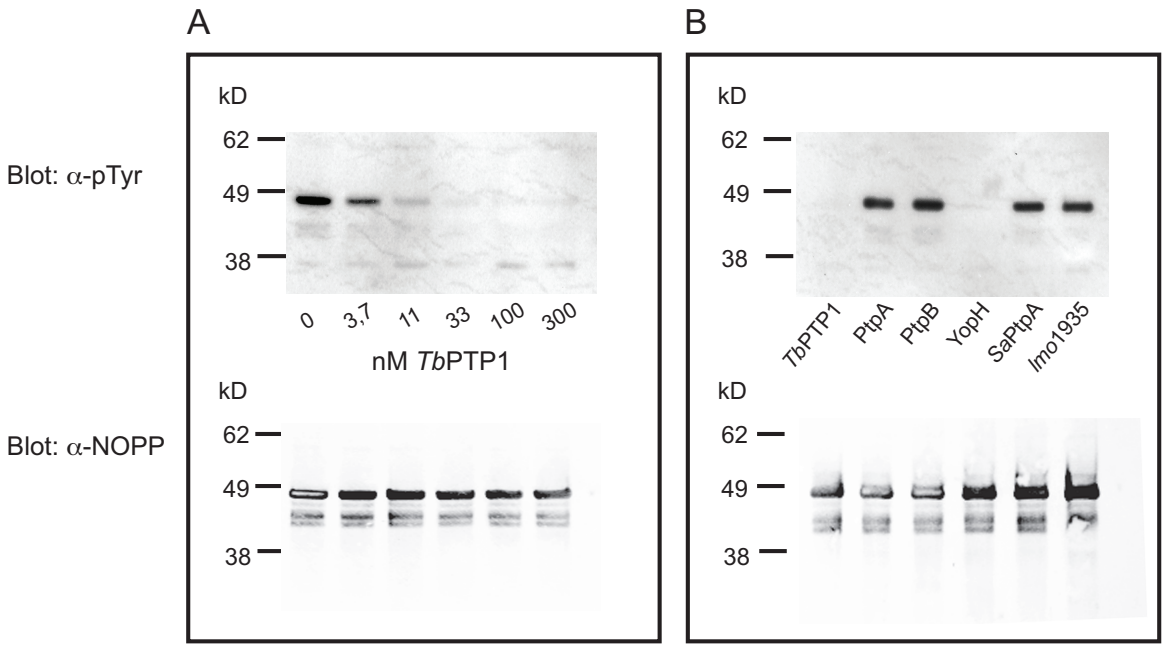


Figure A.3. NOPP44/46 is phosphorylated on Tyr 181.

A) Schematic of NOPP44/46 indicating domain organization and position of the five tyrosine residues. U: Unique region, J: Junction, A: Acidic region, and R: RGG-repeat region.

B) All five NOPP44/46 tyrosines were individually changed to Phe, and the phosphorylation of NOPP44/46 detected by anti-pTyr antibody (upper panel) and anti-NOPP44/46 control Western (lower panel).

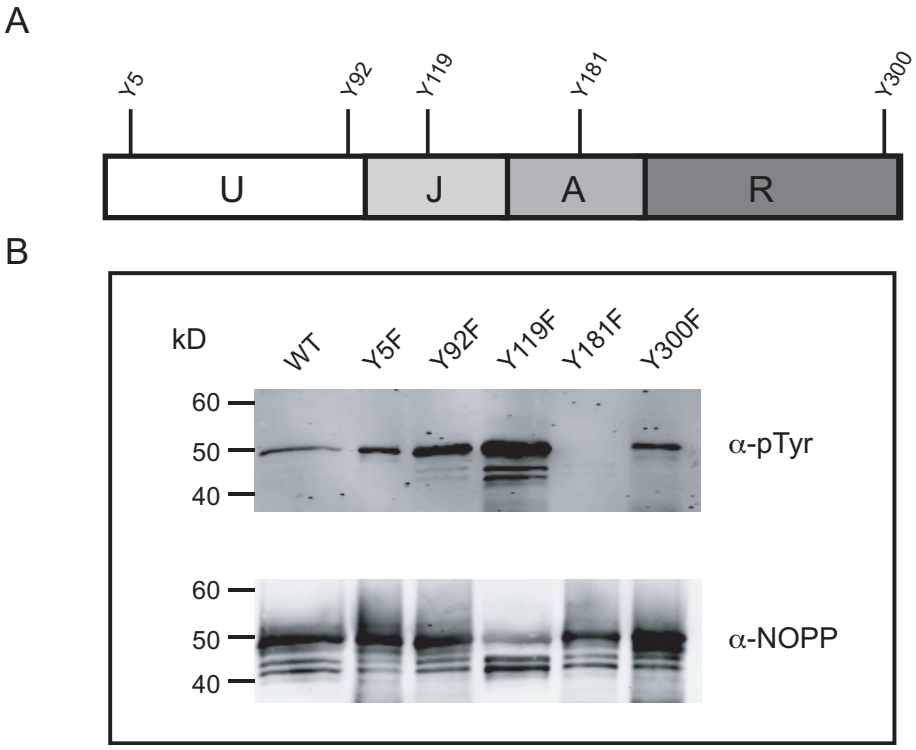


Figure A.4. Overall structure of *Tb*PTP1.

A) *Tb*PTP1 shares the canonical PTP fold. The catalytic motifs P-loop and WPD-loop are highlighted in orange and yellow, respectively. No electron density for residues 65-74 was visible in chain A (dotted line).

B) 2Fo-Fc electron density map of active site showing phosphate (center), contoured at 1.0σ.

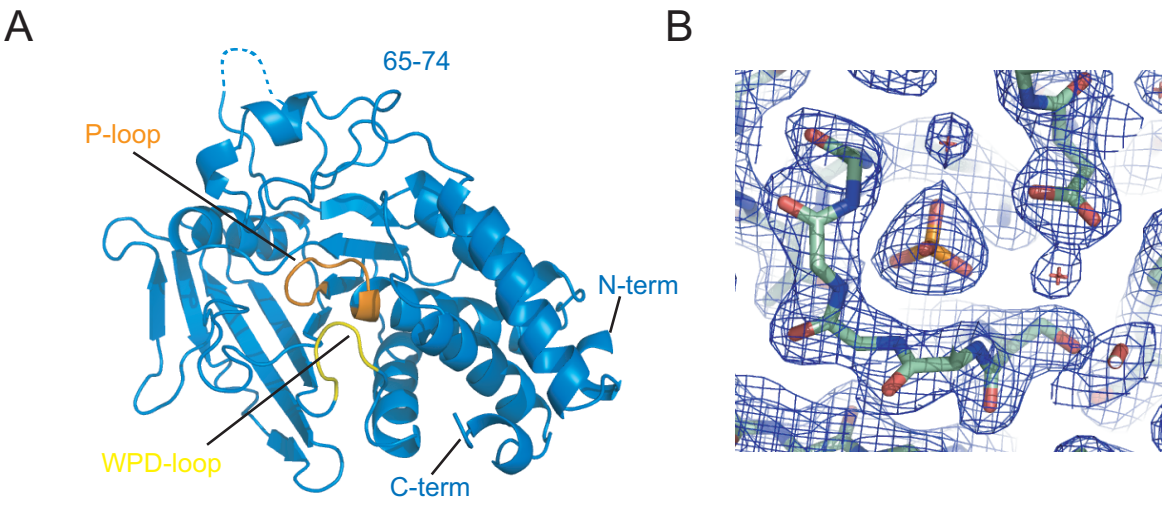


Figure A.5. *Tb*PTP1 has a similar fold to human PTPs.

A) *Trypanosome*-specific sequence motifs (green) map on the surface outside of the active site (P-loop in orange).

B) Superposition of the Cα chain in ribbon representation showing overall strong similarity to human RPTPO.

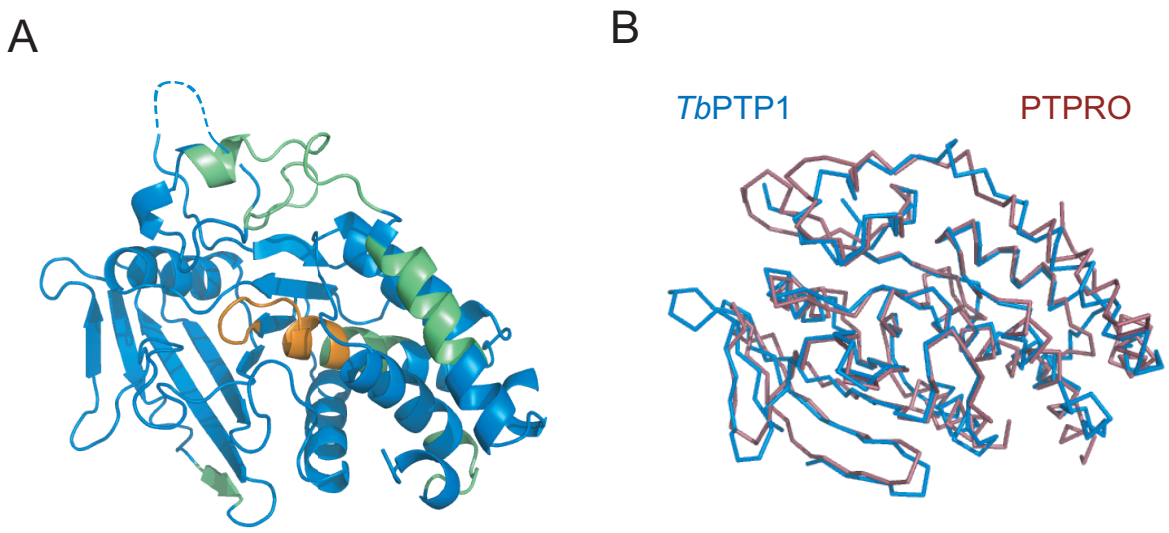
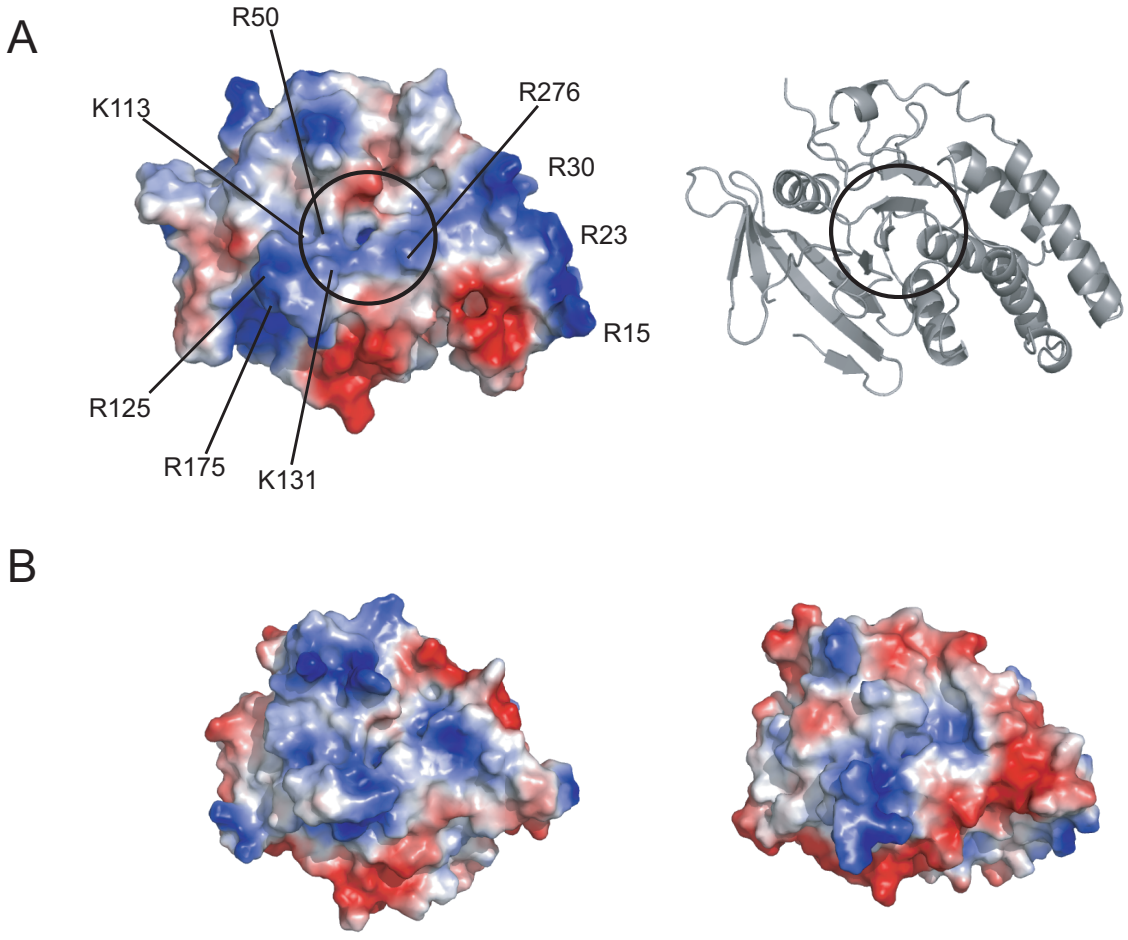


Figure A.6. Electrostatic surface potential of *Tb*PTP1.

(Left) The *Tb*PTP1 surface shows distinct electronegative (red) and positive (blue) regions. The entry to the active site is indicated by the circle.

(Right) Cartoon representation in the same orientation as left panel.



A9. References

1. Fevre, E. M., Wissmann, B. V., Welburn, S. C., and Lutumba, P. (2008) *PLoS Negl Trop Dis* **2**, e333
2. Brun, R., Blum, J., Chappuis, F., and Burri, C. (2009) *Lancet*
3. Fenn, K., and Matthews, K. R. (2007) *Curr Opin Microbiol* **10**, 539-546
4. Parsons, M., Valentine, M., Deans, J., Schieven, G. L., and Ledbetter, J. A. (1991) *Mol Biochem Parasitol* **45**, 241-248
5. Garcia-Salcedo, J. A., Nolan, D. P., Gijon, P., Gomez-Rodriguez, J., and Pays, E. (2002) *Mol Microbiol* **45**, 307-319
6. Li, Z., and Wang, C. C. (2006) *Eukaryot Cell* **5**, 1026-1035
7. Muller, I. B., Domenicali-Pfister, D., Roditi, I., and Vassella, E. (2002) *Mol Biol Cell* **13**, 3787-3799
8. Jensen, B. C., Brekken, D. L., Randall, A. C., Kifer, C. T., and Parsons, M. (2005) *Eukaryot Cell* **4**, 30-35
9. Parsons, M., Ledbetter, J. A., Schieven, G. L., Nel, A. E., and Kanner, S. B. (1994) *Mol Biochem Parasitol* **63**, 69-78
10. Szoor, B., Wilson, J., McElhinney, H., Tabernero, L., and Matthews, K. R. (2006) *J Cell Biol* **175**, 293-303
11. Dean, S., Marchetti, R., Kirk, K., and Matthews, K. R. (2009) *Nature* **459**, 213-217
12. Montalibet, J., Skorey, K. I., and Kennedy, B. P. (2005) *Methods* **35**, 2-8
13. Wirtz, E., Leal, S., Ochatt, C., and Cross, G. A. (1999) *Mol Biochem Parasitol* **99**, 89-101
14. Jensen, B. C., Kifer, C. T., Brekken, D. L., Randall, A. C., Wang, Q., Drees, B. L., and Parsons, M. (2007) *Mol Biochem Parasitol* **151**, 28-40
15. Panigrahi, A. K., Schnaufer, A., Carmean, N., Igo, R. P., Jr., Gygi, S. P., Ernst, N. L., Palazzo, S. S., Weston, D. S., Aebersold, R., Salavati, R., and Stuart, K. D. (2001) *Mol Cell Biol* **21**, 6833-6840
16. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) *Nat Biotechnol* **17**, 1030-1032
17. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007) *Protein Sci* **16**, 1569-1576
18. Otwinowski, Z., and Minor, W. (1997). in *Methods in Enzymology*
19. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763
20. Stein, N. (2008) *J. Appl. Cryst.* **41**, 641-643
21. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954
22. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
23. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21

24. Flint, A. J., Tiganis, T., Barford, D., and Tonks, N. K. (1997) *Proc Natl Acad Sci U S A* **94**, 1680-1685
25. Zhang, Z. Y., Clemens, J. C., Schubert, H. L., Stuckey, J. A., Fischer, M. W., Hume, D. M., Saper, M. A., and Dixon, J. E. (1992) *J Biol Chem* **267**, 23759-23766
26. Barford, D., Das, A. K., and Egloff, M. P. (1998) *Annu Rev Biophys Biomol Struct* **27**, 133-164
27. Salmeen, A., Andersen, J. N., Myers, M. P., Tonks, N. K., and Barford, D. (2000) *Mol Cell* **6**, 1401-1412
28. Barr, A. J., Ugochukwu, E., Lee, W. H., King, O. N., Filippakopoulos, P., Alfano, I., Savitsky, P., Burgess-Brown, N. A., Muller, S., and Knapp, S. (2009) *Cell* **136**, 352-363
29. Andersen, J. N., Mortensen, O. H., Peters, G. H., Drake, P. G., Iversen, L. F., Olsen, O. H., Jansen, P. G., Andersen, H. S., Tonks, N. K., and Moller, N. P. (2001) *Mol Cell Biol* **21**, 7117-7136
30. White, N. J. (2008) *Malar J* **7 Suppl 1**, S8
31. Nett, I. R., Davidson, L., Lamont, D., and Ferguson, M. A. (2009) *Eukaryot Cell* **8**, 617-626
32. Das, A., Peterson, G. C., Kanner, S. B., Frevert, U., and Parsons, M. (1996) *J Biol Chem* **271**, 15675-15681
33. Hellman, K., Prohaska, K., and Williams, N. (2007) *Eukaryot Cell* **6**, 2206-2213
34. Park, J. H., Brekken, D. L., Randall, A. C., and Parsons, M. (2002) *Mol Biochem Parasitol* **119**, 97-106
35. Park, J. H., Jensen, B. C., Kifer, C. T., and Parsons, M. (2001) *J Cell Sci* **114**, 173-185
36. Allende, M. L., Amsterdam, A., Becker, T., Kawakami, K., Gaiano, N., and Hopkins, N. (1996) *Genes Dev* **10**, 3141-3155
37. Kinoshita, Y., Jarell, A. D., Flaman, J. M., Foltz, G., Schuster, J., Sopher, B. L., Irvin, D. K., Kanning, K., Kornblum, H. I., Nelson, P. S., Hieter, P., and Morrison, R. S. (2001) *J Biol Chem* **276**, 6656-6665
38. Lerch-Gaggl, A., Haque, J., Li, J., Ning, G., Traktman, P., and Duncan, S. A. (2002) *J Biol Chem* **277**, 45347-45355
39. Maiorana, A., Tu, X., Cheng, G., and Baserga, R. (2004) *Oncogene* **23**, 7116-7124
40. Du, Y. C., and Stillman, B. (2002) *Cell* **109**, 835-848
41. Das, A., Park, J. H., Hagen, C. B., and Parsons, M. (1998) *J Cell Sci* **111 (Pt 17)**, 2615-2623
42. Inder, K. L., Lau, C., Loo, D., Chaudhary, N., Goodall, A., Martin, S., Jones, A., van der Hoeven, D., Parton, R. G., Hill, M. M., and Hancock, J. F. (2009) *J Biol Chem* **284**, 28410-28419
43. Okuda, M., Horn, H. F., Tarapore, P., Tokuyama, Y., Smulian, A. G., Chan, P. K., Knudsen, E. S., Hofmann, I. A., Snyder, J. D., Bove, K. E., and Fukasawa, K. (2000) *Cell* **103**, 127-140