Allosteric Regulation of Promiscuous Surfaces

by

Kristopher M Kuchenbecker

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements:

I would not have discovered my joy in science without great encouragement and guidance from family. Dad, Mom, Katherine, and Karalyn: Thank you for providing a wonderful environment for my scientific journey.

I would be remiss to not mention Dr. Robert Glaeser and his inspiring introduction to protein structure in a lecture for upper division biochemistry at UC Berkeley. That is the moment at which I decided to pursue graduate studies in biophysics. I thank Dr. David Jablons for providing my first research opportunity in his lab at the UCSF Cancer Center.

In graduate school, Rebecca Brown, has been instrumental in assuring the smooth running of the iPQB and Biophysics program. My lab neighborhood, the Macromolecular Structure Group, is a great community. I am thankful to the members of the Agard, Stroud, and Cooke labs for many wonderful discussions and the sharing of equipment and reagents.

My thesis committee has been a wonderful sounding board for my ideas and their guidance has been critical to my success. I am thankful to Dr. Holly Ingraham for helping me find words that translate my biophysical questions to cellular function. Dr. Matt Jacobson is the person I turn to whenever I am uncertain about a physical phenomenon.

Dr. James Fraser and I began working together in April of 2011. His expertise in ambient temperature crystallography inspired the structural aspect of my thesis project. His contagious enthusiasm is a source of continued motivation.

The Fletterick lab was a wonderful environment to foster my scientific development. Peter Hwang has introduced me to many techniques. His teachings form the foundation of my knowledge in Xray crystallography and surface plasmon resonance. I am very thankful for his willingness to run overnight shifts at ALS with me. Jeremy Wilbur, Sam Pfaff, and John Bruning

have been the initial filter for many of my scientific endeavors. Their discussion and insight have helped to shape many of the central themes in this work.

Dr. Robert Fletterick has been a wonderful advisor. During my graduate work, his support and encouragement of my numerous collaborations were indispensable. His office was always open to me, and every conversation we shared left me with a profoundly new view of the world.

Collaborative Works:

I am extremely fortunate to have had the opportunity to work with classmates in the iPQB, CCB, and Tetrad graduate programs at UCSF. The cooperative discovery process fascinates me, and I am grateful that our scientific paths have crossed. A foreword presenting the underpinning scientific work that enabled these studies precedes the final published versions of these projects, which are included as separate sections.

To all involved in the completion of these projects: Thank you for allowing my contributions. To Tanja Kortemme, Geeta Narlikar, and Keith Yamamoto: Your students have been excellent teachers. To Daniele Canzio, Ryan Ritterson, and Lisa Watson: Our work together accounts for many of my fondest memories of graduate school. There are no words to adequately thank you for your gracious sharing of scientific discovery.

Chapter 3 -- The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals.
Nature Structural & Molecular Biology. April 9, 2013. (DOI: 10.1038/nsmb.2595)

Appendix 4 -- Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly.
Molecular Cell. January 7, 2011. (DOI: 10.1016/j.molcel.2010.12.016)

Appendix 5 -- Design of a Photoswitchable Cadherin
Journal of the American Chemical Society. August 7, 2013. (DOI: 10.1021/ja404992r)

Abstract

Allosteric Regulation of Promiscuous Surfaces

Kristopher M Kuchenbecker

Every signaling event in biology begins with a molecular recognition process. In order for information to directly pass between two molecules, they must come in contact with each other for the requisite time to convey a given message. Specialized pathways often involve cascades of high fidelity interactions, near deterministically arriving at the same outcome for a given input. Other signals are capable of exerting disparate effects depending on the cellular context. Promiscuity in molecular recognition plays a central role in the capacity to produce different responses to the same signal.

For the steroid hormone receptors, the input signal is a small molecule ligand. Upon binding of the hormone, these molecules move to the nucleus, localize to different genomic response elements, and regulate the transcription levels of many different genes by recruitment of transcriptional machinery. The abilities to recognize multiple DNA sequences and to recruit different coregulators require that these interaction surfaces adapt to and engage with diverse structures over a relatively narrow spectrum of interaction energies.

The essential functions of DNA recognition and coregulator binding were investigated using dynamic structural analyses and biophysical measurements of interaction energies. For both, the manifestations of structural remodeling and energetic perturbation at sites removed from the interaction surfaces serve as evidence for allosteric regulation. In DNA recognition, we learn that nucleotide sequence modulates the structure of the dimerization surface of the glucocorticoid receptor, affecting both the cooperativity of complex formation and the kinetics of dissociation. In coregulator binding to the androgen receptor, we find distant structural remodeling and energetic coupling to a site of post translational modification. Together, these allosteric mechanisms help explain functional discrimination in a background of energetic degeneracy.

Table of Contents:

Chapter 1 _____ 1
Promiscuity in nuclear receptor signaling

Chapter 2 _____ 7
Surface plasmon resonance measurement of molecular recognition processes

Chapter 3 _____ 33
DNA sequence modulates dimerization energetics for the androgen and glucocorticoid receptor DNA binding domains

Chapter 4 _____ 91
Structural consequences of binding events on the androgen receptor ligand binding domain

Conclusion _____ 154
Allosteric controls of molecular recognition

Appendix 1 _____ 159
Ambient temperature datasets of the androgen receptor ligand binding domain

Appendix 2 _____ 169
Software for the parameterization of steady-state binding data

Appendix 3 _____ 178
Software for the analysis of ensemble models of protein structure

Appendix 4 _____ 189
Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly.

Appendix 5 _____ 270
Design of a photoswitchable cadherin

List of Tables:

Chapter 1

Three binding processes of nuclear receptors ____ 4

Chapter 2

Kinetic constants for individual 1:1 fits ____ 30

Chapter 3

DNA Recognition Properties of the Androgen Receptor

DNA sequences used in the study of androgen receptor DNA binding ____ 36
Steady state parameters for androgen receptor-DNA interactions ____ 38

DNA Recognition Properties of the Glucocorticoid Receptor

SPR fit parameters for WT and A477T DBD ____ 82
Correlation between transcriptional activity and binding parameters ____ 83

Chapter 4

HBP Ligand Affects Coregulator Binding

Coregulator peptide sequences ____ 95

Ambient temperature structures of the androgen receptor

Structures for comparison of NTD bound states ____ 114
Structures for comparison of coregulators with the DHT liganded receptor ____ 117

Ensemble perspective of structure

Ensemble refinement statistics ____ 125

Appendix 1

Ambient Temperature Datasets of the androgen receptor ligand binding domain

Sequences of coregulator fragments ____ 161
Ambient temperature datasets ____ 163→168

Appendix 2

Software for the Parameterization of Steady-State Binding Data

Parameterization and errors from individual and global fitting ____ 174

Appendix 5

Design of a Photoswitchable Cadherin

Characterization of library mutants ____ 295

List of Matlab Functions:

Appendix 2

Software for the Parameterization of Steady-State Binding Data

Langmuir Model Bootstrap: **LANGstrap** ____ **175**
Hill Equation Bootstrap: **HILLstrap** ____ **176**
Adsorption Energy Distribution: **AEDist_solver** ____ **177**

Appendix 3

Software for the Analysis of Ensemble Models of Protein Structure

Import ensemble model / create dihedral & order parameter variables ____ **181**
Calculate Phi distribution: **kk_PhiDist_getter** ____ **182**
Calculate Psi distribution: **kk_PsiDist_getter** ____ **183**
Calculate Chi1 distribution: **kk_Chi1Dist_getter** ____ **184**
Calculate Chi2 distribution: **kk_Chi2Dist_getter** ____ **185**
Calculate Chi3 distribution: **kk_Chi3Dist_getter** ____ **186**
Calculate Chi4 distribution: **kk_Chi4Dist_getter** ____ **187**
Bootstrap order parameters: **kk_OrderParam_STRAP** ____ **188**

Presentations:

Principles of the SPR measurement 9

List of Figures:

Chapter 1

High fidelity interactions vs. promiscuous surfaces 3
Stages of nuclear receptor signaling 3
Interactions of the androgen receptor 4

Chapter 2

Raw traces for 2011 Benchmark Study 25
Kinetic fits with a 1:1 binding model 26
Simple exponential fits reveal two processes 27
Dissociation reveals two bound states 27
Scatchard plots are revelatory (pt 1) 28
Steady state parameterization 29

Chapter 3

DNA Recognition Properties of the Androgen Receptor

Differential cooperativity suggests allostery 34
Experimental system for DNA binding studies 36
Raw traces of DNA recognition 37
Parameter estimation for cooperative DNA binding 38
Kinetic fits of dimer assembly on the DNA 40
Kinetics of differential cooperativity 41

DNA Recognition Properties of the Glucocorticoid Receptor

Non-specific GBS bases modulate GR structure and activity 64
GBS spacer affects the conformation of the D-loop 65
Structural consequences of disrupting the dimerization surface 66
A477T impairs dimerization but not monomer DNA binding 67
A477T disrupts cooperativity and GBS-specific dissociation 68
Sequence-specific conformation requires intact dimerization surface 69
NMR assignment of DNA-bound GR-DBD complex 75
Chemical shift difference analysis of GBSs differing in half-site/spacer 76
DNA-bound A477T conformation differs from unbound WT DBD 77
Chemical shift difference analysis of GBSs with different transcription levels 78
Conformational exchange within lever arm and dimer surface 79
Chemical shift variance across the different GBS complexes 80
DNA bending by FRET 81

Chapter 4

HBP Ligand Affects Coregulator Binding

Experimental system for ligand comparison 96
Raw coregulator binding traces for DHT, TES, & R1881 liganded receptors 97→99
Dissociation fits for GSN, NTD, and PAK 100
Energetic and kinetic parameters for coregulator binding 102
Ambient Temperature Structures of the Androgen Receptor
Ringer correlation heatmaps for DHT, TES, & R1881 structures 108→113
Coregulator binding remodels the HBP 115
NTD binding collapses the structural ensemble of the HBP 116
Two classes of coregulator binding 118
Coregulator binding impacts H10 bend 119
Comparison of H10 bend for NTD and PAK 120
Structural heterogeneity of Loop9/10 121
Ensemble Perspective of Structure
Ensemble structural perturbations 124
Ensemble processing by order parameter and correlation coefficients 126
Loop9/10 Integrates Coregulator Identity
Comparison of the cryo and ambient ensembles 127
Order parameter difference vectors for cryo vs. ambient ensembles 128
Redefinition of the coregulator binding surface 129
Correlation coefficient of cryo vs. ambient ensembles 131
SPR measurement of NTD and PAK binding to the DHT liganded receptor 132
Ensemble binding modes of NTD & PAK 135
Coregulator perturbations relative to the apo model 136
Order parameter difference vectors for NTD & PAK 137
Shared perturbations to the ensemble 138
Coregulator specific perturbations to the ensemble 141
Raw isotherm data for comparison of L9/10 mutants 147
L9/10 mutation differentially affects coregulator binding 148
Alternative fitting strategies for comparison of L9 mutants 150
Additional L9/10 mutations 150
Representative SPR traces for NTD & PAK binding 151

Appendix 1
Ambient Temperature Datasets of the androgen receptor ligand binding domain
Representative crystal forms and diffraction patterns 160

Appendix 2
Software for the Parameterization of Steady-State Binding Data
Raw traces and isotherm scaling 172
Steady-state fitting with different models 173
Associated errors by model and flow cell 174

Appendix 3
Software for the Analysis of Ensemble Models of Protein Structure

Workflow for Matlab processing of ensemble refined structures 179
Structure variables for processing of ensemble models 180

Appendix 4

Chromodomain-mediated oligomerization of HP1 suggests a nucleosome bridging mechanism for heterochromatin assembly

HP1 protein undergoes four binding processes 190
Protein purity controls data quality 191
Scatchard plots are revelatory (pt 2) 192
Two-site BET adsorption model 193
Swi6 recognizes the H3K9 methyl mark 232
Swi6 forms distinct oligomeric states in the absence of chromatin 233
Swi6 displays lower specificity for the H3K9me^3 mark in mononucleosomes 234
The core unit of Swi6 binding to mononucleosome is a tetramer 235
Amplification of Swi6 specificity occurs on nucleosome arrays 236
The chromodomain contains the Swi6 tetramerization interface 237
Increased tetramerization of Swi6 translates into increased silencing 238
Raw data and analysis of the surface plasmon resonance assay 250
Loss of higher-order oligomerization in L315D Swi6 mutant 251
Nucleosome binding affinity by gel mobility shift assay 252
Details of two analysis methods for mononucleosome:Swi6 AUC data 253
Quality control for nucleosome arrays 254
Swi6 deletion analysis 255
Fr A cassette silencing phenotype 256

Appendix 5

Design of a photoswitchable cadherin

Two-dimensional concentration dependence of cadherin dimerization 271
Reduced model of calcium dependent cadherin dimerization 272
Numerical solution of cadherin species 273
Schematic of the SPR experiment 273
Model of SPR binding surface 274
Cartoon showing basis of design 288
Photoswitchable calcium binding affinity 289
Photoswitchable dimerization binding affinity 290
Computational methodology to create mutant library 296
ESI-MassSpec of biotinylated and conjugated cadherin protein 297
Extent of photoswitching is titratable 298
Calcium affinity as measured by mass spectrometry 299
Representative mass spectra showing calcium binding 300
Illumination and relaxation of photoswitchable protein 301
Single exponential fit to relaxation of photoswitch 302
SPR measurement of calcium titrations 303
Raw traces from the SPR experiment 304
Bootstrap analysis of fit parameters for calcium titration 305

Top 100 parameter combinations by bootstrap analysis **306**
Size-exclusion chromatography of cis and trans X-EC12 **307**
Dissociation of cadherin dimers reveals a single process **308**

Chapter 1:

Promiscuity in nuclear receptor signaling

The act of two molecules coming together to form a functional complex is stochastic. Beginning with the semi-random diffusion of a molecule through the crowded environment of the cytosol and culminating with the act of dissociation, every molecular recognition event can be parsed into a series of coin flips.

The quantitative description of a molecular interaction captures two physical characteristics. The probability of finding two molecules in complex is a function of the absolute concentrations of the constituents. The concentration dependence is a function of the energy difference between the free and bound states. It is often reported as the concentration of molecules at which we expect 50% to be in the bound state ($K_{1/2}$).

The amount of time that molecules remain in complex is independent of concentration. This rate of dissociation is a function of the energy barrier that separates the bound and free states. Similar to the concentration dependent descriptor, it is often useful to report the time dependence as the amount of time required for a population of bound molecules to decrease by half ($t_{1/2}$). Together, the $K_{1/2}$ and the $t_{1/2}$ enable a quantitative description of how molecules behave in solution.

The number of protein-protein interactions in humans is estimated at 650,000, representing only 0.2% of all possible pairwise interactions.[1] For a binary, orthogonal protein-protein interaction (A interacts with B and only B), the recognition surfaces are optimized for high specificity. Energetic gains can be found by increasing contact area and furthering electrostatic complementarity. As a surface with the sole purpose of recognizing only one other molecule, the evolved specificity effectively lowers the $K_{1/2}$ to a point where all other interactions are rendered moot. For promiscuous surfaces, there is a trade-off in the energetics of interaction with the ability to engage many different interaction partners.[2] Simply, if a surface

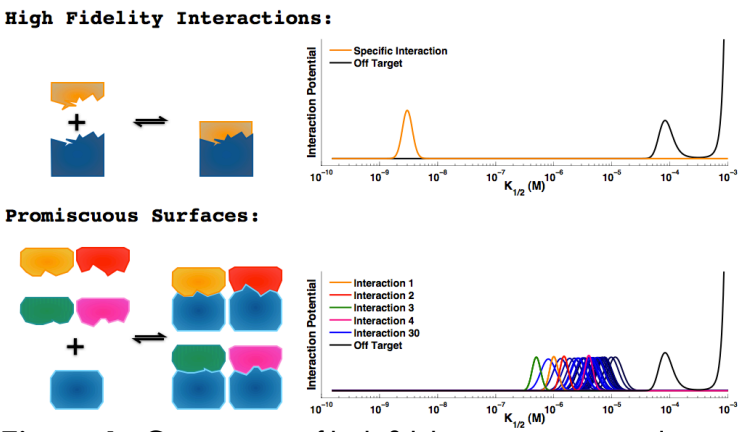


Figure 1. Comparison of high fidelity interactions and promiscuous surfaces. For molecules with single interaction partners, elaborate complementarity between binding surfaces can result in a high affinity interaction. For binding surfaces that engage multiple partners, binding energy is compromised in exchange for structural plasticity.

functions in the recognition of ten different molecules, there is less room for elaboration as equivalent energetic gains across all complexes are unlikely. This constraint leads to a crowded and dynamic interaction spectrum for promiscuous surfaces.

While high fidelity interaction pairs are rare in biology, extreme promiscuity is an essential feature of transcription factor function. Possessing no catalytic activity, nuclear receptors (NRs) respond to small molecule ligands and drastically alter cell fate by regulating the transcription levels of specific genes. The foundation of this function can be divided into three different binding processes: small molecules, proteins, and DNA. At the first step in the activation of NR signaling, the small molecule hormone binds and changes the receptor to an activated conformation. Protein-protein interactions are a near constant for NRs during all stages of signaling as they recruit multiple enzymatic functions to promote gene activation.[3] DNA recognition serves to position the NR at specific response elements upstream of target genes.

These three binding processes are well separated in both energetics ($K_{1/2}$) and kinetics ($t_{1/2}$). Forming the structural core of the C-terminal ligand binding domain (LBD), the binding of hormone

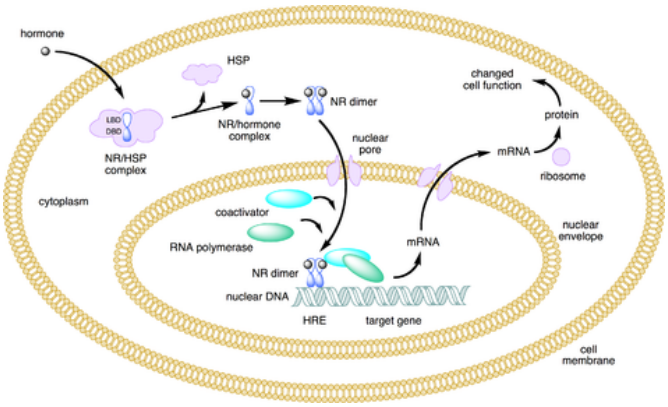


Figure 2. Stages of nuclear receptor signaling.[10]

Molecule	Binding Process	Possible Interactants	Affinity Range ($K_{1/2}$)	Kinetics ($t_{1/2}$)
small molecule hormones	hydrophobic ligands bind buried site in LBD	1-3 natural ligands (ie- DHT, TES)	100 pM ↓ 600 pM	minutes
DNA	inverted repeats of 6bp with 3 bp spacer dimer of NRs bind cooperatively	10's of unique sequences 100s of genomic regions	1 nM ↓ 40nM	10's of seconds
Protein	NR boxes (φxxφφ) helical segments bind shallow groove of LBD	100's of unique motifs 100's of coregulator proteins	300 nM ↓ 30 μM	~1 second

Table 1. The three binding processes of nuclear receptors.

helps fold the receptor and dissociation is very slow ($t_{1/2}$ on the order of minutes).[4] DNA recognition is strongly cooperative and the DNA bound NR complex can persist for 10s of seconds.[5] Protein-protein interactions mediated by the NR LBD are comparatively weak and short lived with $t_{1/2}$'s on the order of 1 second. Because of this kinetic scaling across the three binding processes of the NR, a single hormone binding event can create an activated receptor capable of interacting with many different genomic response elements; in turn, each DNA binding event anchors the receptor for multiple protein-protein interactions. This temporal separation of ligand binding, DNA recognition, and coregulator recruitment motivates the independent study of these processes.

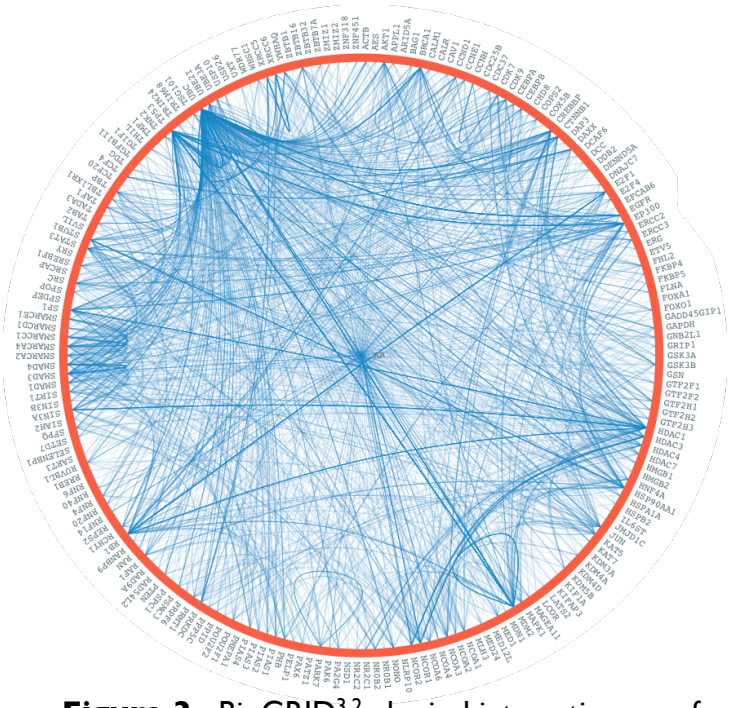


Figure 3. BioGRID[3.2] physical interaction map for the androgen receptor.[7]

While there are only a handful of natural ligands for most NRs, the DNA binding and protein

recognition functions engage multiple partners. Returning to the energetic discussion on specificity and promiscuity, it follows that as the binding processes decrease in energy, the number of interactants increases. For DNA binding, there are over 1,700 androgen responsive genes, consisting of tens of different response element sequences.[6] Documented protein-protein interactions for the androgen receptor provides a lower estimate of 203 unique interaction partners.[7]

For the different hormone ligands that bind to the androgen receptor, different transcriptional outcomes have been attributed to changes in coregulator recruitment.[8,9] Our understanding of this process is principally based on whole cell measurements. In short, we are able to measure receptor function, but not how function affects the receptor. The picture provided does not inform the detailed mechanisms by which the receptor navigates its numerous, energetically-equivalent interaction partners. From the perspective of the receptor, are different DNA or coregulator binding events equivalent? Or, are there dynamic mechanisms at work that inform the receptor of its current interaction profile?

Returning to the stochastic picture of molecular recognition, the energetics of DNA recognition and coregulator binding suggest a near random selection of interaction partners, which is inconsistent with the observation of ligand specific transcriptional responses. This work offers kinetic insight to these binding processes, providing an additional axis for discrimination of binding partners. Further, structural studies herein show how domain structure of the receptor is coupled to identity of interaction partners.

References:

1. Stumpf, M. P. H. *et al.* Estimating the size of the human interactome. *Proc Natl Acad Sci USA* **105**, 6959–6964 (2008).
2. Beltrao, P. & Serrano, L. Specificity and Evolvability in Eukaryotic Protein Interaction Networks. *PLoS Comput Biol* **3**, e25 (2007).
3. Olefsky, J. M. Nuclear Receptor Minireview Series. *Journal of Biological Chemistry* **276**, 36863–36864 (2001).
4. Rich, R. L. *et al.* Kinetic analysis of estrogen receptor/ligand interactions. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8562–8567 (2002).
5. Watson, L. C. *et al.* The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nature Structural & Molecular Biology* **20**, 876–883 (2013).
6. Jiang, M. *et al.* Androgen-Responsive Gene Database: Integrated Knowledge on Androgen-Responsive Genes. *Molecular Endocrinology* **23**, 1927–1933 (2009).
7. Stark, C. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
8. Norris, J. D. *et al.* Differential Presentation of Protein Interaction Surfaces on the Androgen Receptor Defines the Pharmacological Actions of Bound Ligands. *Chemistry & Biology* **16**, 452–460 (2009).
9. Baek, S. H. *et al.* Ligand-specific allosteric regulation of coactivator functions of androgen receptor in prostate cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3100–3105 (2006).
10. http://en.wikipedia.org/wiki/File:Nuclear_receptor_action.png

Chapter 2:

Surface Plasmon Resonance Measurement of Molecular Recognition Processes

Overview of surface plasmon resonance:

When I began my graduate studies in the Fletterick Lab, a brand new Biacore T100 instrument had just arrived. This flagship product for surface plasmon resonance (SPR) measurement represented the state of the art in biophysical characterization of the thermodynamics and kinetics of molecular recognition processes. For the study of nuclear receptor function, I can not imagine a more well suited instrument.

Many protein families are suited for classical biochemical techniques that monitor enzymatic activity by measurement of substrate turnover. Nuclear receptors possess no intrinsic catalytic activity. Instead, their functional activity is in transient molecular association. Because of the difficulties in monitoring binding events, our study of their function has largely relied on cell based transcriptional activation studies. These measures of gene expression only inform the output of massively complex signaling systems and molecular details of function are largely inaccessible. The Biacore T100 represented a window into the black box of nuclear receptor function by precise and sensitive label-free observation of receptor binding events.

While my use of SPR has always been aimed at furthering our understanding of biological processes, the ability to make these measurements requires significant technical expertise. This knowledge was hard-fought through countless hours on my own research projects as well as offering guidance to others at UCSF. I have included a lecture on SPR that I presented for a mini-course at UCSF. This section of my thesis is intended to serve as a reference guide for future users.

In 2011, I was invited to participate in a benchmark study for SPR. My submitted writeup for this project is included as an example of identifying and interpreting non-ideality in SPR measurement.

Goals For This Lecture

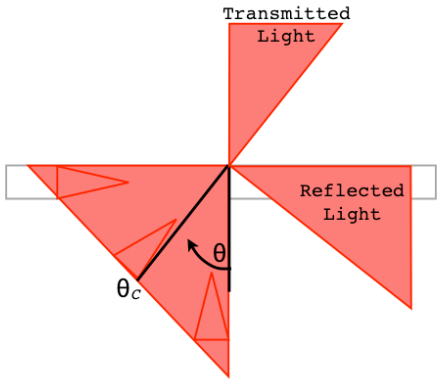
Principles of the SPR Measurement
Value of Dissociation Constant
Identify Experimental Errors
SPR in the Literature
Machines at UCSF

MAJOR GOAL:
Ability to appraise / judge SPR data

Light at a Surface:

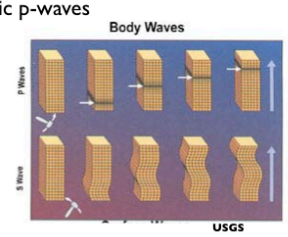
Angle of incidence, θ
- Measured with respect to normal
- At low θ, light transmits

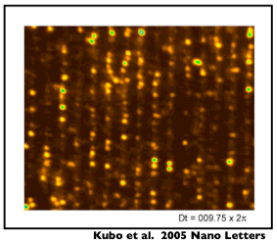
Critical Angle, θ_c
- Angle at which light reflects



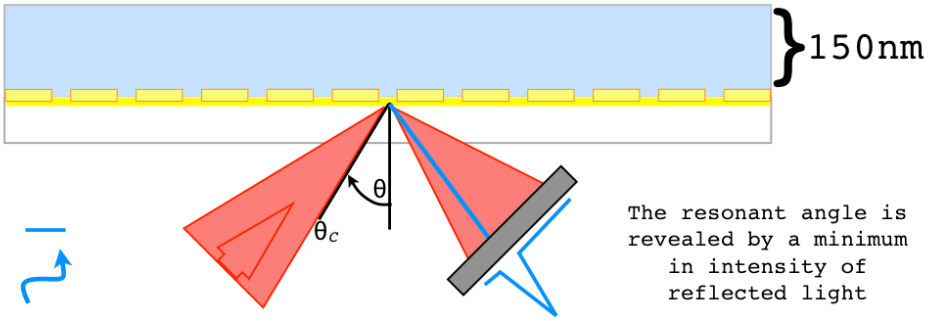
Plasmons:

Collective charge density fluctuations
Electronic p-waves





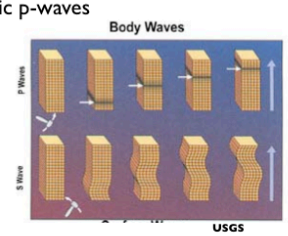
Kubo et al. 2005 Nano Letters

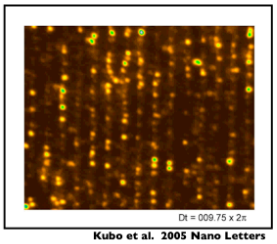


The resonant angle is revealed by a minimum in intensity of reflected light

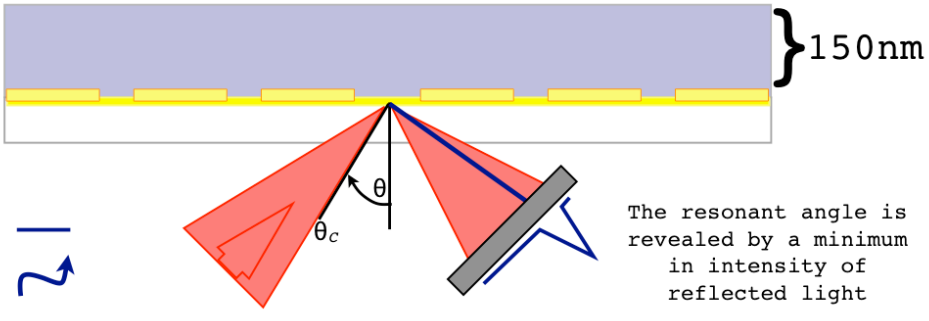
Plasmons:

Collective charge density fluctuations
Electronic p-waves





Kubo et al. 2005 Nano Letters



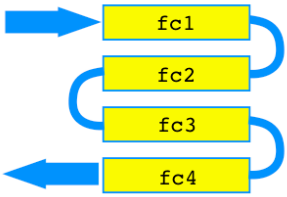
The resonant angle is revealed by a minimum in intensity of reflected light

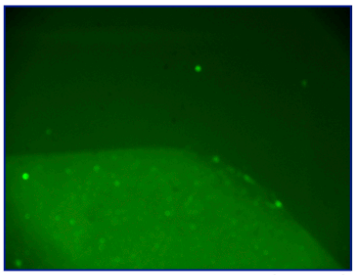
Overview of SPR Principles:

- Measure reflectance as a function of angle
- In resonance, incident light couples to the SPW
  - Evanescent wave probes refractive index
  - Resonant SPW is selected by refractive index
- Measure the resonant angle
  - 1 RU = 10^{-4} degrees ~ Δn of 6e-7
  - ~ 1 pg mm^{-2}
- Truly Label Free
  - Watch all species
- Bulk measurement of single molecule behavior
  - ~10^{10} molecules
  - Independent lattice sites

The SPR Instrument:

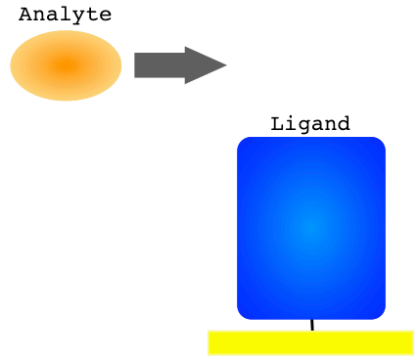
- BiaCore T100
  - $350k Flagship Instrument
  - T200
- Four Flow Cells
  - 1.5 mm^2, 60nL
  - Serial Reference and Sample
- Time Breakdown
  - ~ 1 Month Assay Development
  - > 96 Samples in 24 hrs
- Cost of Consumables
  - ~$300 for SPR Chip
  - Buffer / Tubes / Plates
  - Very Low Consumption of Samples





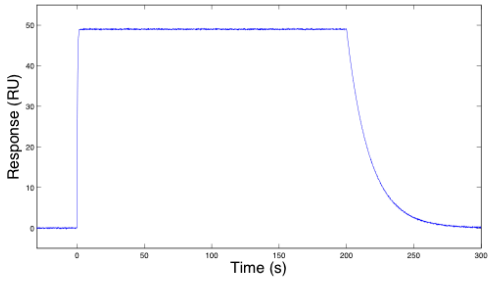
VM (Agard Lab)

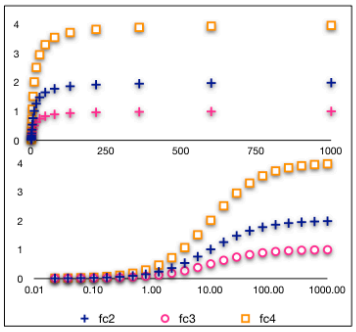
Jargon:

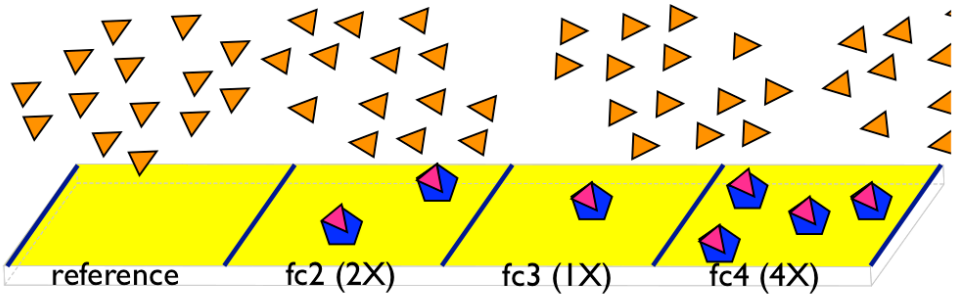


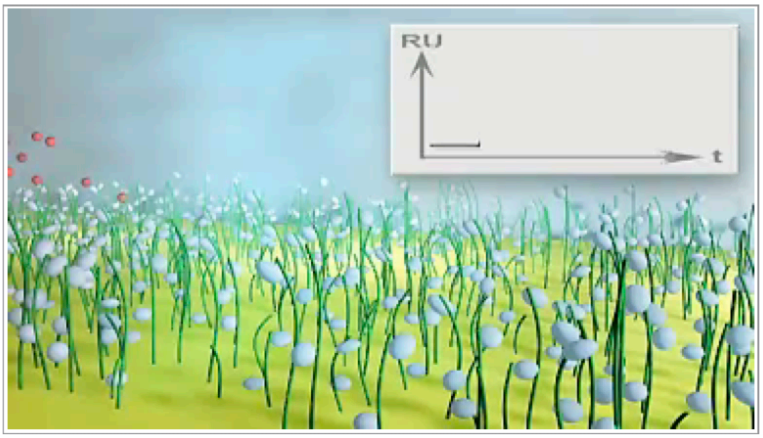
- Terminology
  - Ligand: Immobilized Molecule
  - Analyte: Injected Sample
- Instrument Parameters
  - Temperature
    - 4C to 40C
  - Flow Rate
    - 1μL min^{-1} to 100μL min^{-1}
- Experimental
  - RI: Refractive Index
    - n_{water} ~ 1.33, n_{prot} ~ 1.5, n_{DNA} ~ 1.6
  - Serial Dilutions
    - Equivalent Spacing in Concentration
  - Steady State
    - Equilibrium Measurement of Occupancy
  - Kinetics
    - Curve fitting of on- / off-rates

The SPR Experiment:









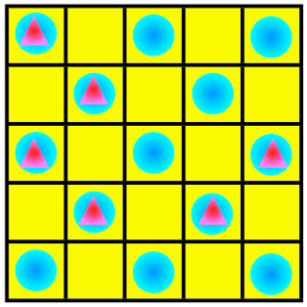
http://www.sandelinanimation.com/projects_biacore.html

Label-Free Measurements:

- The most prevalent methods for determining affinity:
  - Fluorescence Polarization
    - Must be mindful of probe concentration
    - Fluorescent molecules are sticky
  - Gel Shift
    - Not at equilibrium - Leads to underestimation
- Label-Free is industry keyword to distinguish community
  - Isothermal Titration Calorimetry
  - Analytical Ultra Centrifugation
  - Surface Plasmon Resonance

Label-Free Measurements:

- Key Assumptions of Surface Plasmon Resonance
  - All Lattice Sites are Equivalent
  - Each Site Can Hold at Most One Molecule
  - No Interactions Between Adjacent Molecules
  - Concentration of Solution is not Depleted by Interaction



Typical SPR Experiment Has About 10^{10} Immobilized Molecules
At 1μM and flow rate of 30μL min^{-1} provides ~2000 fold analyte:ligand molecules

Strategies for Immobilization:

- Covalent
  - Amine Coupling via EDC/NHS Chemistry
  - Thiol-Coupling via Maleimide Chemistry
- Temporary
  - His-tag via NTA or Ab
  - GST-tag via Ab
  - Protein A/G for Capture of Abs
- ~Permanent
  - Streptavidin - Biotin
    - Biotinylation via NHS-Biotins
    - Site-Specific via aviTag technology

Pros	Cons
Permanent	Random Orientation
	Can Inactivate

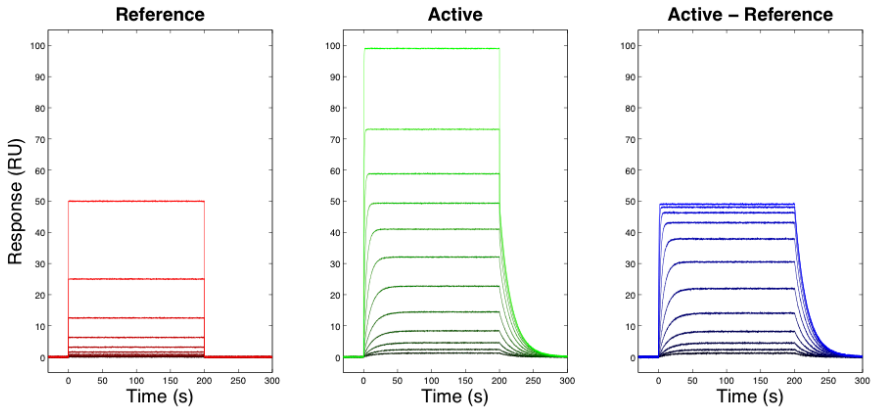
Uniform Presentation	Drifting Baseline
Can Recapture	Eats Sample

Uniform Presentation	
Stable Baseline	

Why Measure Affinities?

- Biochemistry is built upon enzymology
  - Historically, it has been easier to watch chemistry than binding events
- But a lot of biological processes occur in the absence of substrate turnover (ie- scaffolding)
  - Recognition is a necessary first step for events
- An affinity in isolation means nothing
  - Very difficult to connect in vitro affinities to concentration levels in the cell
- Very few proteins have a single interaction partner
  - Comparison of a panel of interactants allows extrapolation to competition within the cell
- Drug targets are no longer just enzymes
  - Characterization of binding surface properties and HT-screening

Raw Data:



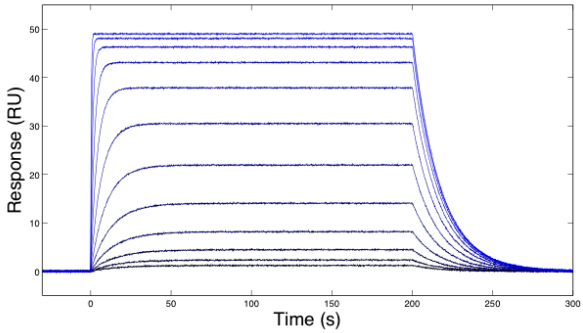
- All data gets double referenced
  - Reference 1: Subtraction of Reference Flow Cell
  - Reference 2: Subtraction of Buffer Injection

Good Data Reveals Kinetics and Thermodynamics:

- Model System

  1μM Interaction, ka = 6e4 $M^{-1}s^{-1}$, kd = 6e-2 s^{-1}

  12-point, 0.5-fold Titration from 50μM



- What to look for:

  Concentration dependent time to equilibrium

  Compression of response with increasing concentration

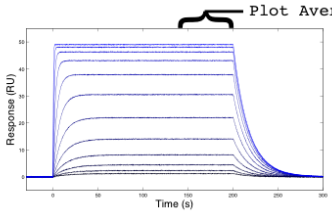
  Complete dissociation, return to baseline

The Equilibrium Model:

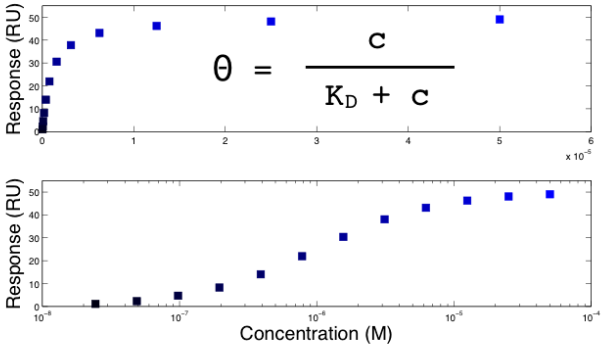
- Irving Langmuir (1881 - 1957)
  - Gas Filled Incandescent Lamp, Hydrogen Welding
  - First Industrial Chemist to Nobel Prize (1932)
- Langmuir 1:1 Binding

$$A + L \rightleftharpoons A{\cdot}L \qquad K_D = \frac{[A][L]}{[A{\cdot}L]}$$

Plot Average Response @ Equilibrium Against Concentration



$$\theta = \frac{c}{K_D + c}$$



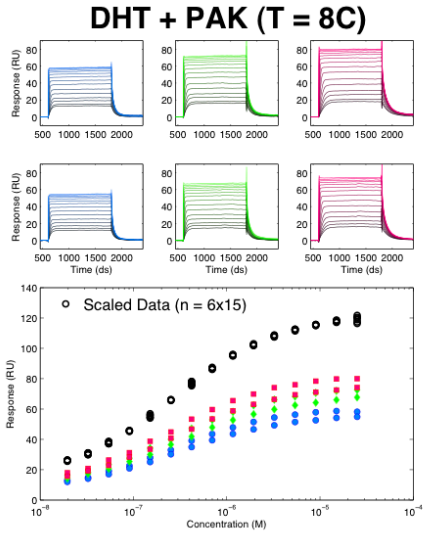
$K_A = 1 / K_D$
$K_A{*}A{*}L = A{\cdot}L$
$f_{L,A\cdot L} = A{\cdot}L / (L + A{\cdot}L)$
$f_{L,A\cdot L} = K_A{*}A{*}L / (L + K_A{*}A{*}L)$
$f_{L,A\cdot L} = K_A{*}A / (1 + K_A{*}A)$
$f_{L,A\cdot L} = A / (K_D + A)$

Fitting Steady State Data:

- Determine mean response at equilibrium
- Data from multiple experiments is scaled and merged
- Scaled data set is fit to the Langmuir model by minimization of SSE
- Fit parameters
  - K_D, Max Response

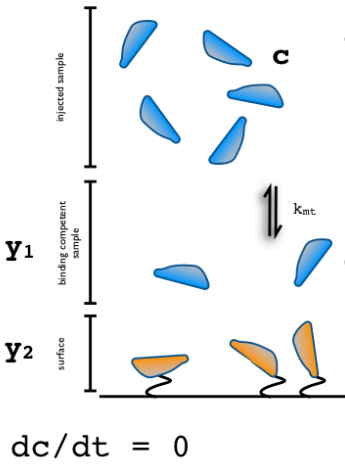
$$\theta = Max * \frac{c}{K_D + c}$$

- Always start with SS fits
  - Fewer parameters
  - More difficult to be fooled



The Kinetic Model:

- 1:1 Binding ± Mass Transport



Parameter	Units	Description
k_{mt}	s^{-1}	Diffusion Rate Between Flow and Surface
ka	$M^{-1}s^{-1}$	Forward Rate of Reaction
kd	s^{-1}	Reverse Rate of Reaction
max	RU	Signal at Saturation In practice, it is the fit value for y_2

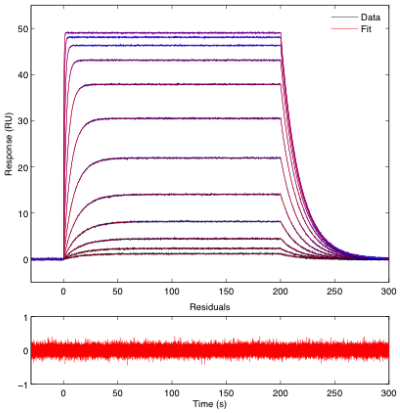
$$dc/dt = 0$$
$$dy_1/dt = kmt*c + kd*y_3 - kmt*c - ka*y_1*y_2$$
$$dy_2/dt = kd*y_3 - ka*y_1*y_2$$
$$dy_3/dt = ka*y_1*y_2 - kd*y_3$$

Fitting Kinetic Data:

- There is an analytical solution to the prior set of differential equations

$$R(t) = \left[\frac{k_a c}{k_a c + k_d} - R_0\right]\left(1 - e^{-(k_a c + k_d)t}\right) + R_0$$

- However, numerical solution is very fast with compiled diff eqs (function calls < 1ms)
  - Also, allows treatment of mass transport



Minimize SSE by adjusting k_a, k_d, k_{mt}, max
More parameters than SS analysis
More information

Kinetic Fits Must Show Data and Fit
Also, Good Practice to Show Residuals

The Most Common Mistakes:

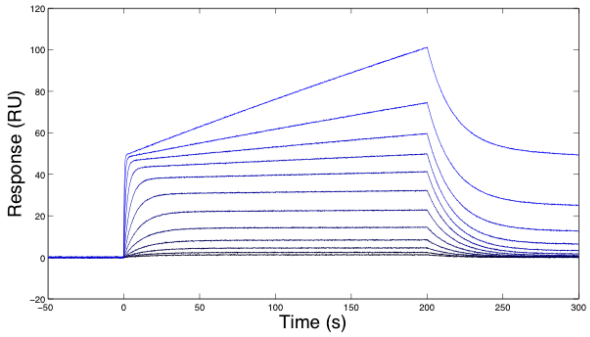
(1) Non-Specific Interactions

(2) Failure to Account for Bulk Shifts

(3) Operating Under Mass Transport Limitations

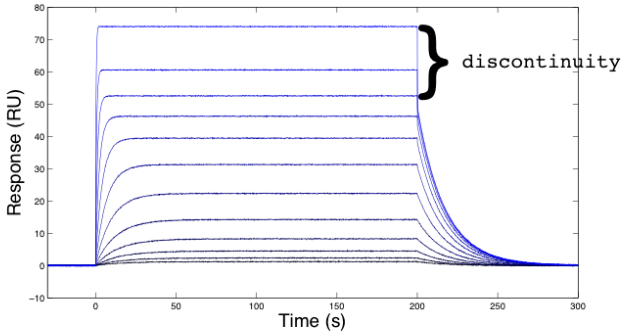
(4) Improper Presentation of Data

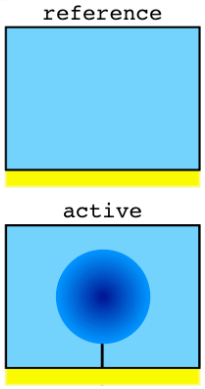
Non-Specific Interactions



- Same Model System with Trace Levels of a Contaminant
  - slow on, slow off, high max binding
  - 99% pure, but 1% co-purified bacterial chaperone or aggregate
- Optimize buffer conditions to minimize non-specifics
  - vary salt, ± BSA / tween20
- Purify the protein
  - SPR-grade >> XTAL-grade
  - Always recommend ion-exchange

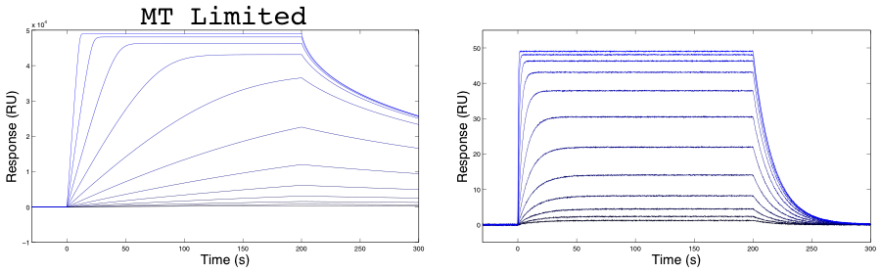
Failure to Account for Bulk-Shifts:





- Large RI differences between running buffer and injected samples
- Concentration dependent step-function skews association phase
- Discontinuous association / dissociation phases
- Due to excluded volume differences between active and reference surfaces
- Minimize the RI difference by dialysis of protein samples against running buffer
- Can be experimentally addressed by including solvent correction cycles

Operating Under Mass Transport Limitations



Association is limited by diffusion of analyte

- Model system with high levels of ligand
- Kinetics depend on amount of immobilized ligand
- Linear phase in association
- Dissociation shows two phases

- Immobilize lower levels of ligand
- Increase flow rate
  - This is the best way to determine if you are working under MT limit

Improper Presentation of Data:

- Tables Without Data
- Kinetic Fits to a Single Concentration
- Steady State Fits to a Single Isotherm
- Un-Steady State Fits
- Fit K_Ds Greater than High Concentration

State Of The Field:

The plan was to Pubmed search “Surface Plasmon Resonance” and count papers backwards till I found a good example...

Got to eight, before I gave up

bs is short for biosensor:

- Two Fundamental Flaws:
  - Poor experimental design
  - Misinterpretation of data
- Lack of Respect:
  - Show me the source for your table
  - A fit without data means nothing
- Direction of Science:
  - Structure often precedes function
  - Functional characterization is a hasty afterthought
- Commercialization
  - Easier to sell instruments under the impression that data collection/analysis is easy
  - Software is too friendly and eager to please
- A vast literature flooded with bad experiments

"Most Biacore users suck!"
-David Myszka

Grading the commercial optical biosensor literature—Class of 2008: 'The Mighty Binders'

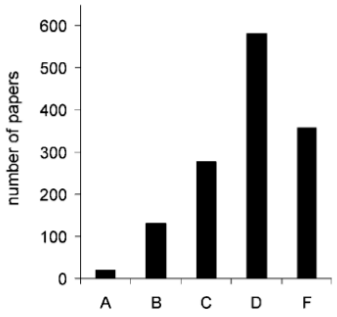
Rebecca L. Rich[a] and David G. Myszka[a]*

Review

Molecular Recognition

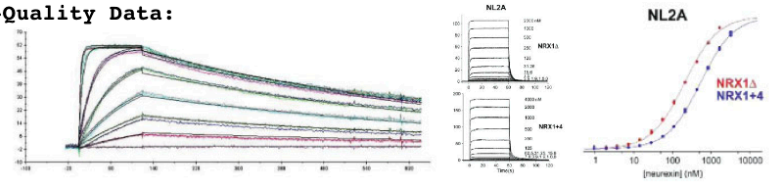
Received: 14 October 2009, Accepted: 19 October 2009, Published online in Wiley InterScience: 2009

- D Myszka has been publishing reviews of SPR in the literature since 1997
  - Unfortunately, Journal of Molecular Recognition doesn't have a huge readership
- Over the years, his tone has become increasingly frustrated:
  - His 2008 review is a must read
- Grades each of the 1413 papers that use SPR
  - Mean grade: Low D
- The field needs this:
  - Incited controversy
  - Subsequent comments on this paper ranged from strongly derogatory to great praise



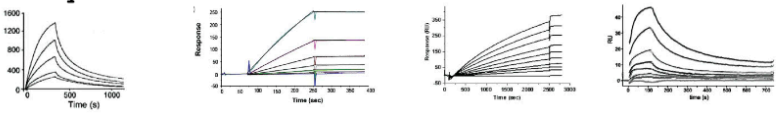
Representative Graded Data:

- **A-Quality Data:**

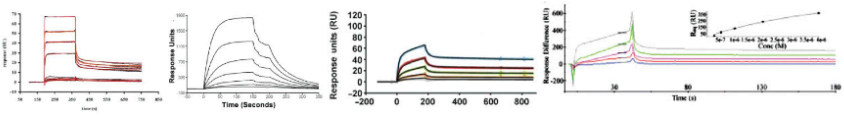


- **B-Quality Data:**

- **C-Quality Data:**



- **D-Quality Data:**



Some highlights:

- Off the wall critique draws attention to the lack of knowledge in proper implementation of the technique
- In my opinion, I think many fields would benefit from these sorts of reviews

Biosensor Pledge

1. I pledge to do my experiment right.
2. I pledge to report how I did my experiment.
3. I pledge to publish easily interpretable figures of my data that include replicates and an overlay of the fit to a simple model.
4. I pledge to report the binding constants and standard errors correctly.

Who can generate poor-quality data?
(we can, and we do)
Who publishes whatever comes out of the machine
Often without even showing the data?
(we can, and we do)
All the time!

Goldilock's responses	responses that are neither too big nor too small, but just right.
Hal-ing your results	publishing whatever the computer tells you.
hate constants	reported rate constants that do not describe the binding responses.
Michael Jackson data	data that are bad, bad really bad,—you know it.
Myszka's Razor	'Stop using a conformational-change model to fit your poor-quality data'.
Thpppfbt	sound signifying derision directed towards anyone claiming biosensor-determined binding parameters never match those from solution-based experiments.

Participation in 2011 Surface Plasmon Resonance Benchmark Study:

Beginning in 2003, David Myszka (University of Utah) has organized benchmark studies for biosensor characterization of biomolecular interactions. In this format, participating labs from around the world receive the necessary reagents to characterize an interaction. While originally motivated to establish the robustness of SPR characterization, this is a wonderful service to the community that brings together many users.

In 2011, I was invited to participate. A couple weeks later, I received a package containing a single spr chip and two tubes labeled analyte and ligand. Provided instructions specified the assay buffer, molecular weights of the species, and the simple directive of "characterize the interaction."

When I began assaying the system, I realized that the organizers had abandoned straightforward systems in favor of a more complex interaction. Simply, the system under investigation displayed a number of subtle non-ideal characteristics that could readily be ignored. In discussion with other participants in the study, most collected single datasets and reported superficial analyses. In response to the non-ideality, I took this benchmark as a personal challenge and refused to give the perfunctory standard analysis. While the system was, to me, clearly non-ideal, calling foul on the organizers required a level of audacity that was attained by in depth data analysis. After submission of my writeup, I received an email from the organizers offering a potential source for the non-ideal character of the system. Later, I learned that the system was deliberately contaminated in order to test users. I consider my participation in this study as affirmation of the power of SPR to provide deep mechanistic insight. What follows is my submitted writeup and figures for the 2011 benchmark study.

2011 SPR Benchmark Study

Experimental Design

Initial attempts to capture the biotinylated ligand failed. Upon receiving instructions to resolubilize the ligand in the BSA containing buffer, the second attempt at capture went as expected. Experiments were conducted on the supplied Xantec 3D streptavidin chip. Because of the ~2-fold difference in MW of the analyte to ligand, low density surfaces were targeted. Final capture levels of 25, 63, and 140 RU were reached. It should also be noted that the analyte was diluted from its stock concentration of 20μM without an initial dilution in the BSA containing buffer.

Before beginning a method run, test injections of the analyte were run to gauge the kinetics and concentration range for adequate data sampling. Because of the exceedingly slow dissociation, regeneration solutions were scouted. While the formed complex was resistant to high salt (1M KCl), the complex readily dissociated in the presence of 1M Gdn-HCl dissolved in the running buffer. Repeat injection series of analyte followed by the Gdn solution showed good reproducibility with minimal loss of binding capacity. Further, because of the well matched refractive indices of the sample and running buffer and continuity of the association and dissociation phases, a solvent correction was deemed unnecessary.

The characterization of the interaction was divided into four steps: Three conventional dose response (DR1 - DR3) experiments, and two rounds of single-cycle kinetics (SCK). For DR1 and DR2, a high concentration of 500nM was serially diluted by 0.6 fold to generate 15 concentration points spanning 391pM to 500nM. DR1 and DR2 consisted of 250s associations at flow rates of 15 and 10μLmin^{-1}, respectively. DR1 was given a 1000s dissociation time, while

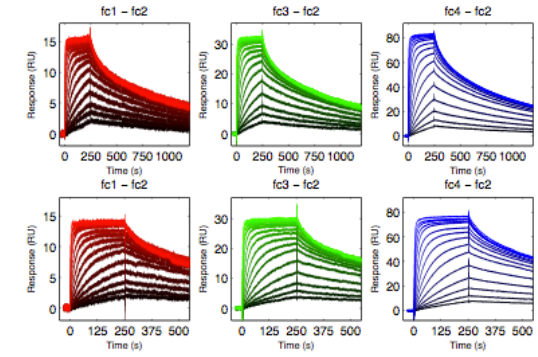
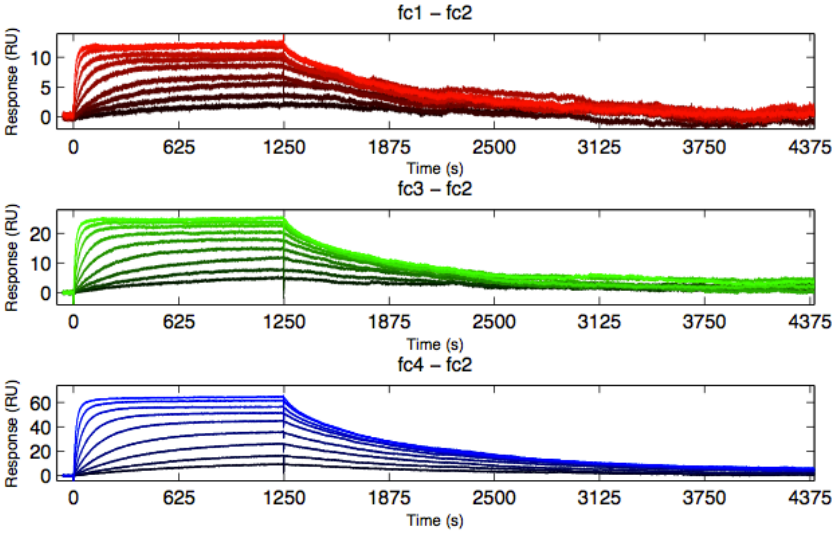
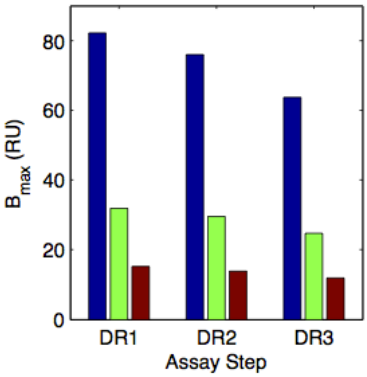


Figure 1. Raw traces and surface activity for the system. Titrations for DR1, DR2, and DR3 across the three active flow cells are shown in the top, middle, and bottom panels, respectively, on the left of this figure. Table with immobilization levels for the three active flow cells and predicted B_{max} values based on the molecular weights of the interactants is shown. Bar plot of surface activity over the course of the experiment.



	Immobilized (RU)	Predicted B_{max} (RU)	Expt'al B_{max} (RU)
fc1	25	46	15
fc3	63	116	30
fc4	140	260	80



DR2 was run with a 300s dissociation time. For DR3, the high concentration of 50nM was serially diluted by 0.5 fold to generate nine concentration points spanning 195pM to 50nM. In attempt to establish equilibrium of the interaction, DR3 was given a 1250s association time and 3200s was used to follow dissociation and allow eventual return to baseline. While the SCK method has proven a reliable strategy for SPR experiments, when implemented in the Biacore software, the user is not able to specify a time between injections. Because of this shorter than ideal dissociation time, fit k_d values are overestimated. For this reason, data from the SCK experiments were not included in the data analysis.

Data Analysis

First pass data processing was conducted in Scrubber 2.0. Each of the three DR experiments across each of the three flow cells was fit to a 1:1 binding model including a term for mass transport. This yielded nine sets of kinetic parameters listed in Table 1. Residual plots demonstrated marked systematic error in fitting to a 1:1 model. Additionally, the large spread of these kinetic parameters across experiments and flow cells cast considerable doubt on the validity of fitting the data to this model. Nonetheless, if the aim of this exercise is to report a single set of k_a, k_d, and K_D values, they are 2.21 x 10^6 M^{-1} s^{-1}, 1.9 x 10^{-3} s^{-1}, and 900 pM respectively. If the aim of this exercise was to omit / choose not to collect high concentration data or to entirely disregard the vast discrepancy between experimental and predicted B_{max} values and report the same set of parameters, the K_D value would be slightly tighter (~650 pM).

In attempt to address the exceptional mis-fitting of the data, further analysis was undertaken to explore the source of non-ideality. Inspection of residuals showed two flaws with the 1:1 model: 1. Fits were too quick to saturate at a lower than observed B_{max} value; and 2.

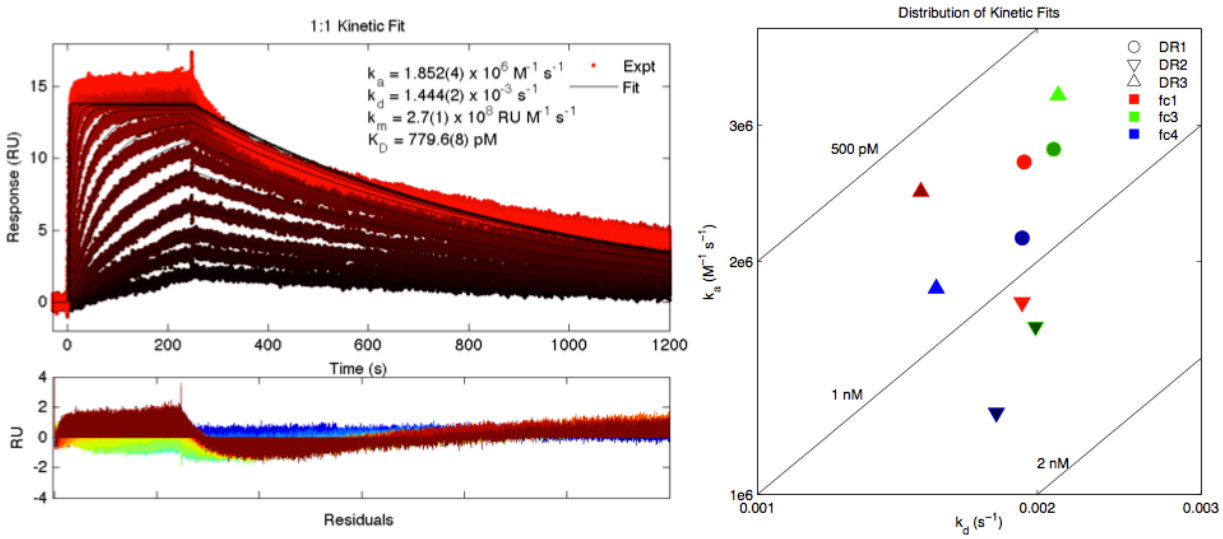


Figure 2. Representative kinetic fit using a 1:1 binding model. Data and fit from fc1-DR1 is plotted on the left with residuals below. The distribution of kinetic fits across the nine individual titration experiments is plotted on a log-log scale.

The fit dissociation constant was an apparent compromise between the slow rate observed at low occupancy and the fast rate that appeared at high occupancy. As this was a possible consequence of global fitting, each individual concentration curve was broken into association and dissociation phases and fit with simple exponentials to extract cycle specific kinetic parameters. Were a simple 1:1 model to explain the interaction, we would expect to find gaussian distributions. Not surprisingly, the histograms of these parameters show large spread and are non-uniform. In particular, the prominent secondary peak in the k_d distribution indicates the presence of a secondary bound species with a distinct half-life.

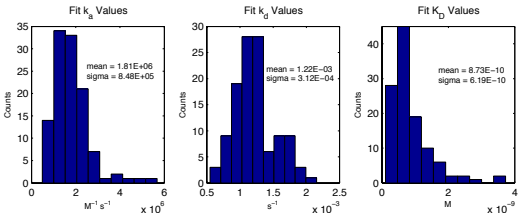


Figure 3. Histogram of simple exponential fits to all individual traces. Note the presence of a shoulder in the fit k_d distribution indicating two processes.

In order to resolve the number of bound species, sums of 1, 2, or 3 exponential decay processes were fit to the dissociation traces. As mentioned previously, the single exponential decay did not fit data. By addition of a second exponential term, the SSE decreased by five fold. Fitting of the data to a three exponential decay model only marginally improved the fit and suppressed the amplitude of the third rate to negligible values. This result is interpreted as a minimum of two bound species with appreciably different off-rates.

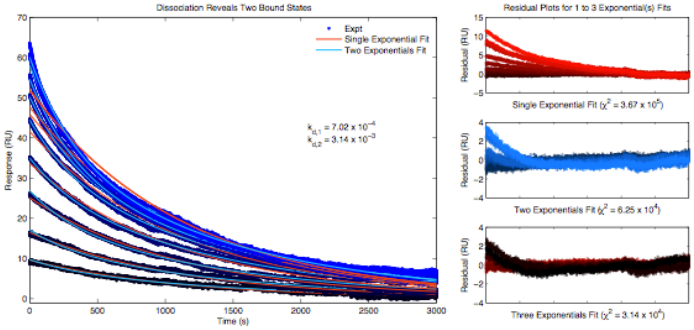


Figure 4. Dissociation reveals two bound states. Traces and fits for a single exponential or a sum of two exponential processes are shown on the left. Residuals for one, two, or three exponential decay processes are shown on the right.

One of the major surprises of the data was the much lower than expected B_{max} values. The immediate interpretation of this discrepancy is stoichiometry. Assuming 100% activity of the immobilized protein and a linear scaling of refractive index to molecular weight, the analyte appears trivalent. As these experiments were conducted on relatively sparse surfaces, it is

surprising that a single analyte molecule could bridge three different ligand molecules. This bridging could be a byproduct of working on a three dimensional surface that allows for flexibility of the immobilized molecules. It would be interesting to study this interaction on a planar sensor surface. However, the appearance of secondary rates at high concentration indicate that some of the analyte molecules are not fully satisfied with their interactions, and are, perhaps, only bound to single ligands.

The elusive piece of data was a true isotherm. The long time to equilibrium precluded direct measure of steady state occupancy at low concentrations. However, in fitting of each individual curve, local concentration B_{max} values were obtained. While it is, of course, preferable to obtain true equilibrium data, these fit B_{max} values can serve as effective predictors of equilibrium response for a given concentration. Scaling these values across all experiments resulted in an extrapolated isotherm spanning 195pM to 500nM. This data was first used to fit a 1:1 langmuir model. The data is clearly non-cooperative as it transitions across the 50% occupancy with a slope < 1. Further, the isotherm seems segmented with saturation of a high affinity site followed by titration of a weaker complex. While a Scatchard plot is, in the eyes of some, an archaic approach to fitting data, the theory is sound and it has the potential to reveal deeper character of an interaction. From the 1250s association data, a Scatchard plot was constructed with forecasted equilibrium responses for the four lowest concentrations. In agreement with the kinetic interpretation, this plot reveals two distinct binding processes in the absence of cooperativity.

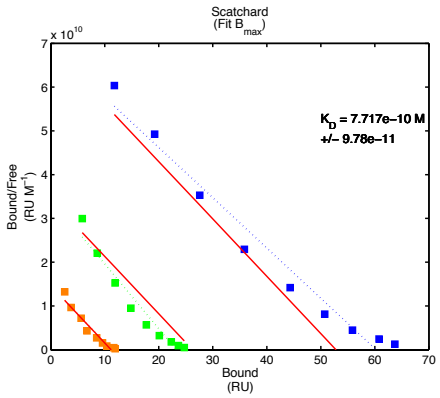
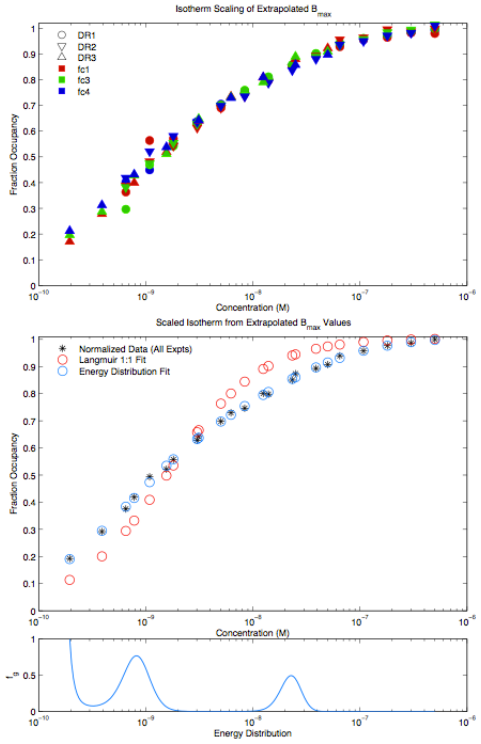


Figure 5. Scatchard plot reveals two binding processes.

Without knowledge of the expected stoichiometry for the interaction, the scaled isotherm from all experiments was fit using an adsorption energy distribution (AED). The AED model is a near perfect fit to the data. Importantly, the AED fit reveals a minimum of two concentration dependent binding phenomena.



1:1	k_a ($M^{-1}\,s^{-1}$)	k_d (s^{-1})	K_D (M)
	2.21×10^{6}	1.9×10^{-3}	9.08×10^{-10}
AED	K_{D1}	K_{D2}	K_{D3}
	812 pM	23 nM	< 195 pM (suspect)

Figure 6. Steady state parameterization of the system. All data was scaled across flow cells and individual points are plotted by fc and experiment. AED fit to mean response reveals two energetic processes. Inset table compares kinetic fit parameters with energies of interaction by the AED.

Discussion

It was fun. Felt like a good brain teaser. I'll be the first to acknowledge that my interpretation of this interaction could be entirely off-base. But the instructions were so vague, it seemed a bit open ended. So, I got a big kick out of working on this.

In critique of my data (raw traces), I see no fault and believe them to be true representations of the interaction under investigation. In critique of my experimental design, the 1M Gdn-HCl may have been a little harsh. Initial tests did not damage the immobilized molecules, but over the course of the thirty hour assay, there was clear loss of binding activity. I do not attribute this to dissociation of the biotinylated ligand from streptavidin. Rather, I think these molecules just died and stop participating in the experiment. The nice scaling of the extrapolated B_{max} values across all experiments indicates that the molecules that were still participating at the end of the experiments were representative of the freshly immobilized

molecules at the beginning. In critique of the extrapolation of B_{max} values, I don't think there is precedence for this and it is fundamentally wrong. However, it is a valuable piece of information and the mechanistic conclusions, while suspect, are relevant. Additionally, I take issue with my analytical fitting of sums of exponentials to the dissociation process. In so doing, I am neglecting mass transport effects. With that said, fitting single curves to mass transport models or

	FC1	FC3	FC4
	DR1		
k_m (RU M^{-1} s^{-1})	4.00(2) x 10^7	7.21(3) x 10^7	1.529(6) x 10^8
k_a (M^{-1} s^{-1})	2.689(6) x 10^6	2.793(7) x 10^6	2.145(5) x 10^6
k_d (s^{-1})	1.936(4) x 10^{-3}	2.082(4) x 10^{-3}	1.926(3) x 10^{-3}
R_{max} (RU)	14.105(3)	29.225(6)	75.94(1)
K_D	720.1(7) pM	745.4(7) pM	897.7(8) pM
σ_{res}	0.541	1.073	2.437
	DR2		
k_m (RU M^{-1} s^{-1})	1.(2) x 10^{10}	9.1(8) x 10^8	2.576 x 10^9
k_a (M^{-1} s^{-1})	1.771(5) x 10^6	1.647(4) x 10^6	1.275 x 10^6
k_d (s^{-1})	1.926(3) x 10^{-3}	1.991(3) x 10^{-3}	1.807 x 10^{-3}
R_{max} (RU)	12.856(3)	27.373(7)	70.38
K_D	1.088(2) nM	1.209(2) nM	1.418 nM
σ_{res}	0.541	1.138	2.445
	DR3		
k_m (RU M^{-1} s^{-1})	4.25(3) x 10^7	4.19(1) x 10^7	1.188(3) x 10^8
k_a (M^{-1} s^{-1})	2.463(7) x 10^6	3.280(8) x 10^6	1.846(2) x 10^6
k_d (s^{-1})	1.499(4) x 10^{-3}	2.106(5) x 10^{-3}	1.557(2) x 10^{-3}
R_{max} (RU)	10.876(3)	23.520(5)	60.675(7)
K_D	608.7(7) pM	642(2) pM	843.2(4) pM
σ_{res}	0.771	1.070	1.647

Table 1. Kinetic constants for individual 1:1 fits with a term for mass transport. The individual titrations for each flow cell were fit independently. Errors are given as the standard deviation of the residual vectors.

sequential dissociation ODE models to the traces did not improve fit. In another system I study, the number of required exponentials is well correlated to structural changes upon complex formation.

What do I think it is? Because of the tight binding and its resistance to KCl and susceptibility to Gdn, I think the interaction involves significant burial of hydrophobic surface. Why is the data so poorly fit by a 1:1 model? Most likely it is because of dirty samples with a significant amount of dead ligand and competing analytes in solution. Second option is that it's not a 1:1 interaction. This option is supported by a glut of evidence clearly showing multiple bound states. Without direct knowledge of the interactants, it is difficult to finalize a stoichiometry for the interaction. But the phenomenological result of the AED fit warrants brief discussion. The input data was less than ideal for two reasons. First, the prediction of B_{max} values for low concentrations that do not reach equilibrium is not okay. Second, the concentration range did not extend low enough. This second reason is manifest in the appearance of a tight binding shoulder. Were the data to extend to lower concentrations, this artifact would likely be suppressed or resolved into a discrete third concentration dependent event. The absence of a weak binding shoulder speaks to the specificity of this interaction and argues against contamination in the analyte solution. I have spent a good amount of time feeding models into the AED algorithm, and the appearance of multiple peaks requires specific character to the data. In application of AED analysis to oligomeric systems, I have observed peaks corresponding to oligomerization affinities of the molecules in solution. This is unlikely, as one would observe larger than predicted B_{max} values. The most facile way to generate three peaks in an AED fit is with three non-competitive bound states. Assuming trivalency, the different bound states could represent the three possible stoichiometries.

I really feel like a bit of a nut for going so deep in this. But, I think the utility of SPR is too often viewed as just a way to quantitate someone's favorite interaction. In so doing, the mechanistic insight provided by the data is neglected. When presented with something that is clearly not fit by a model, it is our responsibility to find a model that does adequately describe the data. Without direct knowledge of the interactants, I decline to propose a specific model. With that said, I do believe that my analysis reveals certain characteristics essential to a model that could capture the data.

Thanks,

Kris

Chapter 3:

DNA sequence modulates dimerization energetics for the androgen and glucocorticoid receptor DNA binding domains

In collaboration with:

Lisa Watson, Benjamin J Schiller, John D Gross, Miles A Pufall, and Keith R Yamamoto

Reprinted with permission from:

Watson LC, Kuchenbecker KM, Schiller BJ, Gross JD, Pufall MA, Yamamoto KR. The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nat Struct Mol Biol. 2013 Jul; 20(7):876-83.

Foreword:

DNA sequence-specific kinetics and cooperativity are achieved via intermolecular signaling at the dimerization surface of the DNA binding domain.

Cooperativity does not require allostery:

Differential cooperativity strongly suggests allostery:

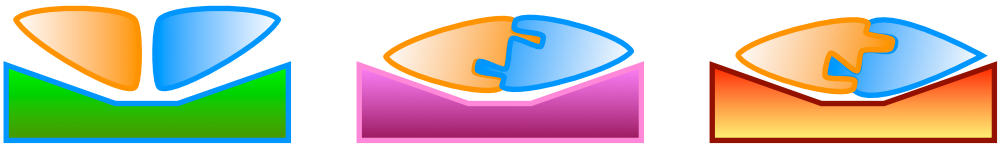


Figure f1. This schematic represents DNA recognition by a dimer. For the three DNA sequences, energetic cooperativity is possible in the absence of structural rearrangements at the dimerization interface. Alternatively, sequence specific cooperativity could manifest via structural rearrangements at the dimerization interface.

My thesis project focused on the fast time scale coregulator interactions of nuclear receptorligand binding domains. When Lisa Watson (Tetrad, Keith Yamamoto) and I first discussed the potential for surface plasmon resonance measurement of DNA recognition by nuclear receptor DNA binding domains, the project piqued my interest for two reasons: (1). The complexity of the dimer binding process; and (2.) The opportunity to study the intermediate time regime in nuclear receptor signaling.

Historically, estimates of DNA binding affinity have been made using gel shift assays. Because dissociation / dilution begins upon electrophoresis, this non-equilibrium technique will

systematically underestimate affinity. Simply, if two complexes have equivalent binding energies, but one dissociates more rapidly, a gel-shift measure will give a tighter binding affinity for the longer-lived complex. As surface plasmon resonance is a true equilibrium technique that also enables kinetic measurements, Lisa's question of differential binding energetics was well suited. However, the lore of the surface plasmon resonance community would strongly discourage the use of the technique because of possible complications due to the presence of dimers. Despite this, the intrigue of kinetic information on the cooperative binding process was too great. The only way to overcome these concerns was by consistent measurement of a well behaved system.

From a biophysical perspective, the time scale of the three molecular recognition processes of nuclear receptors (ligand binding, dna binding, and coregulator binding) are vastly different. These processes scale such that a single ligand binding event would result in multiple DNA recognition events; and each DNA recognition event would provide time for sampling of multiple coregulator interactions. The primary function of the nuclear receptor occurs when bound to DNA and is the recruitment of transcriptional machinery. From an energetic perspective, an accurate measure of binding energy would inform the probability of gene regulation at specific DNA sequences. Information on the half-life for a nuclear receptor at different DNA sequences would inform the activity potential for a transcriptional regulatory event occurring at a specific locus. Also, the potential for DNA-specific cooperativity might provide insight into graded and switch like transcriptional responses.

Lisa's story focused on the glucocorticoid receptor. It made beautiful use of a mutation that disrupted the dimerization surface. By NMR structural analyses, the study shows this disruption reduces the structural ensemble. Surface plasmon resonance measures of

cooperative binding energetics and kinetics of dissociation revealed that this collapse in the structural ensemble also reduces the biophysical space that can be used in the discrimination of DNA sequences.

As a control in our experiments, DNA binding by the androgen receptor was assayed in parallel. While these results were not included in the publication, I include here a deconstruction of the poster I presented at the 2012 Bay Area Label Free Users Group.

abstract:

The androgen receptor (AR) is a ligand activated transcription factor. The dna binding domain (DBD) of AR recognizes inverted repeats of six bases. This work describes the dna recognition properties of the isolated DBD. Surface plasmon resonance was used to investigate the dynamic range of free energy and kinetics in dna recognition by the AR DBD.

DNA	Sequence
PAL	AGAACA TTT TGTTCT
GHA	GGAACA TAA TGTTCC
GILZ	AGAACA TTG GGTTCC
SGK	AGAACA TTT TGTCCG
FKBP5	AGAACA GGG TGTTCT

Table f1. DNA oligos used in this study. With respect to the canonical palindromic DR3 (PAL), differences in sequence are highlighted in red.

Experimental Design:

Protein Purification: The DNA binding domain of androgen receptor was recombinantly expressed in E.Coli as a 6x-His-TevSite-AR (aa's 531-646)-STOP fusion protein. Expressed protein was purified by IMAC, size exclusion,

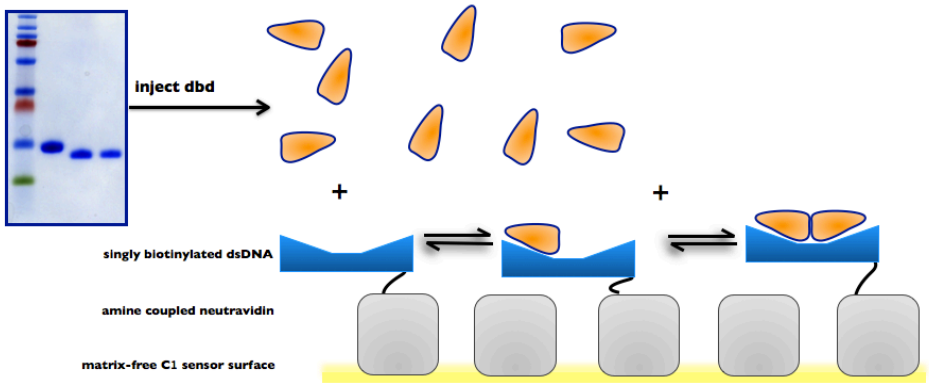


Figure f2. Schematic of the experimental system.

TEV protease treatment, and ion exchange. Protein concentration was determined by A_{280} in 6M Guanidine-HCl.

SPR Chip Preparation:

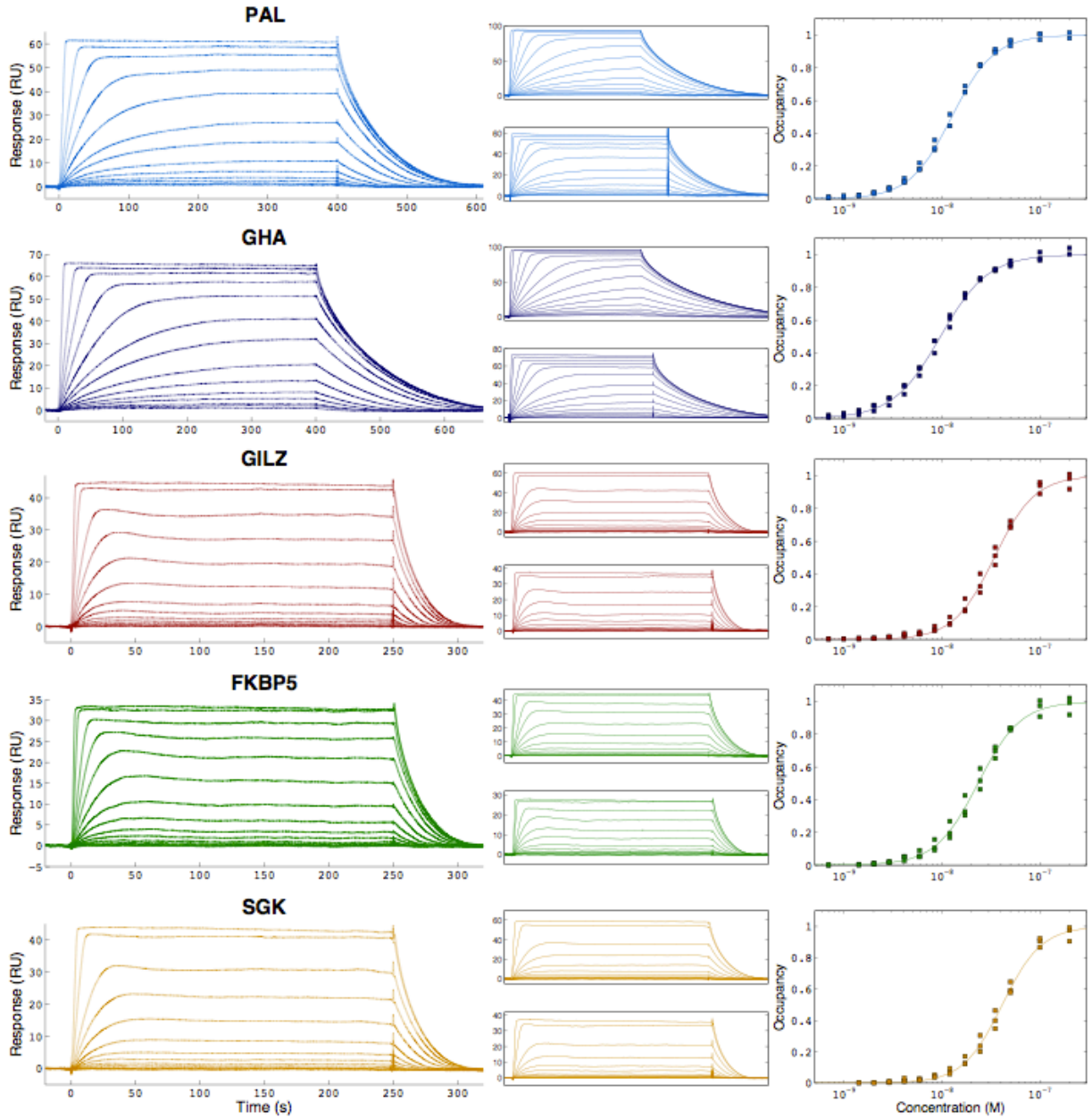


Figure f3. Raw data and scaled isotherms. The interactions between AR and the five response elements were measured in triplicate using the same 15-point concentration series of 200nM-100nM-50nM, followed by 1.4-fold dilutions to 700pM. Response at steady state was recorded and scaled between replicates. In the isotherms on the right panel, scaled data from all replicates is plotted, while fits to the Hill equation are plotted as lines.

Neutravidin was covalently coupled to BiaCore C1 sensor chips using conventional EDC/NHS mediated amine chemistry. Singly biotinylated dsDNAs were subsequently captured at immobilization levels between 20-50 RU.

Experimental Conditions:

All experiments were conducted in a standard phosphate buffer supplemented with 0.01% BSA to prevent loss at low protein concentrations. All experiments were undertaken on a BiaCore T200 at 8C with a flowrate of 30μL min^{-1}.

Data Processing:

All analysis was performed in Matlab.

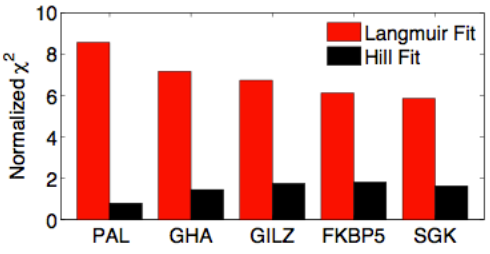
	Fit $K_{1/2}$ (mean±2σ)	Fit n_H (mean±2σ)
PAL	12.5 ± 0.7nM	2.05 ± 0.13
GHA	9.4 ± 0.4nM	1.79 ± 0.15
GILZ	35 ± 3nM	2.11 ± 0.26
SGK	43 ± 8nM	2.10 ± 0.34
FKBP5	24 ± 3nM	2.03 ± 0.26

Table f2. Steady state estimates of affinity ($K_{1/2}$) and cooperativity (n_H).

DNA Recognition is Strongly Cooperative:

Langmuir Model: $B_{\max} \times \frac{c}{c + K_D}$

Hill Model: $B_{\max} \times \frac{c^{n_H}}{c^{n_H} + {K_D}^{n_H}}$



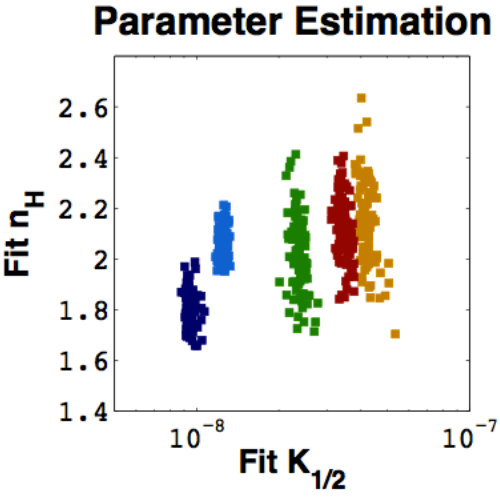


Figure f4. The isotherms reveal cooperative interaction. For all five DNAs, the isotherms are significantly better fit using the Hill equation than the Langmuir model. Bootstrap with replacement was used to estimate fit parameters. The comparatively large spread in fit n_H for the three weak sites is due to inadequate sampling along the midpoint of the concentration curve. Parameter estimates are colored by DNA site (see Figure f3.)

Equilibrium measurement of DNA binding allows accurate determination of affinity. In comparison to gel-shift assays, equilibrium SPR measurement reveals much tighter affinities for these interactions. Interestingly, the discrepancies between EMSA and SPR are greatest for the fastest dissociating complexes.

Further, the reproducibility and sensitivity of the SPR measurement allows determination of relative cooperativity with moderate precision. A weak anti-correlation between affinity and Hill coefficient is apparent.

Error Treatment:

A bootstrap with replacement strategy was implemented in assessing the error in parameter estimation. In short, considering a scaled isotherm of *n* points, the fitting process is initiated by random selection of *n* points with the possibility of selecting the same point multiple times. These *n* points are used in a round of conventional least squares minimization. This process is then repeated 500 times, and sets of parameters are ranked by sum of squared error. Fit $K_{1/2}$ and n_H are presented as the mean ± two standard deviations of the 250 best sets of fit parameters for the different interactions.

Improvements to Experimental Design:

Failure to adjust the concentration series across the panel of interactants led to large errors in fit Hill coefficients for the weak binding interactions. In these three cases, the fit Hill coefficients are likely overestimates as there is an inadequate number of concentration points sampling the steepest part of the isotherm.

Minimization Strategy for Kinetic Fits:

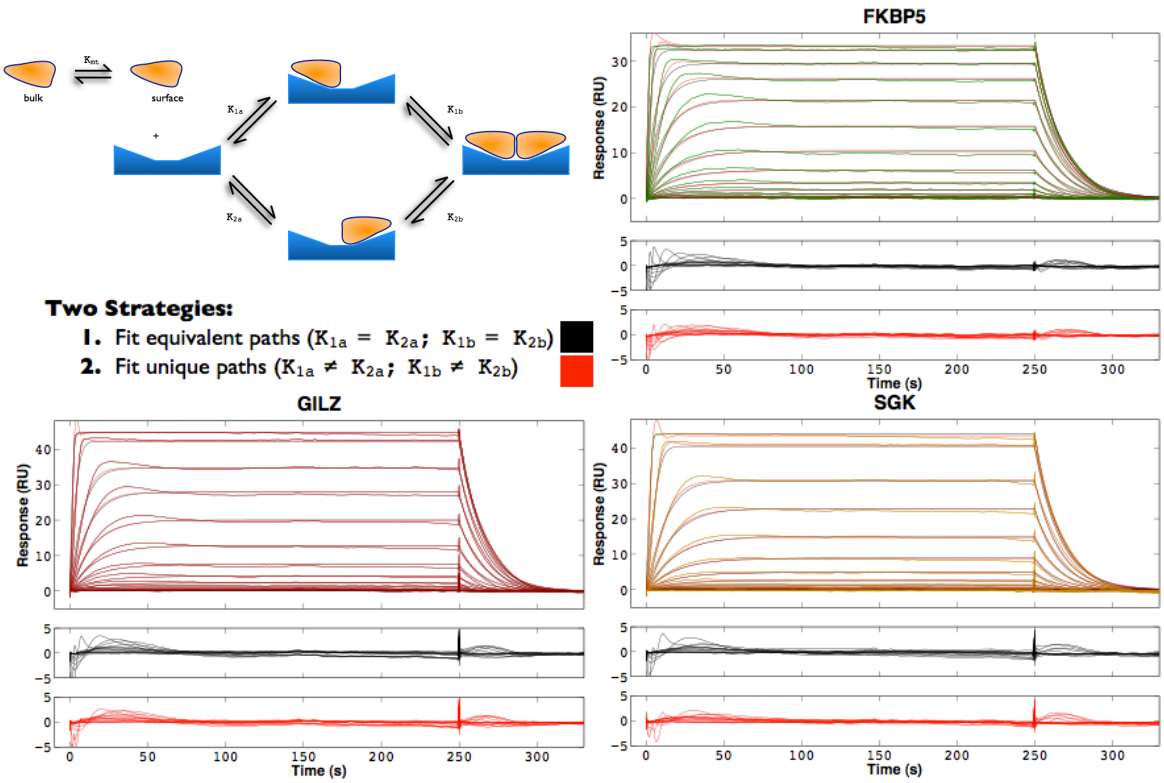


Figure f5. Model for kinetic evaluation of dimer assembly on the DNA. Data was fit comparing equivalent vs. unique paths of assembly. Residuals for the equivalent path are shown in black while residuals for the unique path model are shown in red. While the unique paths lowers the sum of squared error, there is no change to the structure of the residuals, and the more elaborate model is not justified.

To search parameter space for the full binding model, an annealing algorithm was implemented in conjunction with conventional least squares minimization. In short, following a search, a randomly selected parameter was kicked before re-searching. Acceptance rate for new parameters was tuned down over the course of 20 pseudo-annealing steps.

Quality of Kinetic Fits:

Structure of residuals is not improved by allowing unique paths. PAL and GHA were omitted from kinetic analysis because of severe mass transport limitation. This is likely also affecting the fits for GILZ, FKBP5, and SGK. The quality of data is not suited to kinetic analysis, and these are not good fits. This should be improved by data collection at sparser surface densities and optimization of buffer conditions.

Kinetic Assessment of Cooperativity:

Enhancement of binding energetics is largely determined by slowed monomer dissociation from the assembled ternary complex. Extrapolating to assembly, monomer bound species are longer lived and less likely to result in full complex formation for FKBP5. In contrast, the GILZ site is more switch-like. By steady state analysis, the Hill coefficients for GILZ and FKBP5 were statistically indistinguishable. By comparison of the fold enhancement in kinetics for the two binding events, these DNAs differentially affect cooperativity. While these interactions occupy unique regions of kinetic space for the first binding event, the kinetics of dimer assembly on the DNA are very similar. As this enhancement is non uniform, these results suggest sequence specific energetics for dimer assembly.

Cooperativity does not require allostery:

Cooperativity can manifest in the absence of context dependent structural change. If the dimerization surface of the DBD were static, one would still expect deviation from the Langmuir model as an energetic consequence of surface burial. However, considering a static dimerization surface, this energetic gain would be constant across different DNA sequences. While the kinetic analysis is less convincing, the differential cooperativities (Hill coefficients) in binding to GHA and PAL strongly suggest that the DBD is able to tune its dimerization energetics in a DNA sequence specific fashion.

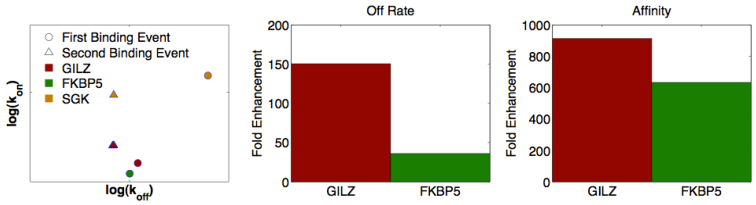


Figure f6. Kinetics Reveal Differential Cooperativity. Kinetic parameters (k_{on} / k_{off}) are plotted on a log-log scale comparing the first and secondary binding events. The primary mechanism for cooperativity is a decrease in off rate upon dimer assembly.

The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals

Lisa C. Watson[1,2], Kristopher M. Kuchenbecker[3,4], Benjamin J. Schiller[1,2], John D. Gross[5], Miles A. Pufall[2,6], Keith R. Yamamoto[2]

[1]Tetrad Graduate Program, University of California, San Francisco, San Francisco, California, USA.

[2]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California, USA.

[3]Biophysics Graduate Group, University of California, San Francisco, San Francisco, California, USA.

[4]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California, USA.

[5]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California, USA.

[6]Present address: Department of Biochemistry, University of Iowa, Iowa City, Iowa, USA.

Correspondence should be addressed to K.R.Y. *email: yamamoto@cmp.ucsf.edu*

In collaboration with:

Lisa Watson, Benjamin J Schiller, John D Gross, Miles A Pufall, and Keith R Yamamoto

Reprinted with permission from:

Watson LC, Kuchenbecker KM, Schiller BJ, Gross JD, Pufall MA, Yamamoto KR. The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nat Struct Mol Biol. 2013 Jul; 20(7):876-83.

Abstract:

Glucocorticoid receptor binds to genomic response elements and regulates gene transcription with cell- and gene-specificity. Within a response element, the precise sequence to which the receptor binds has been implicated in directing its structure and activity. We use NMR chemical shift difference mapping to show that non-specific interactions with particular base positions within the binding sequence, such as those of the "spacer", affect the conformation of distinct regions of the rat glucocorticoid receptor DNA binding domain. These regions include the DNA-binding surface, the "lever arm" and the dimerization interface, suggesting an allosteric pathway that signals between the DNA binding sequence and the associated dimer partner. Disrupting this path by mutating the dimer interface alters sequence-specific conformations, DNA-binding kinetics and transcriptional activity. Our study demonstrates that glucocorticoid receptor dimer partners collaborate to read DNA shape and to direct sequence specific gene activity.

Introduction

Gene expression is tailored to the needs of specific tissues and in response to environmental and developmental changes. Transcriptional regulators coordinate this task by integrating input signals at specific genomic regions[1,2] to effect precise transcriptional outputs at target genes. This intricate process relies on combinatorial control, in which distinct combinations of factors assemble into functional regulatory complexes that control the transcriptional activity of associated genes. However, the determinants that define the gene-specific assembly and activity of these regulatory complexes are poorly understood.

Glucocorticoid receptor (GR), a glucocorticoid-activated member of the nuclear receptor superfamily, utilizes combinatorial control to regulate hundreds to thousands of genes in a cell- and gene-specific manner. In part, this specificity arises from context-dependent GR binding regions (GBRs), which can be defined *in vivo* using genome-wide approaches. Some, but not all, GBRs appear to function *in vivo* as glucocorticoid response elements (GREs), which confer context-specific glucocorticoid regulation upon nearby genes. While GBR and GRE activities are clearly separable, both rely on the effects of multiple signals, such as hormonal ligands, other regulatory factors, and post-translational modifications. Each of these signals drives distinct conformational changes in the receptor[3-8], thereby modulating its transcriptional regulatory activity[9-11]. For example, two GR ligands, dexamethasone (dex) and RU486, differentially affect the formation of a coactivator interaction surface of the ligand-binding domain[8] and induce different transcriptional profiles.

GBRs and GREs are composite elements consisting of binding motifs for non-GR transcriptional regulatory factors and often one or more GR binding sequence (GBSs)[12]. GBSs are bound with high affinity by purified GR *in vitro and* mutational studies have confirmed that

GBSs within a particular GBR are responsible for GRE activity[13], (Thomas-Chollier, M., unpublished). GBSs vary loosely around a 15 base pair (bp) consensus sequence consisting of two hexameric half-sites separated by a 3 bp spacers[13]. GR binds to a GBS as a homodimer with each dimer partner specifically contacting, at most, three bases within each GBS half-site. Structural studies of free and DNA-bound GR DBD suggest that DNA binding imparts structural changes in the second zinc finger of the DBD, forming the dimerization interface[14-16].

Our lab previously demonstrated that DNA binding sequences serve as distinct signals that direct GR structure and activity[17,18]. Crystallographic studies comparing GR bound to different GBSs revealed alternate protein conformations that are dependent on the precise DNA binding sequence[17]. The observed alternate conformations were localized to a loop region within the DNA binding domain (DBD) termed the lever arm, which does not itself contact the DNA. Moreover, GBSs that produced different lever arm conformations were invariant at all nucleotide positions that make direct contacts with GR, indicating that non-specific bases affect GR structure. The presence of alternate lever arm conformations suggests that GBS-specific conformational dynamics play a role in GR gene-specific regulation.

These crystallography studies motivated the following questions: (1) how does GR detect sequence differences among GBSs, (2) do GBSs drive distinct "allosteric paths" of conformational changes that extend into and through the lever arm, and (3) how do GBS-dependent differences in GR conformation impact GR activity? To address these questions, we used solution techniques to assess the effects of changing precise nucleotide positions within the GBS and perturbing a functional surface of the GR DBD.

Results

GBS spacer affects GR occupancy, activity and structure

We sought to determine the degree of sequence variability among endogenous GBSs to estimate the potential for DNA sequences to be unique signals that produce distinct GR activities. We identified GR binding regions in U2OS cells exogenously expressing full-length rat GR, using GR chromatin immunoprecipitation (ChIP) followed by deep sequencing of the precipitated DNA fragments. An unbiased search for sequence motifs within 1,000 GR binding regions with the highest number of reads revealed a GBS motif composed of imperfect palindromic hexamers separated by a 3 bp spacer (**Fig. 1a**), similar to previously identified motifs based on smaller sample sets (*e.g.,* [12]). Scanning for this motif among the observed 30,000 GR binding regions revealed that 90% of GBSs are unique, suggesting that there is sufficient diversity for each to be a gene-specific signal. This model would require that non-specific bases contribute to sequence specificity. The GBS positions with the highest information content (>1 bit) correspond to the six bases that are directly contacted by the GR dimer[16,17]. The remaining nine nucleotide positions each contain less than 1 bit of information, with half-site positions 3 and 13 nearly devoid of sequence preference (0.05 and 0.1 bits, respectively). Notably, however, GR displayed appreciable base-preference at positions that it does not contact directly: pyrimidines at spacer positions 7-9, as well as A and T at positions 6 and 10, respectively, adjacent to the spacer.

It was previously shown that GBSs differentially modulate GR transcriptional induction using luciferase reporters consisting of a single GBS upstream of a minimal promoter[17]. To investigate how varying the GBS spacer affects GR transcriptional induction, we compared

reporter activity in the presence of 100 nM dex for GBSs that differ only in spacer sequence (**Fig. 1b-c**). Changing the spacer of Sgk from TTT to GGG resulted in a 69% decrease in transcriptional activation and changing the spacer by only one base (from TTT to TTG) resulted in an intermediate 42% decrease. Alternatively, changing the spacer of Fkbp5 from GGG to AAA (Pal-F), though not TTT (Pal-R) resulted in decreased transcriptional activity. Thus, spacer sequence, within the context of the whole GBS, influences GR activity.

As GR does not directly contact the GBS spacer, we investigated other potential structural mechanisms by which spacer sequence influences GR function. Though prior crystallography studies did not include the Sgk-ggg GBS, we aligned structures of GR DBD bound to the Pal-F and Fkbp5 GBSs, which differ only in spacer (**Fig. 1d**). These structures revealed that the minor groove of the Pal-F spacer is narrower than that of the Fkbp5 spacer, with average widths of 3.8 Å and 6.4 Å, respectively, as measured by Curves+[19]. As the Pal-F and Fkbp5 GBSs have spacer sequences of AAA or GGG, respectively, the sequence-specific difference in minor groove width is consistent with previous studies showing that short A-tracts narrow the minor groove[20]. This led us to hypothesize that DNA shape, defined by the nucleotide sequence of the GBS spacer, serves as a signal that regulates GR activity. Thus, we predict that structural features associated with the DNA spacer impart structural changes that are propagated through the GR DBD. Consistent with this, our examination of GR contacts with the DNA phosphate backbone near the GBS spacer indicated that the orientation of the side chain of Lys490 is dependent on the spacer sequence (**Fig. 1e**).

D-loop conformation depends on spacer but not half-site

To further investigate how GBS spacer affects GR structure, we monitored GR DBD conformation by ^{15}N-HSQC, which measures the chemical environment of the amide bond of individual amino acid residues. For GR DBD bound to a high-affinity GBS (Gha), we assigned over 90% of the chemical shift peaks to their corresponding residues (**Supplementary Fig. 1**). We overlaid the HSQC spectra for GR bound to GBSs that differ only at non-specific bases of the spacer or at half-site positions 13 and 15 (hereafter referred to as half-site$_{13:15}$). In both comparisons, the spectra displayed non-overlapping peaks indicating that each GBS complex is structurally distinct (**Fig. 2a**), and suggesting that both the sequence of the spacer and half-site$_{13:15}$ influence GR structure.

We used chemical shift perturbation, which is sensitive to local changes in conformation[21-23], to identify regions of GR DBD that are affected by spacer sequence compared to those affected by half-site$_{13:15}$. GBSs that differ only in the spacer resulted in substantial peak shifts mapping to Ala477 and Gly478 of the GR dimerization loop (D-loop) (**Fig. 2b**), despite a distance of at least 18 Å between the GR backbone and the nucleotide bases of the spacer. In contrast, half-site$_{13:15}$ had little effect on Ala477 and Gly478, but instead affected residues at other surfaces of the DBD (**Supplementary Fig. 2**). Thus, sequence variation within particular regions of the GBS corresponds with conformational changes in distinct subdomains of the DBD.

Extending this analysis across the GR DBD further demonstrated that changing the spacer, but not half-site$_{13:15}$, induced peak shifts mapping to the D-loop of GR (**Fig. 2c-d**). Alternatively, half-site$_{13:15}$, but not the spacer, influenced peaks mapping to outward-facing surfaces of the DBD near the DNA (**Fig. 2d** and **Supplementary Fig. 2**). Additionally, the DNA recognition helix (H1) and the lever arm were affected by sequence changes in either the

spacer or half-site[13:15]. This effect of GBS on the lever arm conformation further corroborates the alternate conformations observed by crystallography[17].

How might information in the GBS spacer be transmitted across the substantial distance to the D-loop to elicit specific rearrangements? As the lever arm structurally links the DNA recognition helix to the D-loop, the simplest model suggests structural coupling of the D-loop to the DNA recognition helix via the lever arm.

The A477T mutation disrupts D-loop conformation

To investigate the functional role of spacer-specific structural changes, we tested whether perturbing the D-loop affected GR activity in a GBS-specific manner. We focused on Ala477 of the D-loop, which makes one of the four dimerization contacts within the GR DBD[16] —a backbone hydrogen bond between the carbonyl of Ala477 and the amide of Ile483 from the associated dimer partner. This mutation, A477T, has been shown to alter GR activity in a gene-specific manner[24]. As reported previously[17,25], the extent to which A477T transcriptional induction differed from wild type GR (WT) was GBS-specific. Among eight GBSs tested in reporter assays, the A477T mutation resulted in increased, decreased or unchanged transcriptional induction compared to WT (**Fig. 3a**). In contrast to WT, A477T activity was indistinguishable for two GBSs that differ in spacer (Sgk and Sgk-ggg). Thus, this mutation in the dimerization interface did not abolish GR activity, but instead resulted in reinterpretation of GBS signals by the mutant GR.

To assess the mechanism by which this mutation differentially affects GBS-specific activity, we characterized the structural and biophysical impacts of A477T. We compared the HSQC spectra of WT and A477T bound to the Fkbp5 GBS (**Fig. 3b-c**). The A477T DBD

spectrum showed many peaks that overlaid well with GR WT, but over 30% of residues were shifted as a result of the A477T mutation. Importantly, these peaks did not overlay with those corresponding to the unbound WT DBD, ensuring that the protein is completely bound to DNA (**Supplementary Fig. 3**).

We quantified the chemical shift difference as the distance between each peak in the GR WT spectra to the nearest peak in the A477T spectra[25]. Comparison of WT- and A477T-bound Gha, Fkbp5, and Sgk complexes (**Fig. 3d**) revealed A477T-specific shifts that mapped to the D-loop and the residues surrounding Ile483, consistent with a disruption of the dimerization interface. Additional chemical shift differences mapped to the N-terminal region of the lever arm and the recognition helix of GR. Thus, the A477T mutation generates local structural changes as well as structural reorganization in regions outside of the dimerization surface. Together with the observation that WT GR produced GBS-specific structural changes in the dimerization interface (**Fig. 2b-c**), our findings with the A477T mutation indicated that the dimerization and DNA interfaces are structurally coupled.

A477T affects cooperativity but not stoichiometry

To pursue the mechanism by which the A477T mutation affects the GBS-specific activity of GR, we assessed DNA binding by an electrophoretic mobility shift assay (EMSA). Comparison of WT DBD and A477T DBD binding to the Pal-R and Sgk GBSs revealed that GR dimer complexes were formed at saturating concentrations of both WT and A477T, though the mutant showed reduced DNA-binding affinity (**Fig. 4a**). For the WT, the transition from free DNA to dimer complex occurred at lower concentrations of GR, with only a minor population of DNA-

bound monomer, indicative of strong positive cooperativity. In contrast, A477T displayed little cooperativity, having nearly saturated the DNA as a monomer prior to dimer formation. To distinguish whether the reduced overall affinity of A477T was due to impaired DNA recognition resulting from the mutation, we compared binding of GR WT and A477T to a GBS half-site (**Fig. 4b**). We found that half-site binding was equivalent for WT and A477T, indicating that the A477T mutation does not disrupt the DNA-binding ability of the monomer (**Fig. 4b-c).** Thus, the reduction in overall affinity of the A477T mutant is due to diminished cooperativity.

As the EMSA is a non-equilibrium measure, surface plasmon resonance (SPR) was additionally used to monitor the effects of binding sequence on the DNA recognition properties of WT and A477T, under similar conditions as used for the NMR studies. We compared WT DBD or A477T DBD binding to two GBSs whose transcriptional induction in reporter assays was reduced (Pal-R) or unaffected (Gilz) by the A477T mutation (**Fig. 5a**). Isotherms constructed from maximal binding of WT or A477T DBD to GBS-immobilized surfaces at equilibrium showed that the mutation results in a decrease in binding affinity by a factor of 10 and 5 for Pal-R and Gilz, respectively (**Fig. 5b** and **Supplementary Table 1**). Consistent with the EMSA, maximal binding responses were similar for WT and A477T, indicating that the mutant binds to DNA at the same stoichiometry as WT (**Fig. 5a**). As the concentration dependence was non-Langmuir, the SPR binding isotherms were fit to the Hill equation. Fit Hill coefficients for WT were 1.83 ± 0.28 and 2.13 ± 0.26 for the Gilz and Pal-R GBSs, respectively (**Fig. 5b** and **Supplementary Table 1**). The A477T isotherms were well described by Hill coefficients of 1.34 ± 0.16 (Gilz) and 1.41 ± 0.1 (Pal-R) indicating that cooperativity was reduced but not abolished by the A477T mutation. As the Hill coefficient represents an exponential component of the Hill equation, these differences in Hill coefficients between the

two GBSs (0.3 for WT and 0.07 for A477T) are substantial for WT and diminished for A477T. This suggests that differential cooperativity contributes to the discrimination between binding sites for WT, but appears to play less of a role for A477T.

As the transcriptional activity of GR is affected by the fractional occupancy of the active dimer complex on a given response element, we also compared the dissociation kinetics of WT and A477T DBD – GBS complexes (**Fig. 5a**). To simplify the comparison between WT and A477T across the different binding sites, dissociation traces were fit to a single exponential decay process and the fit parameters were presented as $t_{1/2}$ values (**Supplementary Table 1**). For WT, fit $t_{1/2}$ values of 55 ± 4 s and 23 ± 2 s were determined for Pal-R and Gilz, respectively. In contrast, A477T dissociated from both Pal-R and Gilz GBSs with a $t_{1/2}$ of ~5 s; a 90% and 80% decrease, respectively, relative to WT. Under these conditions, the dissociation of WT was dependent on the GBS, whereas A477T kinetics appeared undiscriminating of sequence. This suggests that an intact dimerization surface is critical for interpreting GBS-specific signals that modulate GR dissociation.

To assess more broadly the biophysical parameters that might influence GBS-specific transcriptional activity, we extended our analysis to include five additional GBSs. Across all GBSs, A477T had lower affinity, faster dissociation and reduced cooperativity compared to WT (**Supplementary Table 1**). Interestingly, while Pal-R and Gha shared similar binding parameters, their transcriptional regulatory activities in reporters were not aligned (p-value = 0.004). To assess the relationship between DNA-binding properties and GBS-specific transcription, we compared transcriptional activity to the $K_{1/2}$ or $t_{1/2}$ values across this panel of GBSs (**Fig. 5c**). We found that transcriptional induction of GR WT did not vary as a simple function of DNA affinity. For example, a mutation resulting in reduced binding affinity enhanced

transcriptional induction at the Gha GBS (**Fig. 3a**). Additionally, GBSs with similar GR binding properties such as Gha and Pal-R or Fkbp5 and Gilz differed in transcriptional activity (**Fig. 1a**), as well as conformation (**Supplementary Fig. 4**). These results are consistent with those of Bain et al.[26], who also describe GBSs with binding affinities and regulatory activities that are not aligned. Unexpectedly, the authors of that work asserted that DNA binding affinity defines transcriptional activity at GBSs. While DNA binding affinity clearly plays a role in the activities of transcriptional regulators, our results as well as those of Bain et al. demonstrated that other factors must also contribute substantially. Having previously proposed that GBS-specific structural changes determine transcriptional activity[17,18], we assessed the relationship between binding and activity in the A477T mutant. In comparison to WT, A477T displayed a stronger correlation between transcription and affinity measured by multiple correlation analyses (**Supplementary Table 2**). Thus, disrupting the dimer interface appears to dampen allosteric signaling, rendering GBS affinity a stronger determinant.

A477T disrupts signaling between dimer partners

To further dissect the structural mechanism by which A477T alters GBS-specific activity, we used ^{15}N-HSQC to monitor the effects of the A477T mutation on GBS-specific conformations of GR. We focused on the lever arm region because its structure is sensitive to GBS[17], and found that, for WT, Gly470 displayed peak-splitting in a GBS-dependent manner (**Fig. 6a**)**.** Peak-splitting indicates two unique chemical environments for a single residue and can reflect two possibilities: (1) the GR dimer partners have non-equivalent conformations, or (2) GR dimers undergo slow conformational exchange between two distinct states. To distinguish

between these, we used ZZ-exchange NMR to detect conformational exchange. This experiment is similar to an HSQC except that a mixing period is introduced between recording the ^{15}N chemical shift and the ^{1}H chemical shift for each amide. Chemical exchange occurring during the mixing period is detected as cross-peaks in the NMR spectrum corresponding to the ^{15}N chemical shift of conformation A and the ^{1}H chemical shift of conformation B—and vice versa (**Supplementary Fig. 5a-b**). We individually fit the intensity of exchange cross-peaks at different mixing periods[27] (**Fig. 6b**) and determined that the conformational exchange rate of Gly470 (4.36 s^{-1} and 3.10 s^{-1}) is two orders of magnitude faster than the rate of dissociation from DNA (0.03 s^{-1}) (**Supplementary Fig. 5e**). This suggests that doublet peaks result from conformational exchange of DNA-bound GR dimer partners rather than dissociation from one half-site and subsequent re-binding binding to the adjacent half-site (**Supplementary Fig. 5b**).

Because the Gilz GBS consists of two non-identical half-sites, the simplest view is that Gly470 conformations may be determined solely by the sequence of the GBS half-site to which each GR dimer partner is bound. This would predict that GBSs with one identical half-site and that differ at the other half-site will have one overlaid Gly470 peak and one non-overlaid peak (compare Pal-R and Sgk). We found, however, that GBS complexes that are identical at one half-site had two Gly470 chemical shifts that were unique (Gilz, Sgk, Pal-ttg) (**Fig. 6a**). Thus, the lever arm conformation for each GR dimer depends not only on the sequence of the half-site to which it was directly bound, but also on the sequence bound by the adjacent dimer partner. Taken together with comparisons of Gly470 chemical shifts among different GBS complexes, our data suggest that each GR dimer partner integrates sequence-specific signals from both GBS half-sites as proposed in the model (**Fig. 6c**).

In contrast to WT, the Gly470 peaks were overlaid for A477T bound to different GBSs (**Fig. 6a**). Therefore, disrupting the dimer interface at Ala477 abolished the sequence-specific conformations of Gly470 within the lever arm. We expanded this comparison to investigate the extent to which each residue samples distinct conformations when bound to different GBSs by determining the chemical shift variance across a panel of seven GBSs (**Fig. 6d** and **Supplementary Fig. 6**). For GR WT, residues with considerable chemical shift variance included those in the recognition helix and lever arm, with the highest variance at Arg466, which makes a direct contact with the GBS. Consistent with the impact of A477T on Gly470, we found that A477T displayed reduced chemical shift variance compared to WT throughout the recognition helix and lever arm. Thus, this global chemical shift perturbation analysis supports our conclusion that the GBS-specific conformations of GR depend on an intact dimer interface.

Building upon NMR chemical shift mapping showing that GBS modulates the GR dimer interface, we conclude that the intact dimerization surface allows for allosteric communication between dimer partners and integration of sequence signals from the GBS as a whole. Consistent with this view, disruption of this communication by the A477T mutation reduced the sequence-specific effects of the GBS on GR conformation and simplified the relationship between GR DNA binding affinity and regulatory activity. Thus, we propose that signals are transmitted from the DNA-binding interface through the lever arm, the dimerization interface, and into domains of GR outside of the DBD that confer transcriptional regulatory activity[28,29].

Discussion

Genomic response elements are composed of combinations of sequence motifs that specify binding of distinct transcriptional regulators to execute gene-specific control of transcription. Even combinations of bases within a single binding motif can affect gene-specific activity[12]. Crystallographic study of the GR DBD revealed that the conformation of the "lever arm", a region between the DNA recognition helix and the dimer interface, differs at distinct GBSs[12]. Using NMR analysis of WT and mutant GR DBD – GBS complexes we defined both the origin and the consequence of the lever arm conformational transitions—a path of affected residues including parts of the DNA interface, the lever arm, and the dimerization interface that facilitates allosteric communication between GR dimer partners. This path enables integration of sequence-specific signals from both GBS half-sites, exponentially increasing the informational complexity of the GBS.

We considered the possibility that chemical shift differences in the DNA and dimer interface result from GBS-dependent reorientation of rigid GR dimer partners relative to the DNA or each other, e.g., from differential DNA bending. We found that the magnitude of chemical shift differences did not correlate with the proximity of GR DBD residues to the DNA (data not shown), contrary to the rigid-but-reoriented model. We assessed GBS-dependent DNA bending by FRET analysis. GR binding produced very small increases GBS FRET efficiency (**Supplementary Fig. 7**), suggesting minimal DNA bending. Furthermore, A477T behaved equivalent to WT among the five GBSs tested. Therefore, the chemical shift differences between GBS complexes and between WT and A477T represent GBS-specific differences in GR structure, not DNA bending.

Protein allostery is critical to concepts of combinatorial control. Structural studies of isolated nuclear receptor ligand binding domains (LBDs) suggest how signaling information

residing in small molecule ligands is transmitted to a coregulator recognition surface[8,30-34]. Here we show that GBSs drive structural changes at the DNA-binding interface that are coupled with changes in the GR dimerization interface and dimer partner, as well as correlated with distinct biophysical and transcriptional outcomes. Such structural transitions likely extend into distinct domains of intact GR to specify regulatory complex assembly and activity. These findings provide the clearest mechanistic perspective to date on functional studies showing that binding sequences modulate receptor interactions with coregulators[17,35,36], that ligands modulate interactions with DNA[37] and that both DNA and ligands direct interactions with coregulators[38]. Thus, the DBD residues identified here, together with LBD residues that interpret ligand signals[39] and affect gene-specific regulation[40-42] could begin to define a molecular "map" that, in the cellular context, integrates GBSs, ligands, chromatin, coregulators and post-transcriptional modifications to determine the composition and function of gene-specific transcriptional regulatory complexes.

Other than specific base contacts, what DNA signals might trigger changes in GR structure? Our data suggest that GR "measures" the spacer minor groove width as an indirect readout of spacer sequence. In the Pal complex, the Lys490 side chain reaches across the spacer minor groove to contact the phosphate backbone at the complement of spacer position 7. In the Fkbp5 complex, which has a wider spacer minor groove, Lys490 contacts the phosphate backbone of the proximal strand at position 11 (**Fig. 1e**). Indirect recognition of narrow minor grooves by insertion of positively charged side chains is a general feature of DNA recognition[20] and a contributor to specificity among transcription factors[43,44]. For GR, we observe a distinct mechanism, where minor groove width imposes structural constraints on lysine-mediated backbone contacts to DNA. Thus, flexible regions of the protein may adopt conformations that

accommodate differences in DNA shape, as has been demonstrated for other transcriptional regulators[45].

In addition to identifying regions of GR that can adopt distinct conformational states among different GBS complexes, we found that dimer partners undergo dynamic exchange between two discrete conformations while bound to a particular GBS (**Fig. 6a-b** and **Supplementary Fig. 5**). This is consistent with the structural asymmetry between dimer partners observed by crystallography[17]. Lever arm Gly470 and dimer interface Ile484 residues displayed dynamics with similar timescales, consistent with structural coupling between these two regions of the DBD. We speculate that GR – GBS complexes may differentially access conformational states that interact preferentially with particular transcriptional coregulators, thus providing "assembly instructions" for different regulatory complexes. GBS-specific dissociation kinetics may, in part, impact GR activity through altering the turnover of GR – DNA complexes. Indeed, interactions with response elements are highly dynamic, on the timescale of seconds[46], and regulatory complexes are actively and continuously disassembled[47]. How GR DNA-binding kinetics are regulated and how they impact transcriptional activity remain open questions.

Structural characterization of related nuclear receptors have shown that DNA-binding mediates conformational changes in the dimerization surface, providing a mechanism for cooperative dimerization[48]. We showed that GR cooperativity is affected by the precise binding sequence and impaired by the A477T mutation. Differential cooperativity is well-established as a mechanism for achieving gene-specific activity and suppressing transcriptional noise[49,50]. We propose that sequence-dependent conformational changes in the dimer interface modulate gene-specific cooperativity, in turn regulating the level of transcriptional activation by GR. While

this manuscript was in revision, Hudson et al.[51] reported that GR binds with negative cooperativity at "nGREs" where GR represses transcription. Thus, GBS-mediated allosteric regulation of cooperativity may enable GR to exhibit exceptional specificity of gene regulation ranging from transcriptional activation to repression.

Figure Legends:

Figure 1 Non-specific GBS bases modulate GR structure and activity. (**a**) The GR binding motif identified by GR ChIP-sequencing in U2OS-GR cells using MEME in "zero or one motif per site" mode with a 2nd order background Markov Model based on the top 1,000 peaks. Asterisks indicate specific-bases that are directly contacted by GR. (**b**) Luciferase induction of single GBS reporters with 100 nM dexamethsone (dex) treatment compared to ethanol control in U2OS cells. Error is s.d. of four or more independent experiments. Significant difference in transcriptional response by 2-tailed t-Test ($*p < 0.05$) is indicated for GBSs that differ only by spacer. (**c**) List of the GBS sequences used in this study with spacer (lowercase) and bases that differ from the palindromic Pal-R sequence (red). The 15-bp GBSs were centered within identical flanking sequences resulting in a 24-bp double-stranded DNA. (**d**) Alignment of GR DBD – Pal-F (PDB ID: 3g99, grey) and GR DBD – Fkbp5 (PDB ID: 3g6u, blue) crystal structures. (**e**) Zoomed view of Lys490 interaction with the GBS spacer of the Pal-F GBS (grey) and the Fkbp5 GBS (blue).

Figure 2 GBS spacer affects the conformation of the D-loop. (**a**) Comparison of ^{15}N-HSQC spectra of GR DBD – GBS complexes that differ at spacer only (Sgk and Sgk-ggg) or at spacer and half-site (Sgk and Gilz, Sgk-ggg and Gilz). (**b**) Zoomed spectra showing the chemical shift perturbation of D-loop residues Ala477 and Gly478 resulting from changing specific nucleotides in the spacer or half-site$_{13:15}$. (**c**) The magnitude of combined ^{1}H and ^{15}N chemical shift difference between GR DBD – Sgk and GR DBD – Sgk-ggg spectra for each assigned residue, colored onto the crystal structure (PDB ID:3g6u). Unassigned residues are colored white. (**d**)

Chemical shift difference (C.S.D.) analysis for pair wise comparison of GR DBD complexes with TTT compared to TTG spacer (top), or TGTTCT compared to TGTCCG half-site (bottom). Peaks unambiguously arising from peak splitting were assigned to their corresponding residues and C.S.D. values for both peaks are plotted. Grey bars indicate residues that have a C.S.D. greater or equal to the mean C.S.D. across pair wise comparisons. UA = unassigned.

Figure 3 Disrupting the dimerization interface affects lever arm and DNA recognition helix conformation. (**a**) Comparison of transcriptional induction of GBS luciferase reporters in U2OS cells expressing either GR wild type (WT) or mutant (A477T) after treatment with 100 nM dex. Error is one s.d. of four or more independent experiments and significant differences in transcriptional response between WT and A477T were determined by 2-tailed t-Test (*p < 0.05). The transcriptional response of A477T is equivalent for Sgk compared to Sgk-ggg or Sgk-ttg, (p-value = 0.44 and 0.19, respectively). (**b**) ^{15}N-HSQC comparing WT and A477T DBD bound to the Fkbp5 GBS. (**c**) Magnitude of combined ^{1}H and ^{15}N chemical shift difference between WT and A477T DBD bound to the Fkbp5 GBS, colored onto the WT DBD crystal structure (PDB ID 3g6u). Unassigned WT residues are colored white. (**d**) Chemical shift difference (C.S. Diff.) between spectra of WT DBD and A477T DBD complexes. Grey bars highlight residues with a chemical shift difference ≥0.05ppm between WT and A477T across all three GBSs. UA = unassigned.

Figure 4 A477T impairs dimerization but not monomer DNA-binding. Electrophoretic mobility shift assay (EMSA) monitoring binding of WT or A477T DBD to (**a**) GBSs conjugated

with an Alexa-488 fluorophore at a concentration of 5 nM or to (**b**) mutated GBSs, where one half-site is changed to the least favorable nucleotide at each position, based on the ChIP-seq binding motif in Fig. 1a. (**c**) Quantitative comparison of WT DBD (open squares) or A477T DBD (filled triangles) binding by EMSA. Error bars are one s.d. for 2-4 replicates.

Figure 5 A477T disrupts cooperativity and GBS-specific dissociation. (**a**) Representative SPR binding traces for WT and A477T DBD binding to immobilized GBSs at 35˚C. Comparison of dissociation curves for WT DBD (top right) and A477T DBD (bottom right) (**b**) Binding isotherms for a GR DBD concentration series (0.700–200 nM for WT, 1.4–400 nM for A477T) binding to immobilized GBS surfaces at 35°C from 3 separate titrations, normalized to maximal binding. Isotherms were fit to the Hill equation: Fractional occupancy = $(c/(c+K_{1/2}))^{nH}$. (**c**) Comparison of transcriptional activity (fold induction) versus binding affinity ($K_{1/2}$) or dissociation half-life ($t_{1/2}$), for WT DBD (blue) and A477T DBD (red) across seven GBS surfaces at 8˚C.

Figure 6 Sequence-specific lever arm conformation is dependent on the intact dimerization interface. (**a**) ^{15}N-HSQC zoomed on the Gly470 peak of the lever arm of WT DBD bound to the asymmetric GBSs (Gilz, Sgk, Pal-ttg) and a palindromic site (Pal-R). Overlay of Gly470 peaks from all for GBSs for WT and A477T DBD (far right). (**b**) The peak intensity for Gly470 and Ile484 WT DBD residues in ZZ-exchange spectra at five mixing periods. The auto peak and the corresponding exchange peak are shaded equivalently. Insets show the spectra for auto and exchange peaks at a mixing period of 0.2 s. (**c**) Model showing the role of the dimerization

interface in defining GBS-specific conformations. Both half-sites and the spacer determine the conformation of each GR dimer partner by transmitting information from the adjacent GBS half-site across the dimerization interface (represented by the colored arrows). Disruption of the dimerization interface by A477T impairs signal transmission and results in lever arm conformations that are insensitive to GBS sequence. (**d**) Comparison of WT and A477T chemical shift variance among seven GBS complexes colored by amino acid onto the GR DBD crystal structure.

Figure 1 Non-specific GBS bases modulate GR structure and activity

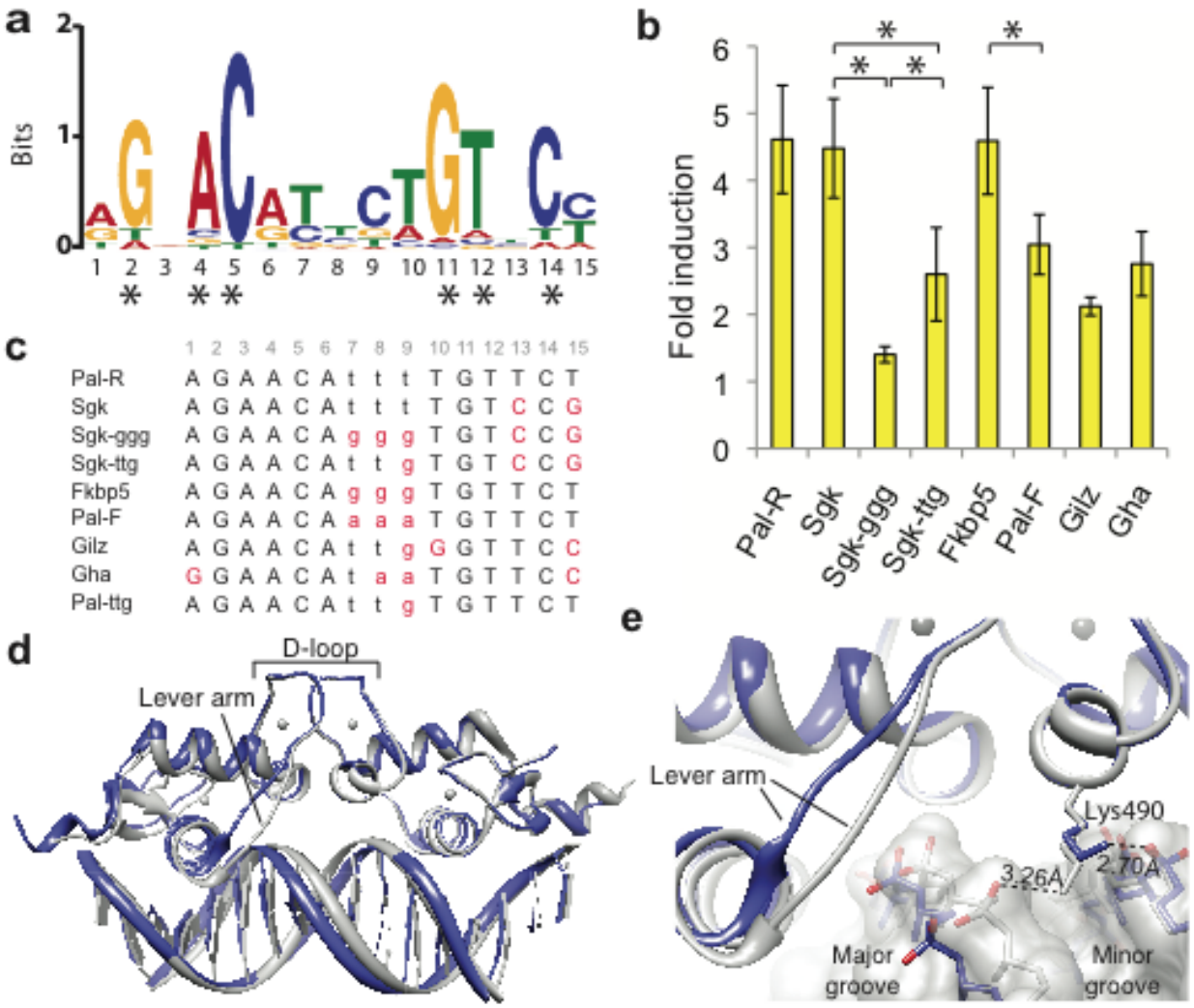


Figure 2 GBS spacer affects the conformation of the D-loop

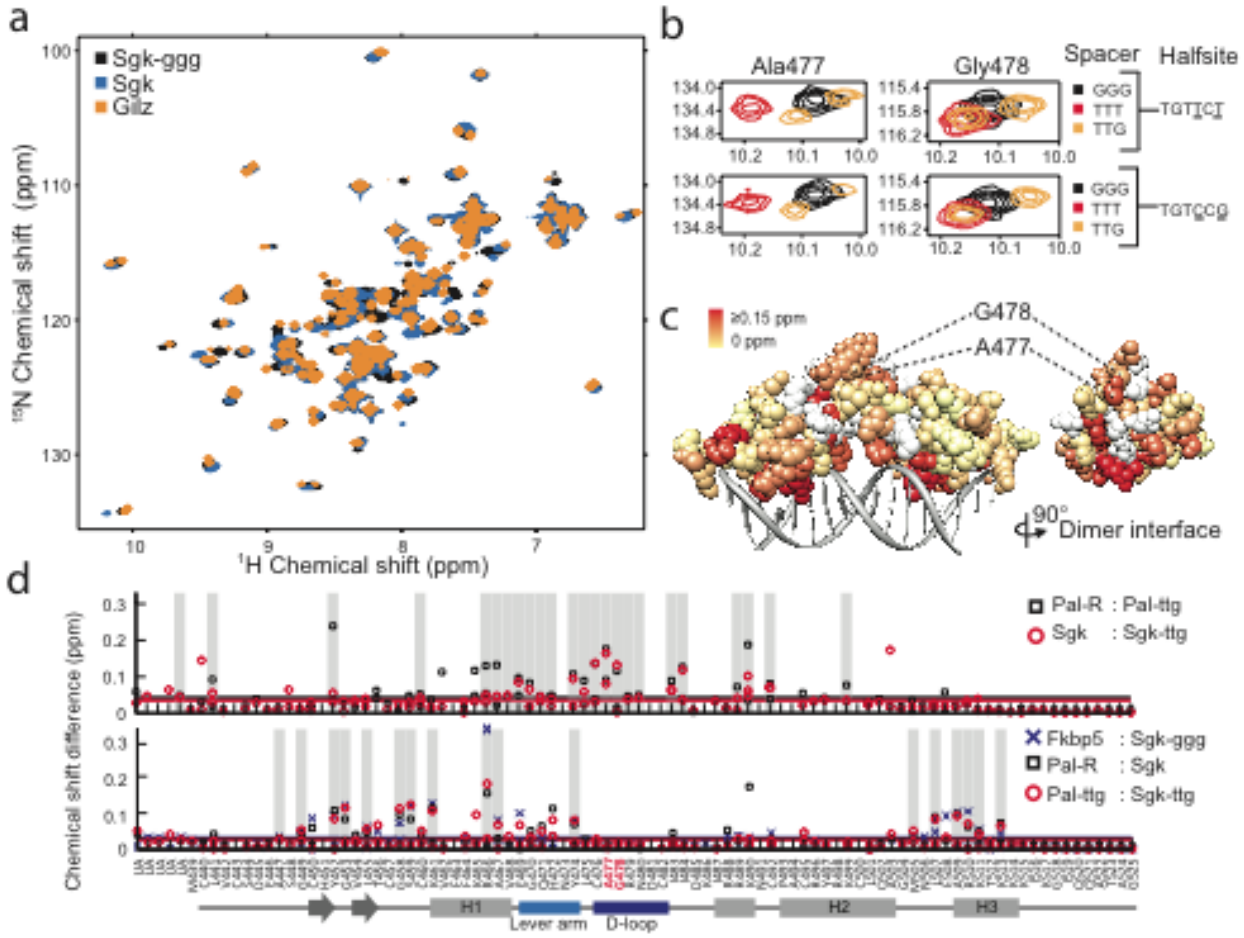


Figure 3 Disrupting the dimerization interface affects lever arm and DNA recognition helix conformation

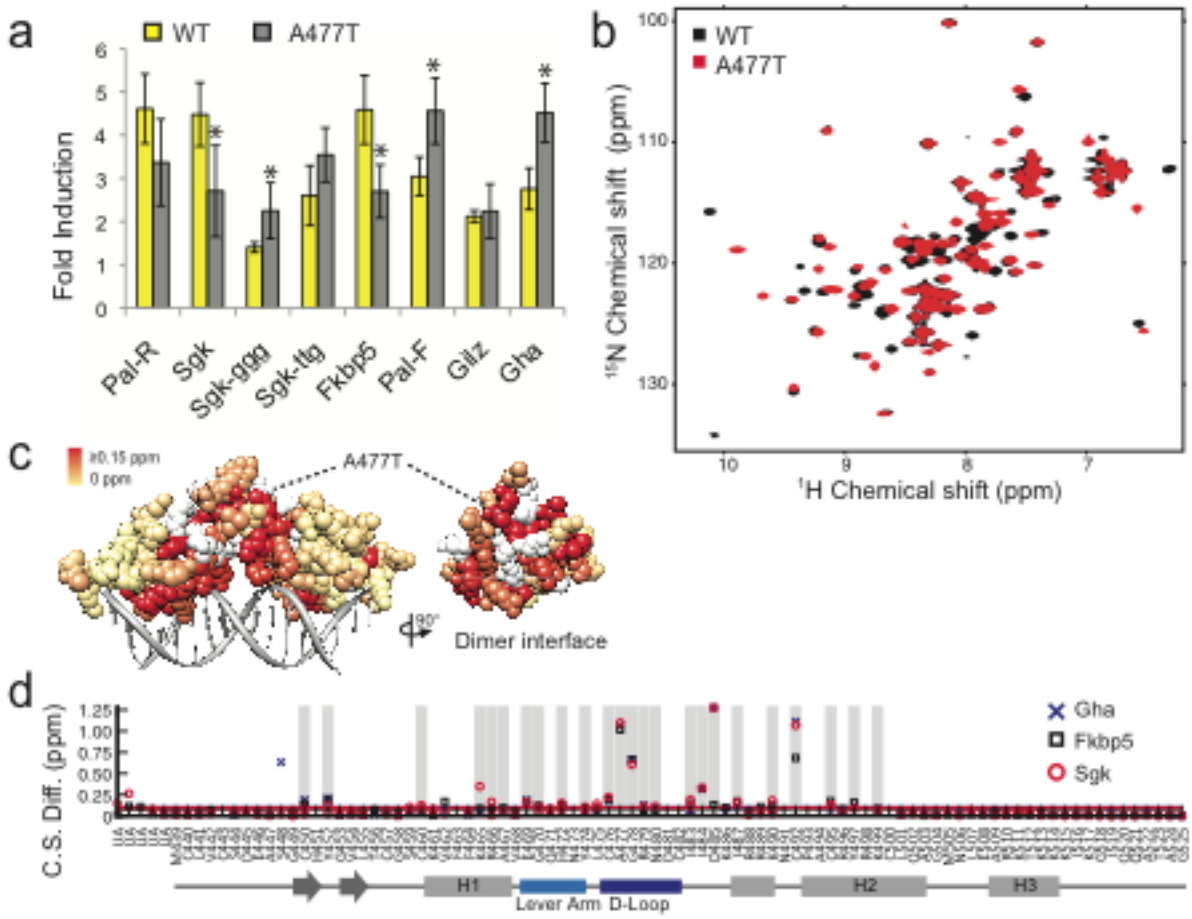


Figure 4 A477T impairs dimerization but not monomer DNA-binding

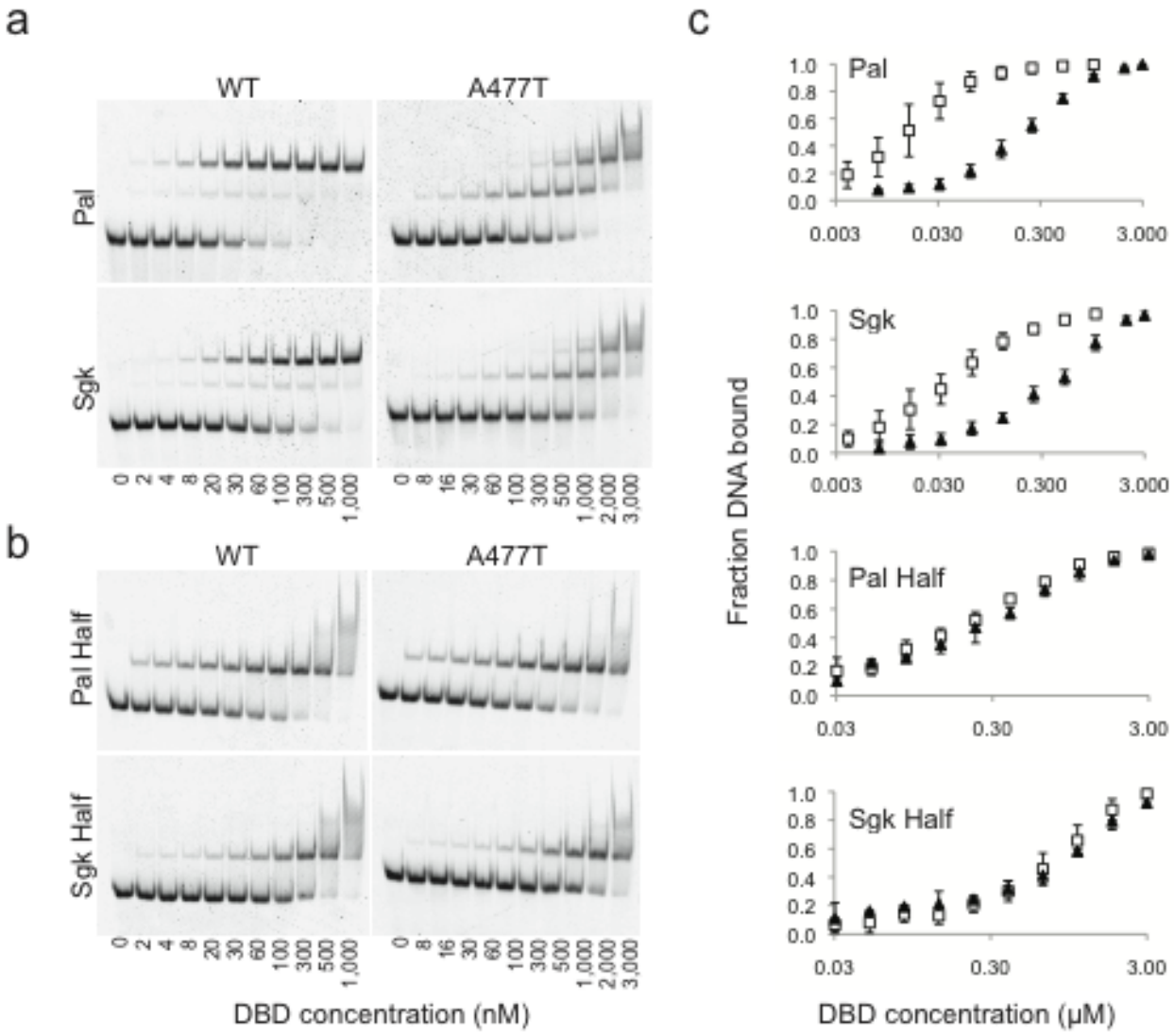


Figure 5 A477T disrupts cooperativity and GBS-specific dissociation

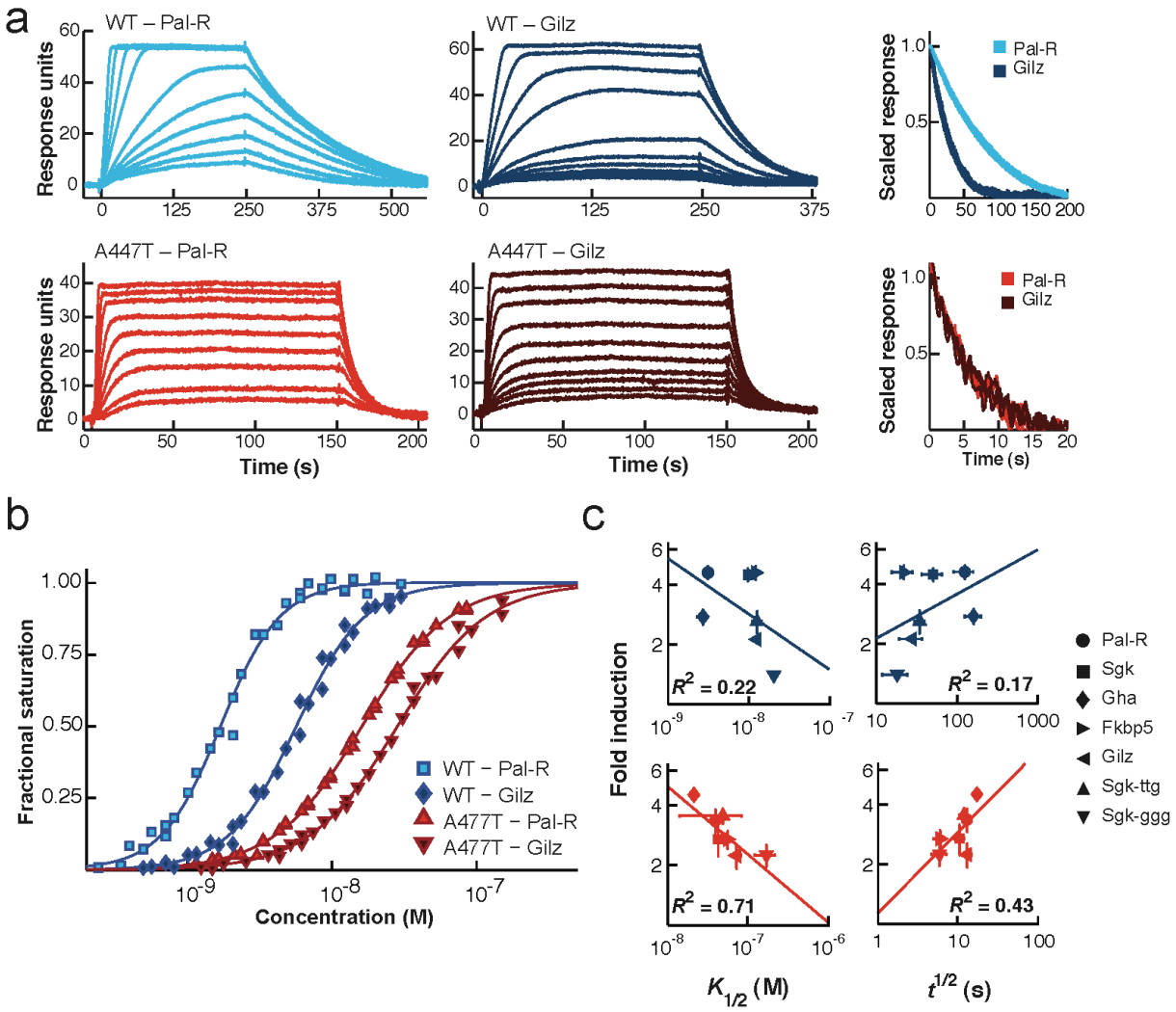
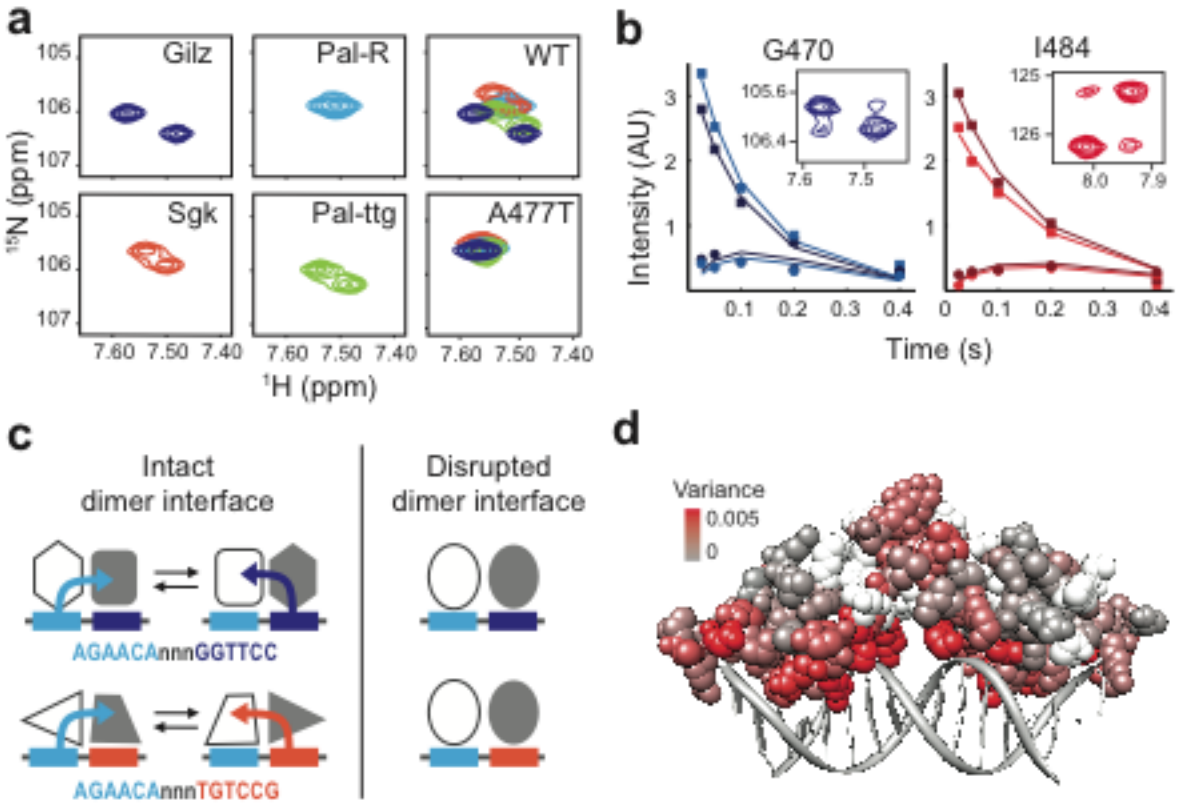


Figure 6 Sequence-specific lever arm conformation is dependent on the intact dimerization interface.



Acknowledgements:

We thank M. Kelly (UCSF) and J. Pelton (UC Berkeley) for NMR support; J. Cheney and S. Floor for providing scripts for NMR analysis; and S. Cooper and S. Meijsing for developing GR ChIP-seq protocols. We thank R. Fletterick, A. Johnson, G. Narlikar and members of the Yamamoto Lab for thoughtful discussions; and M. Knuesel and S. Meijsing for critical reading of the manuscript. This work was supported by US National Institutes of Health (NIH) grant CA020535 (K.R.Y), the Biophysics Graduate Group training grant NIHT32GM008284 (K.M.K), NIH grants GM08537 and 5K99CA149088 (M.A.P.), the Leukemia and Lymphoma Society Fellowship (M.A.P), the Larry S. Hillblom Foundation Fellowship (L.C.W), the Genentech and Sandler Foundation Graduate Fellowship (L.C.W), and the Cancer Research Coordinating Committee Fellowship (L.C.W).

Author Contributions:

L.C.W. designed, performed and analyzed experiments. J.D.G. and L.C.W. designed and performed experiments for NMR sequential assignment. K.M.K. and L.C.W. designed, performed, and analyzed SPR and FRET experiments. B.J.S. and L.C.W. designed and performed ChIP-seq experiments and B.J.S. analyzed the data. M.A.P. and K.R.Y. contributed to the direction of project and L.C.W. and K.R.Y. wrote the manuscript.

References

1. Yamamoto, K., Darimont, B., Wagner, R. & Iniguez-Lluhi, J. Building transcriptional regulatory complexes: signals and surfaces. *Cold Spring Harbor Symposia on Quantitative Biology* **63**, 587–598 (1998).
2. Rosenfeld, M. G., Lunyak, V.V. & Glass, C. K. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes & Development* **20**, 1405–1428 (2006).
3. Ricketson, D., Hostick, U., Fang, L., Yamamoto, K. & Darimont, B. A conformational switch in the ligand-binding domain regulates the dependence of the glucocorticoid receptor on Hsp90. *J Mol Biol* **368**, 729–741 (2007).
4. Garza, A. M. S., Khan, S. H. & Kumar, R. Site-specific phosphorylation induces functionally active conformation in the intrinsically disordered N-terminal activation function (AF1) domain of the glucocorticoid receptor. *Mol Cell Biol* **30**, 220–230 (2010).
5. Bledsoe, R. K. *et al.* Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110**, 93–105 (2002).
6. Frego, L. Conformational changes of the glucocorticoid receptor ligand binding domain induced by ligand and cofactor binding, and the location of cofactor binding sites determined by hydrogen/deuterium exchange mass spectrometry. *Protein Science* **15**, 722–730 (2006).
7. Schoch, G. A. *et al.* Molecular switch in the glucocorticoid receptor: active and passive antagonist conformations. *J Mol Biol* **395**, 568–577 (2010).
8. Kauppi, B. *et al.* The three-dimensional structures of antagonistic and agonistic forms of the glucocorticoid receptor ligand-binding domain: RU-486 induces a transconformation that leads to active antagonism. *J Biol Chem* **278**, 22748–22754 (2003).
9. Shah, N. & Scanlan, T. S. Design and evaluation of novel nonsteroidal dissociating glucocorticoid receptor ligands. *Bioorganic & Medicinal Chemistry Letters* **14**, 5199–5203 (2004).
10. Diamond, M. I., Miner, J. N., Yoshinaga, S. K. & Yamamoto, K. R. Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science* **249**, 1266–1272 (1990).
11. Rogatsky, I., Waase, C. L. & Garabedian, M. J. Phosphorylation and inhibition of rat glucocorticoid receptor transcriptional activation by glycogen synthase kinase-3 (GSK-3). Species-specific differences between human and rat glucocorticoid receptor signaling as revealed through GSK-3 phosphorylation. *J Biol Chem* **273**, 14315–14321 (1998).
12. So, A. Y.-L., Chaivorapol, C., Bolton, E. C., Li, H. & Yamamoto, K. R. Determinants of cell- and gene-Specific transcriptional regulation by the glucocorticoid receptor. *PLoS Genet* **3**, e94 (2007).
13. La Baer, J. & Yamamoto, K. R. Analysis of the DNA-binding affinity, sequence specificity and context dependence of the glucocorticoid receptor zinc finger region. *J Mol Biol* **239**, 664–688 (1994).
14. Baumann, H. *et al.* Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* **32**, 13463–13471 (1993).
15. Härd, T. *et al.* Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* **249**, 157–160 (1990).

16. Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
17. Meijsing, S. H. *et al.* DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–410 (2009).
18. Lefstin, J. A. & Yamamoto, K. R. Allosteric effects of DNA on transcriptional regulators. *Nature* **392**, 885–888 (1998).
19. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves. *Nucleic Acids Res* **37**, 5917–5929 (2009).
20. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248–1253 (2009).
21. Zhuravleva, A. & Gierasch, L. M. Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones. *Proc Natl Acad Sci USA* **108**, 6987–6992 (2011).
22. Selvaratnam, R., Chowdhury, S., VanSchouwen, B. & Melacini, G. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc Natl Acad Sci USA* **108**, 6133–6138 (2011).
23. Masterson, L. R., Mascioni, A., Traaseth, N. J., Taylor, S. S. & Veglia, G. Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci USA* **105**, 506–511 (2008).
24. Heck, S. *et al.* A distinct modulating domain in glucocorticoid receptor monomers in the repression of activity of the transcription factor AP-1. *EMBO* **13**, 4087–4095 (1994).
25. Floor, S. N., Borja, M. S. & Gross, J. D. Interdomain dynamics and coactivation of the mRNA decapping enzyme Dcp2 are mediated by a gatekeeper tryptophan. *Proc Natl Acad Sci USA* **109**, 2872–2877 (2012).
26. Bain, D. L. *et al.* Glucocorticoid receptor-DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity. *J Mol Biol* **422**, 18–32 (2012).
27. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* **4**, 727–734 (1994).
28. Iñiguez-Lluhí, J. A., Lou, D. Y. & Yamamoto, K. R. Three amino acid substitutions selectively disrupt the activation but not the repression function of the glucocorticoid receptor N terminus. *J Biol Chem* **272**, 4149–4156 (1997).
29. Darimont, B. D. *et al.* Structure and specificity of nuclear receptor-coactivator interactions. *Genes & Development* **12**, 3343–3356 (1998).
30. Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H. & Moras, D. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-a. *Nature* **375**, 377–382 (1995).
31. Egea, P. F. *et al.* Crystal structure of the human RXRα ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. *EMBO J* **19**, 2592–2601 (2000).
32. Lusher, S. J. *et al.* Structural basis for agonism and antagonism for a set of chemically related progesterone receptor modulators. *J Biol Chem* **286**, 35079–35086 (2011).
33. Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
34. Brzozowski, A. M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–757 (1997).
35. Hall, J. M., McDonnell, D. P. & Korach, K. S. Allosteric regulation of estrogen receptor

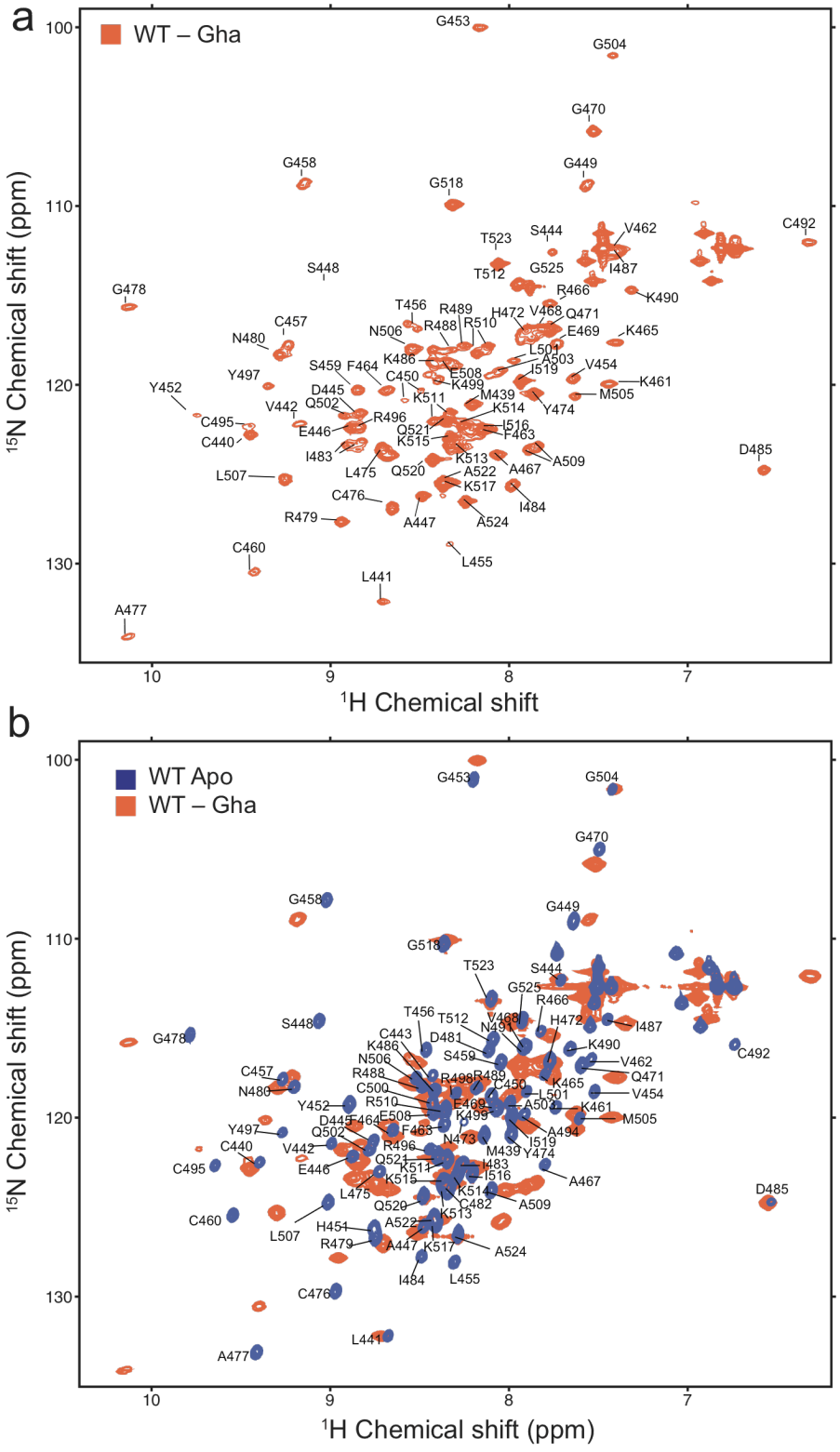
structure, function, and coactivator recruitment by different estrogen response elements. *Mol Endocrinol* **16**, 469–486 (2002).
36. Engel, K. B. & Yamamoto, K. R. The glucocorticoid receptor and the coregulator Brm selectively modulate each other's occupancy and activity in a gene-specific manner. *Mol Cell Biol* **31**, 3267–3276 (2011).
37. Wang, J.-C. *et al.* Novel arylpyrazole compounds selectively modulate glucocorticoid receptor regulatory activity. *Genes & Development* **20**, 689–699 (2006).
38. Zhang, J. *et al.* DNA binding alters coactivator interaction surfaces of the intact VDR–RXR complex. *Nat Struct Mol Biol* **18**, 556–563 (2011).
39. Shulman, A. I., Larson, C., Mangelsdorf, D. J. & Ranganathan, R. Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell* **116**, 417–429 (2004).
40. Rogatsky, I. Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor. *Proc Natl Acad Sci USA* **100**, 13845–13850 (2003).
41. Tao, Y.-G., Xu, Y., Xu, H. E. & Simons, S. S. Mutations of glucocorticoid receptor differentially affect AF2 domain activity in a steroid-selective manner to alter the potency and efficacy of gene induction and repression. *Biochemistry* **47**, 7648–7662 (2008).
42. Lee, G.-S. & Simons, S. S., Jr. Ligand binding domain mutations of the glucocorticoid receptor selectively modify the effects with, but not binding of, cofactors. *Biochemistry* **50**, 356–366 (2011).
43. Joshi, R. *et al.* Functional specificity of a hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530–543 (2007).
44. Slattery, M. *et al.* Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
45. Scully, K. M. *et al.* Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290**, 1127–1131 (2000).
46. McNally, J. G., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**, 1262–1265 (2000).
47. Stavreva, D. A., Müller, W. G., Hager, G. L., Smith, C. L. & McNally, J. G. Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol Cell Biol* **24**, 2682–2697 (2004).
48. Holmbeck, S. M., Dyson, H. J. & Wright, P. E. DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J Mol Biol* **284**, 533–539 (1998).
49. Ackers, G. K., Johnson, A. D. & Shea, M. A. Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci USA* **79**, 1129–1133 (1982).
50. Robblee, J. P., Miura, M. T. & Bain, D. L. Glucocorticoid receptor–promoter interactions: energetic dissection suggests a framework for the specificity of steroid receptor-mediated gene regulation. *Biochemistry* **51**, 4463–4472 (2012).
51. Hudson, W. H., Youn, C. & Ortlund, E. A. The structural basis of direct glucocorticoid-mediated transrepression. *Nat Struct Mol Biol* **20**, 53–58 (2013).

Supplementary Materials

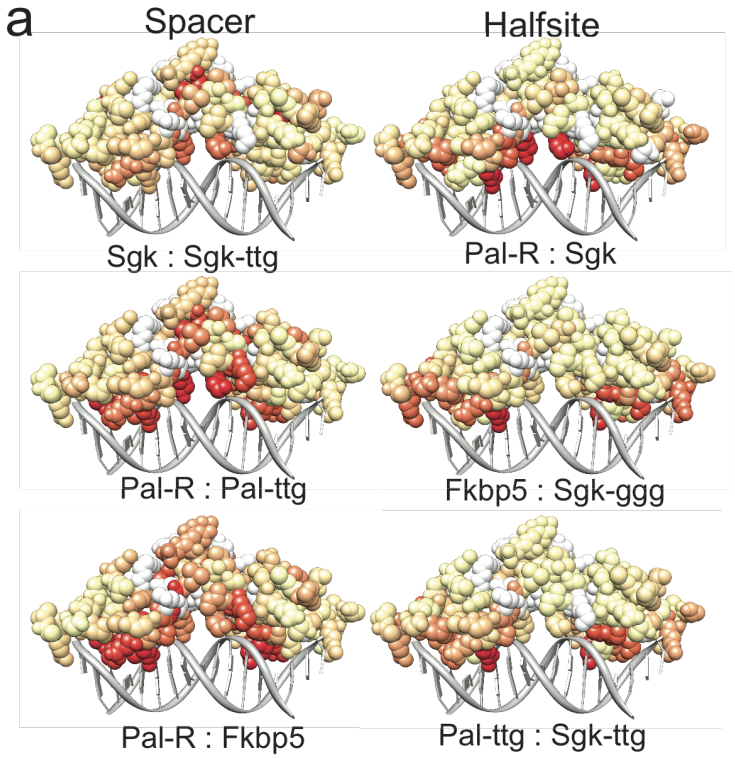
The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals

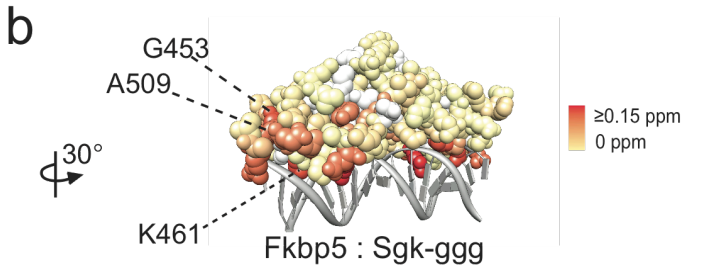
Lisa C. Watson[1,2], Kristopher M. Kuchenbecker[3,4], Benjamin J. Schiller[1,2], John D. Gross[5], Miles A. Pufall[2,6], Keith R. Yamamoto[2]

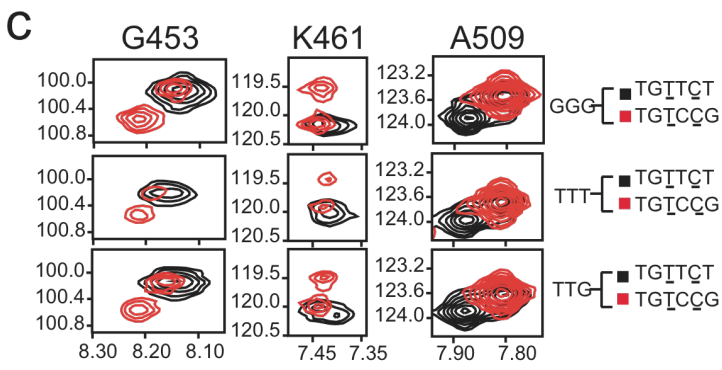
Supplementary Figure 1 NMR assignment of the DNA-bound GR DBD complex. (**a**) HSQC of $^{2}H^{15}N$-lableled GR wild type (WT) DBD bound to the Gha GBS at 35°C. (**b**) Overlay of the ^{15}N-HSQC spectra of the unbound WT DBD (WT Apo) and WT DBD – Gha complex at 25°C.



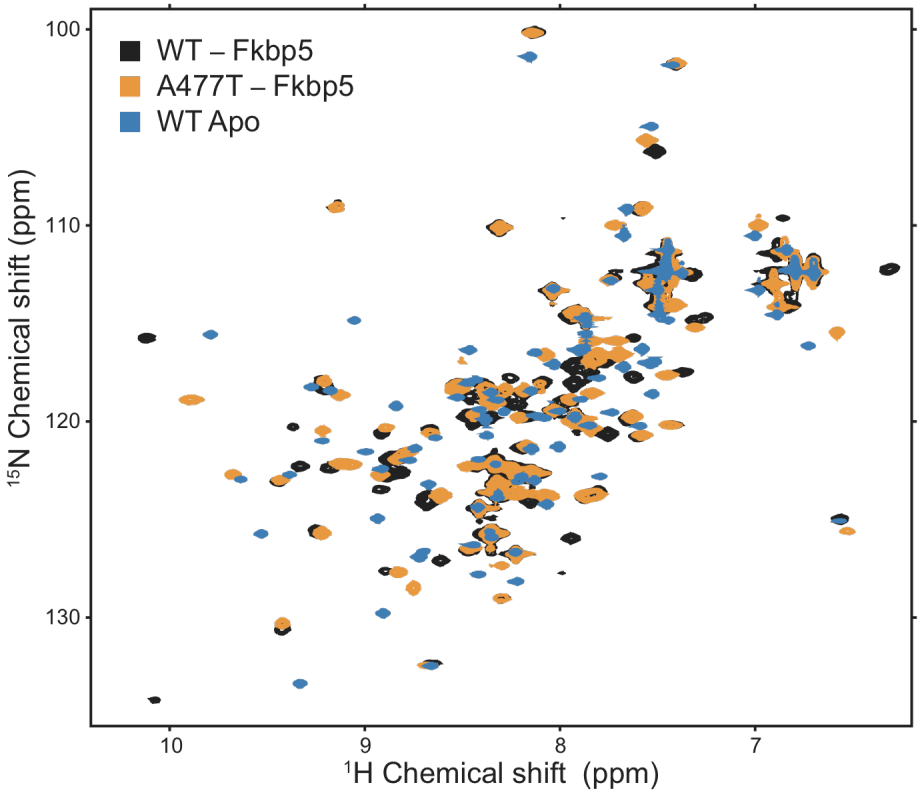
Supplementary Figure 2 NMR chemical shift difference analysis of GBSs that differ in the half-site or spacer sequence. (**a**) Pairwise comparison of six GBS complexes showing the effect of changing either spacer or half-site positions 13 and 15 (half-site$_{13:15}$). NMR chemical shift differences are mapped on to the crystal structure of GR DBD (PDB ID:3g6u). (**b**) The locations within the GR DBD and (**c**) the overlay of ^{15}N-HSQC peaks for residues that differ in half-site$_{13:15}$.



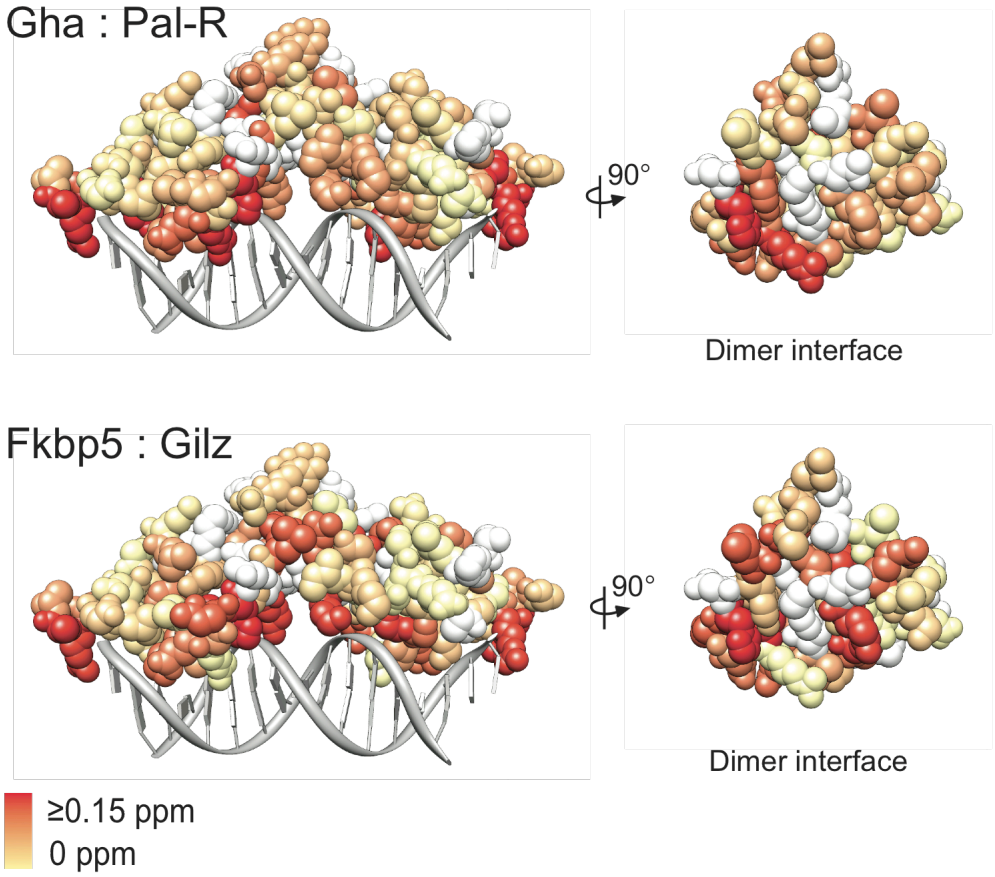




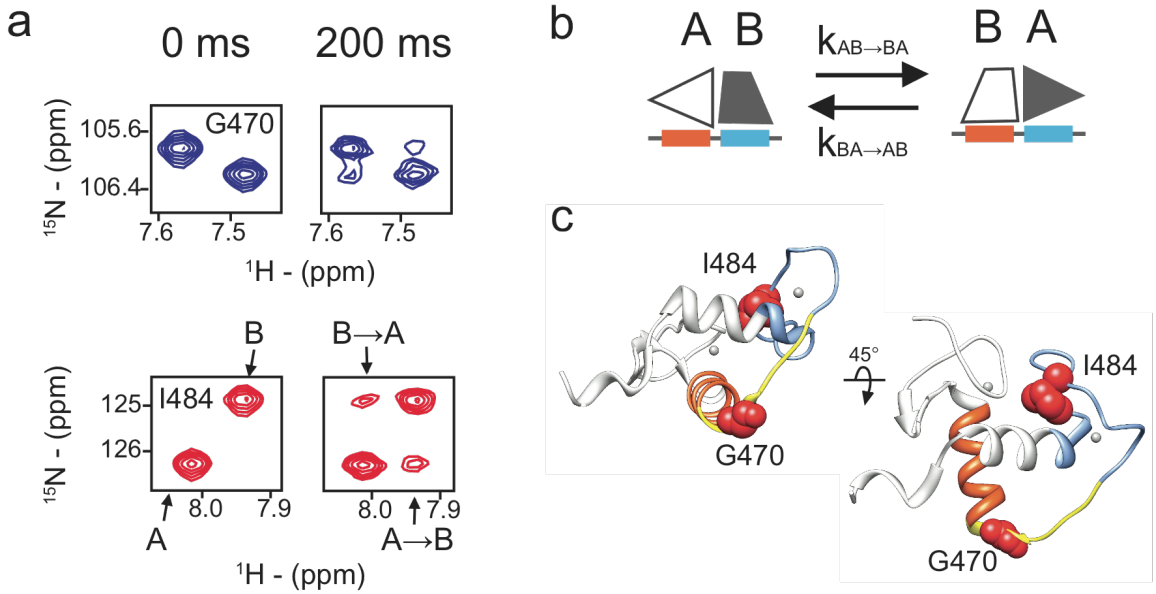
Supplementary Figure 3 DNA-bound A477T conformation differs from that of the unbound WT DBD. The overlay of ^{15}N-HSQC spectra of WT and A477T DBD bound to the Fkbp5 GBS, compared to the unbound DBD (WT Apo) at 35°C.

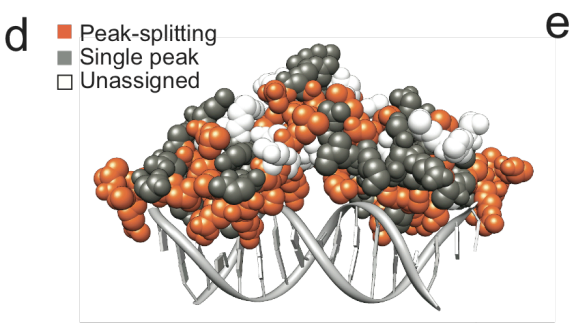


Supplementary Figure 4 Chemical shift difference between WT DBD – GBS complexes with different levels of transcriptional induction by GR. Structures are colored according to the magnitude of combined ^{1}H and ^{15}N NMR chemical shift differences for pairs of GBS complexes that have similar binding affinity, cooperativity and dissociation kinetics. Unassigned residues are colored white.



Supplementary Figure 5 Conformational exchange within the lever arm and dimer interface. (**a**) Zoomed view of ZZ-exchange NMR spectra for Gly470 and Ile484 with a 200 ms mixing period. The peaks indicate two conformational states, arbitrarily labeled A and B, at 0 ms mixing period and the exchange cross-peaks are labeled A⇄B and B⇄A. (**b**) Model for conformational swapping between dimer partners while bound to DNA, based on the observed rate of conformational exchange (k_A and k_B), which is faster than the measured GR dissociation rate from DNA (0.03 s^{-1}). (**c**) The location of Gly470 and Ile484 (red spheres) within the GR DBD monomer. Recognition helix (orange), lever arm (yellow), and zinc finger 2 of the dimer interface (light blue). (**d**) Mapping of ^{15}N-HSQC peaks of the WT DBD – Gilz complex that undergo peak-spitting from slow exchange. (**e**) The parameters from curves fit to ZZ-exchange peaks, where *R* = longitudinal relaxation rate of magnetization, *k* = exchange rate for converting from site A to B, and *I(0)* = longitudinal magnetization at the start of the mixing period[27].

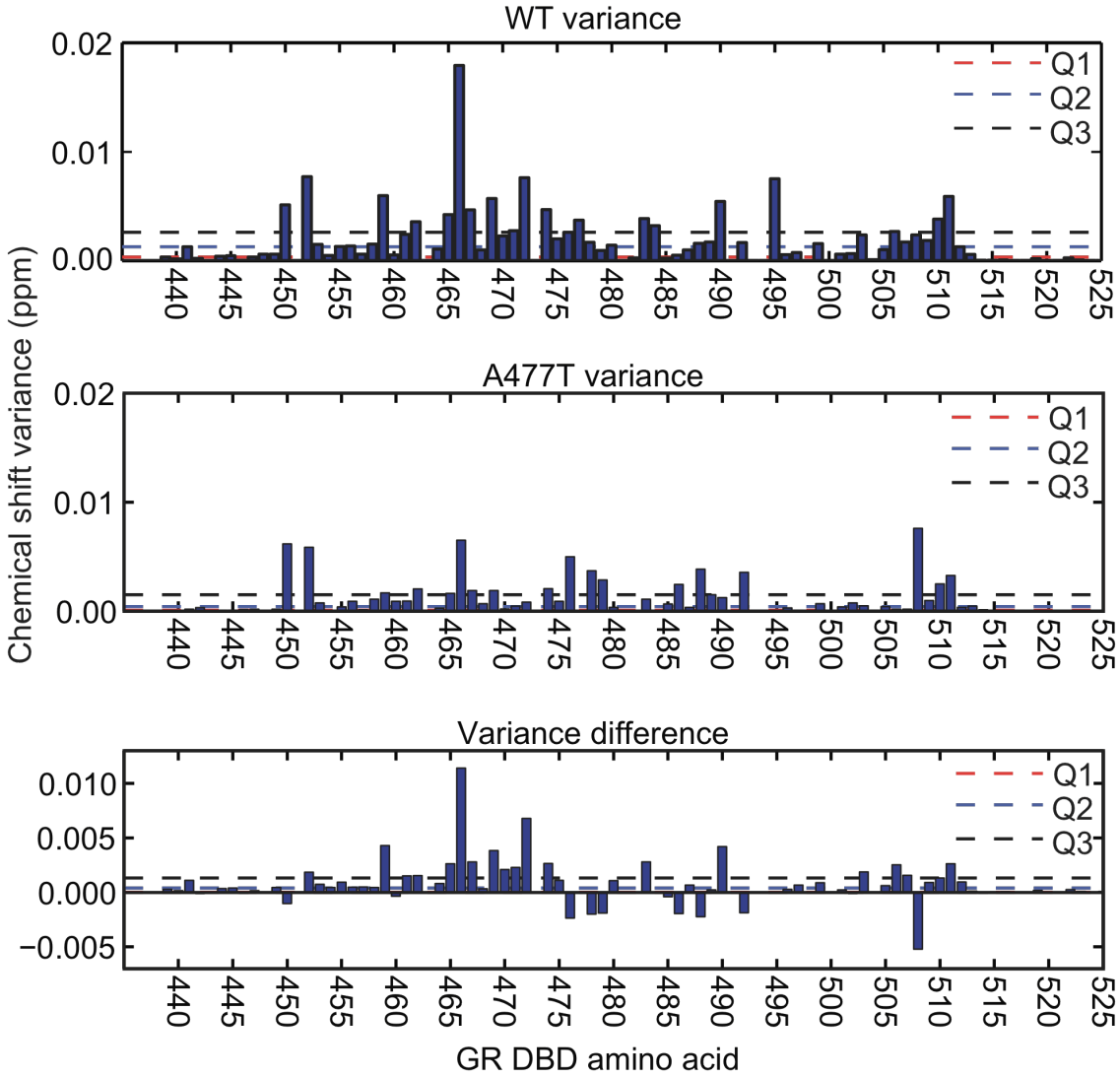




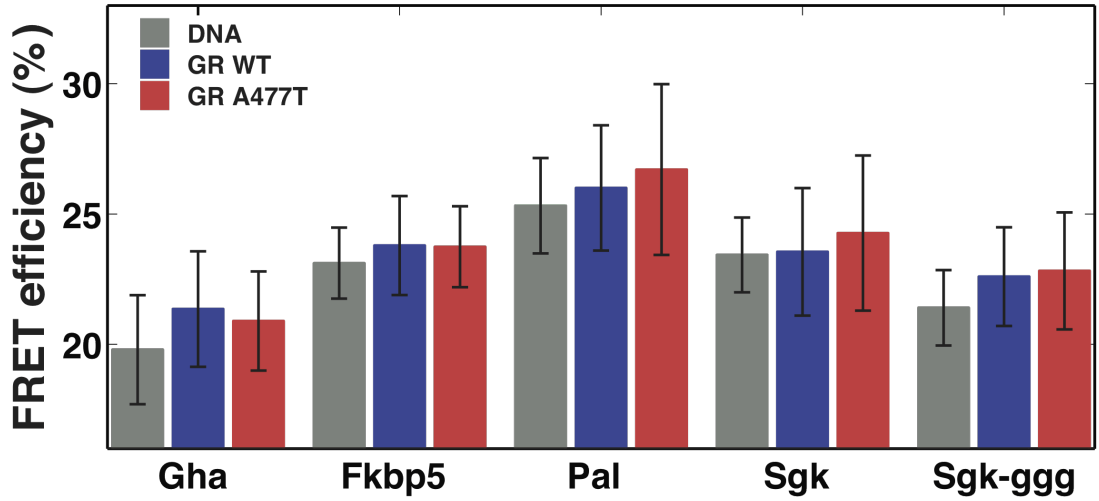
e

	G470	I484
R_A	5.11 s^{-1}	3.92 s^{-1}
R_B	6.63 s^{-1}	4.71 s^{-1}
k_A	4.36 s^{-1}	2.31 s^{-1}
k_B	3.10 s^{-1}	2.13 s^{-1}
$I_A(0)$	3.47 x 10^6	2.82 x 10^6
$I_B(0)$	4.15 x 10^6	3.53 x 10^6

Supplementary Figure 6 NMR chemical shift variance across seven different GBS complexes. WT variance and A477T variance among ^{15}N-HSQC spectra for each amino acid calculated as σ^2_H or $\sigma^2_N = \sum(x - \mu)^2/(n\text{-}1)$, where x = ^{1}H or ^{15}N chemical shift (ppm), μ = the mean chemical shift (ppm), and n = 7; and combined as the variance sum: $\sigma^2_{NH} = (1/5 \bullet \sigma^2_N) + \sigma^2_H$ (top and middle panel). The difference between WT and A477T chemical shift variance (bottom panel). The dotted lines represent the first quartile (Q1), median (Q2) the third quartile (Q3). Unassigned amino acids are plotted as zero.



Supplementary Figure 7 Comparison of sequence-specific DNA bend as determined by FRET assay. Alexa488 and Alexa546 fluorophores were conjugated to opposite 3' ends of 24-bp GBS-containing oligos, serving as the donor and acceptor, respectively. FRET was measured for five GBSs at 50 nM with and without GR DBD (500 nM WT or A477T). By assuming a bend at the center of the GBS, GR binding-induced DNA bend is less than 1° for all GBSs, using the formula $R = R_0[(1/E)-1]^{1/6}$, where R = distance, E = efficiency and R_0 = Förster distance of the FRET pair. All FRET assays were performed in NMR buffer using a Molecular Devices SpectraMax M5 with a fixed excitation of 444 nM and an emission scan from 500-650 nm. Equivalent results were obtained using 1 μM protein, indicating saturated binding under these conditions. Data is shown as the mean ± two s.d. from 12 replicates.



	FRET Efficiency			p-value		
GBS	DNA	WT	A477T	DNA vs. WT	DNA vs. A477T	WT vs. A477T
Gha	19.8±2.1	21.4±2.2	20.9±1.9	0.0018	0.0133	0.2873
Fkbp5	23.1±1.4	23.8±1.9	23.7±1.6	0.0576	0.0470	0.8929
Pal	25.3±1.8	26.0±2.4	26.7±3.3	0.1312	0.0177	0.2423
Sgk	23.4±1.4	23.6±2.4	24.3±3.0	0.7697	0.0917	0.2089
Sgk-ggg	21.4±1.4	22.6±1.9	22.8±2.2	0.0022	0.0013	0.6107

Supplementary Table 1 Summary of SPR fit parameters for WT and A477T DBD binding to GBS surfaces at 35°C or 8°C. Dissociation constant, $K_{1/2}$, and Hill coefficient, n_H, were calculated from equilibrium binding isotherms fit to a the Hill equation. Fitting of the dissociation phases of the SPR traces by conventional models gave systematic residuals consistent with a biphasic-sequential dissociation process, likely resulting from the complex kinetics and strong cooperativity of this system. Instead, the stabilities of the different complexes were reliably determined by a standard exponential decay model, and $t_{1/2}$ values were extrapolated from the apparent off-rates. Errors are the mean ± two s.d. from 2-3 replicates.

	WT			A477T		
GBS	**$K_{1/2}$ (nM)**	**n_H**	**$t_{1/2}$ (s)**	**$K_{1/2}$ (nM)**	**n_H**	**$t_{1/2}$ (s)**
	35°C			35°C		
Pal-R	1.6 ± 0.2	2.1 ± 0.3	55 ± 4	16 ± 1	1.4 ± 0.1	4.8 ± 0.7
Gilz	5.7 ± 0.6	1.8 ± 0.3	23 ± 2	28 ± 1	1.3 ± 0.2	4.7 ± 0.8
	8°C			8°C		
Gha	2.7 ± 0.2	2.2 ± 0.2	160 ± 20	21 ± 1	1.2 ± 0.05	18 ± 1
Pal-R	3.1 ± 0.1	2.1 ± 0.1	120 ± 20	39 ± 1	0.9 ± 0.03	13 ± 0.8
Sgk	9.7 ± 0.2	1.8 ± 0.06	50 ± 7	42 ± 1	1.0 ± 0.02	10 ± 0.7
Fkbp5	12 ± 0.2	2.0 ± 0.06	22 ± 3	55 ± 1	1.4 ± 0.02	5.9 ± 0.5
Gilz	13 ± 0.2	1.9 ± 0.05	28 ± 4	71 ± 5	1.1 ± 0.03	7.6 ± 0.7
Sgk-ttg	13 ± 0.8	2.3 ± 0.2	34 ± 2	48 ± 17	1.4 ± 0.02	12 ± 0.7
Sgk-ggg	20 ± 0.7	1.9 ± 0.06	18 ± 3	166 ± 23	1.1 ± 0.05	5.8 ± 0.6

Supplementary Table 2 Correlation between transcriptional activity and DNA-binding affinity ($K_{1/2}$) or half-life ($t_{1/2}$). Correlations for WT and A477T were calculated for seven different GBS complexes. Because this relationship is expected to be non-linear[26], a statistical analysis with non-linear fitting of data are presented.

	Wildtype		A477T	
Analysis	**$K_{1/2}$**	**$t_{1/2}$**	**$K_{1/2}$**	**$t_{1/2}$**
R^2 (log transformed)	0.22	0.17	0.71	0.43
Pearson's (R^2)	0.31	0.06	0.44	0.54

Online Methods:

Protein expression and purification

Expression vector pET28a containing rat GR WT DBD, residues 440-525, has been described previously[17]. Vector for A477T DBD was derived by PCR site-directed mutagenesis. BL21 Gold E. coli cells were grown in 50mL LB media to optical density of ~0.5-1.0, then pelleted and resuspended in 1L minimal media containing 2 g/L $^{15}NHCl_4$ as the only nitrogen source. Cultures were grown to an optical density of ~0.7 and expression was induced with 0.5 mM IPTG for ~16 hours at 18°C or 8 hours at 30°C (both produced equivalent spectra). Cells were pelleted and resuspended at 40 mL/L cells in lysis buffer containing 50 mM Tris pH 7.5, 500 mM NaCl and 15 mM Imidazole, then frozen in liquid nitrogen, and stored at -80°C. Cells were lysed using an EmulsiFlex C5 homogenizer. Lysate was centrifuged for 45 min at 40,000 rpm at 4°C and run over a Nickel Sepharose (GE Healthcare) column and eluted with a linear gradient to 350 mM Imidazole. Pooled fractions were dialyzed into 20mM Tris pH 7.5, 50 mM NaCl, 2.5 mM $CaCl_2$, and 0.5 mM β-mercaptoethanol and cleaved at 4°C overnight with 50-100 U Thrombin. Protein was further purified over a Resource S ion exchange column with a linear gradient of 50 mM-300 mM NaCl, 20 mM Tris pH 7.5, and 0.5 mM β-mercaptoethanol. Protein was concentrated using Amicon Ultra 5K MWCO (Millipore) and run over a 16/60 Superdex75 gel filtration column in NMR buffer (20 mM Sodium Phosphate pH 6.7, 100 mM NaCl, 1 mM DTT).

Protein-DNA complex formation

Single-stranded GBS oligos (IDT) were purified by 10/10 MonoQ as described previously[17], and resuspended at 2 mM. Oligos were annealed at 1 mM in boiling water and cooled slowly to room temperature. To ensure solubility of GR DBD – DNA complexes, dilute protein was combined with dsDNA 1X diluted in cold NMR buffer. DNA was in ~33% excess to DBD, to ensure saturated binding. Dilute GR DBD – DNA complexes were concentrated slowly at 4°C using a 3K MWCO centrifugal filter (Amicon) to ~300 μM dimer complex and filtered with Ultrafree PVDF 0.22 μm columns (Millipore).

Protein NMR assignment

Triple-labeled GR DBD was prepared as described above except that 50 mL LB cultures were resuspended in 1 L unlabeled minimal media, grown to optical density of ~0.7, then pelleted and resuspended in 1 L of minimal media containing 2 g of $^{15}NHCl_4$, 2g ^{13}C-glucose in 90-100% D_2O. Expression was induced at 30°C for 12 h. Assignments in the absence of DNA were with 1.7 mM GR DBD. The following experiments were run at 25°C on a Bruker 500 MHz spectrometer: ^{15}N-HSQC, HNCO, HNCACO, HNCA, HNCOCA, HNCOCACB[52,53], and CC(CO)NH-TOCSY. A ^{15}N-edited NOESY[54] was ran on a Bruker 800 MHz spectrometer. For assignment of DNA-bound GR, purified DBD and the Gha GBS were combined at a ratio of 2:1 monomer to GBS and concentrated to 500 μM complex (1mM GR DBD). NMR assignments were generated from ^{15}N-edited NOESY[54], TROSY-HNCO, TROSY-HNCOCA, TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HNCACB, and TROSY HN(CA)CO[55] experiments conducted on 600, 800 and 900 MHz spectrometers at 25°C and 35°C because some peaks gave stronger signal at 35°C. Assignments were aided by ^{15}N-HSQC of ^{15}N specific amino acid labeling of Ile,

Leu, Val, Phe, Tyr, Lys, and reverse-labeled Arg using DL39 cells for expression. All NMR data were processed in NMRPipe[56], and analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). WT DBD – GBS assignments were transferred to A477T DBD complexes, in general, according to the minimal combined ^{1}H and ^{15}N chemical shift difference for each assigned GR WT peak to the nearest A477T peak[25]. Assignment transfer was aided by ^{15}N reverse-labeling of Arg or Lys A477T residues.

Chemical shift difference and ZZ-exchange NMR

^{15}N-HSQC spectra were acquired on a Bruker 800 MHz spectrometer at 35°C. Peak assignments were transferred from the WT DBD – Gha complex to additional GBS complexes by measuring the minimal chemical shift difference from each assigned WT DBD – Gha peak using the formula: chemical shift difference, $\Delta\delta = [(\Delta^{1}H \text{ ppm})^2 + (\Delta^{15}N \text{ ppm} / 5)^2]^{1/2}$ for each assigned GR WT peak to the nearest A477T peak[25]. Similarly, peak assignments for each A477T – GBS complex were transferred by minimal chemical shift difference using the A477T DBD – Gha complex as a reference.

For ZZ-exchange, ^{15}N-HSQC spectra were acquired on a Bruker 800 MHz spectrometer at 35°C using a pulse sequence modified to include a mixing time of 0, 0.025, 0.05, 0.1, 0.2 and 0.4 s. The intensities of auto and exchange peaks were plotted against mixing period and curves were fit individually for Gly470 and Ile484 residues according to the formulas described by Farrow *et al.* [57]. The k_{ex} and relaxation rate for each conformation were fit separately.

Transcriptional Reporter Assays

GBS reporter plasmids were either those generated by Meijsing et al.[17] or were constructed equivalently. Reporter assays were performed as previously described[17], except that we used hGR WT and hGR A477T (generated by PCR site-directed mutagenesis). Briefly, U2OS cells were seeded in DMEM + 5% FBS in 24-well plates at ~20,000 cells per well one day prior to transfection. Cells were transfected with 20ng/well GR plasmid, 20ng/well GBS-luciferase plasmid, 200 pg/well pRL Renilla, 120 ng/well empty p6R plasmid, 1 µL PLUS Reagent and 0.7 µL/well Lipofectamine Reagent (Invitrogen) for 4 hours in no-serum DMEM media. Cells were washed and recovered in DMEM + 5% FBS for 3 hrs, then treated with 100 nM dexamethasone or ethanol for 12 hrs. Luciferase induction was measured using Dual-Luciferase Reporter kit (Promega) in 96-well format using a Tecan plate reader. Data is normalized to Renilla for each well, then to ethanol-treated control, and empty pGL3 vector control.

Surface Plasmon Resonance

SPR analysis of WT and A477T DBD interaction with GBSs was carried out on a Biacore T100. Matrix-free surfaces were prepared by injection of Neutravidin (Invitrogen) across a planar saccharide monolayer with covalently coupled biotin (BP chips, Xantec Bioanalytics) at 25°C in 20 mM HEPES (pH 7.2), 150 mM NaCl. Double-stranded GBS oligos with a single 5' biotin-TEG label (IDT) were subsequently captured at immobilization levels ranging from 20 to 65 RU. WT and A477T DBDs were dialyzed overnight in NMR buffer. Following dialysis, 0.1 mg/mL BSA (Sigma) was added to both the assay buffer and the protein samples. Fifteen point concentration series were prepared by serial dilutions spanning 0.700–200 nM for WT and 1.4–400 nM for A477T. Association and dissociation times were selected to ensure equilibrium and

complete dissociation. All data were processed and analyzed in Matlab. Isotherms were fit to the Hill equation: Fractional occupancy = $(c/(c+K_{1/2}))^{nH}$. Errors in $K_{1/2}$ and n_H were determined using a bootstrap method with replacement: after scaling of *n* equilibrium responses for each GBS, a random set of **n** data points was selected with the possibility of selecting the same data point multiple times. After 500 iterations, the 100 best-fit parameters (s.s.e) were used to find mean values and standard deviations. $t_{1/2}$ parameters were fit from the following equations: $R = R_o$ e^(-$t \cdot k_{off}$) and $t_{1/2} = \ln(2)/k_{off}$, where *R* = response units.

Electrophoretic Mobility Shift Assay (EMSA)

Unlabeled GR WT DBD and A477T DBD were purified as described above. Double-stranded GBS with an Alexa488 fluorophore conjugated to one of the 3' ends (IDT) were incubated for 30 min at 5 nM final concentration with GR DBD titrations in Binding Buffer (20 mm Tris pH 8, 50 mM NaCl, 1mM EDTA, 10ug/mL dIdC, 5 mM $MgCl_2$, 200 µl/mL BSA, 5% glycerol, and 1 mM DTT) on ice. Native 8% polyacrylamide gels were run at 200 V in 0.5X TBE at 4°C. Alexa488 signal was imaged on a Typhoon scanner (GE Healthcare) and quantified using ImageQuant. Fraction GR DBD bound was determined as 1 – [DNA_{free}].

GR ChIP-sequencing

Details of GR ChIP-sequencing methods and data will be described in a separate publication**.** Briefly, U2OS cell lines stably expressing GR WT or A477T were treated with 100 nM dex for 90 min, then cross-linked with 1% formaldehyde and quenched with 125 mM glycine. Cells were lysed for 30 min at 4°C, and nuclei pelleted at 600 x g for 5 min and

resuspended in RIPA buffer (10 mM Tris pH8.0, 1 mM EDTA, 150 mM NaCl, 5% glycerol, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100). Chromatin was fragmented by Diagenode Biorupter sonication and incubated for 16 hrs at 4°C with pre-washed anti-GR antibodies bound to Protein G Dynabeads in RIPA buffer containing protease inhibitor cocktail and 2 μg/μL BSA. Complexes were washed extensively with RIPA buffer containing 500 mM NaCl followed by LiCl buffer and cross-links were reversed. DNA was column-purified (Zymogen Clean and Concentrator Kit). ChIP-sequencing libraries were prepared by end-repair, dATP-addition and ligation of sequencing adaptors containing in-house barcodes. Libraries were amplified by 17 cycles of PCR and purified by PAGE. Libraries were sequenced using the Genome Analyzer II (Illumina) with 2 X 75 bp paired-end reads, and aligned with Bowtie[58]. Motif analysis was performed using MEME[59].

DNA oligos for NMR, EMSA, SPR and FRET

GBS	5' → 3'
Fkbp5	gtacAGAACAgggTGTTCTtcgac
Gha	gtacGGAACAtaaTGTTCCtcgac
Gilz	gtacAGAACAttgGGTTCCtcgac
Pal-F	gtacAGAACAaaaTGTTCTtcgac
Pal-R	gtacAGAACAtttTGTTCTtcgac
Pal-ttg	gtacAGAACAttgTGTTCTtcgac
Sgk	gtacAGAACAtttTGTCCGtcgac
Sgk-ggg	gtacAGAACAgggTGTCCGtcgac
Sgk-ttg	gtacAGAACAttgTGTCCGtcgac

References for Online Methods

52. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93–158 (1999).
53. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J Am Chem Soc* **121**, 844–848 (1999).
54. Talluri, S. & Wagner, G. An optimized 3D NOESY-HSQC. *J Magn Reson B* **112**, 200–205 (1996).
55. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA* **95**, 13585–13590 (1998).
56. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293 (1995).
57. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* **4**, 727–734 (1994).
58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**, 357–359 (2012).
59. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

Chapter 4:

Structural consequences of binding events on the androgen receptor ligand binding domain

Abstract:

The androgen receptor ligand binding domain interacts with small molecule hormones and proteins. The hormone binding site is a buried hydrophobic pocket, where ligand binding plays an integral role in the structural core of the receptor. The protein-protein interaction surface is proximal to the hormone binding pocket and serves as an adaptor site for coregulator proteins. While the energetics and kinetics of binding at these two sites are disparate, they are similar in their capacity to adapt and accommodate structurally diverse ligands. Ambient temperature X-ray crystallography enables an ensemble framework to connect structural differences in hormone ligand to local and distant changes in polysteric clusters of the receptor's structure. The energetic coupling between the hormone binding pocket and the coregulator binding surface is explored using surface plasmon resonance, and the role of hormone in differential coregulator recruitment is revealed as a consequence of perturbations to the structural ensemble. In measurement of coregulator binding kinetics, two classes of interaction are revealed. The structural impact of these interactions is superficial for the fast dissociating coregulators, while tight binding interactants penetrate and elicit distant redistributions of the structural ensemble. For the longest-lived coregulator complex, we find evidence of structural and energetic coupling with a distant loop that is a known hotspot for post translational modification. Together, these results suggest ensemble encoding of molecular recognition events.

Ligand Affects Coregulator Binding:

The classical agonists of androgen receptor are testosterone, dihydrotestosterone, and metribolone (R1881). Testosterone is primarily produced in the sex organs (male testicles, female ovaries) and is circulated in complex with the sex hormone binding globulins. While the mechanism of cell entry is unclear, testosterone can be converted to dihydrotestosterone by the enzyme 5-alpha-reductase, or estradiol by the enzyme aromatase, both of which have tissue specific expression patterns. It is generally held that muscle and bone lack 5-alpha-reductase, making testosterone the primary effector molecule in skeletal muscle. The synthetic androgen R1881 is not metabolized and has been easier to procure in research laboratories until recently when it was also classified as a schedule 3 controlled substance. For these reasons, R1881 is often used for studying androgen responsiveness in cell culture transcription assays. This is a common theme in NR studies, as the synthetic glucocorticoid dexamethasone is routinely employed in cell culture activation of the glucocorticoid receptor. While dexamethasone is a significantly more potent agonist than the natural glucocorticoid agonist cortisol, R1881 and dihydrotestosterone have very similar EC50 values. Even with the caveat of testosterone and dihydrotestosterone being metabolized, I believe biological signaling events should be studied with their natural activators.

The hormone binding pocket (HBP) of the androgen receptor is an allosteric site that regulates the presentation of the active site which binds coregulator proteins. As rigid ligands, the three androgenic steroids (DHT, R1881, and TES) can be thought of as keys that unlock the activity of the receptor. While similar in structure, these keys do have different edges that would engage the tumblers of the lock in ligand specific ways. Indirect measurement of ligand binding (IC_{50}) informs us that the different keys have similar binding energies (1.1 nM, 1.4 nM,

4.5 nM for DHT, R1881, and TES, respectively).[1] Measurement of transcriptional response to these ligands tell us that we have unlocked activity. But the path to activation might be different. Specifically, do the different ligands unlock different activated forms of the androgen receptor?

In order to measure the ligand dependent changes in coregulator recruitment, biotinylated AR LBD was expressed and purified in the presence of the three different agonists. Four representative 12-mer coregulator fragments were assayed for interaction with the three liganded forms of AR. Three AR-specific coregulators (gelsolin, AR-n-terminal domain, P21ActivatedKinase-6) and one general transcription factor (TRAP100) were considered. This distinction is made based on the presence of phenylalanine or leucine in the coregulator motif.[2] GSN, NTD, and PAK are the three tightest interactions that I have measured for AR LBD and differ in NR-box motif at the +4 position, with PHE, LEU, or MET, respectively. TRAP 100 is the tightest binding LXXLL motif that I have assayed.

While the interaction of gelsolin with AR was first shown as transcription activating in 2003,[3] the mechanism is unknown. The principal enzymatic function of gelsolin is a potent actin severing protein[4]. Colocalization studies have revealed that gelsolin might be involved in the nuclear import of the androgen receptor. In delivery of this motif via transfection or cell-penetrating peptides, it is possible to inhibit androgen receptor function.[5]

The N-terminal domain peptide corresponds to residues 19-30. *In vitro* intramolecular FRET measurements have implicated this motif in both an intra-molecular N/C-interaction and dimerization.[6] More recently, it has been shown that the androgen receptor does not form dimers at low μM concentrations.[7] However, as this coregulator motif is present at near mM concentrations throughout all stages of AR signaling, it will be a strong competitor for interaction with the coregulator binding surface.

p21-Activated Kinase 6 functions as an inhibitor of both androgen and estrogen receptors.[8] CDK interacting protein 1 (p21) inhibits the activity of cyclins and functions as a regulator of cell cycle progression at G_1 and S phase.[9] As its name suggests, PAK6 is activated by p21 and has been shown to inhibit prostate cancer cell growth by phosphorylation of the androgen receptor and activation of a ubiquitin ligase that targets AR for degradation.[10]

short name	coregulator function	sequence
GSN	actin binding protein	ETPLFKQFFKNW
NTD	N-terminal domain of AR	YRGAFQNLFQSV
PAK	kinase	KRRLFRSMFLST
TRAP	mediator complex	QPSKLMRLLSSN

Table 1. Coregulator peptide sequences used in the comparison of the three agonist bound forms of AR.

TRAP100, also known as MED24, is a component of the mediator complex that brings RNA pol II to gene transcription start sites.[11] I classify TRAP100 as a general transcription factor because it bears an LXXLL motif, which can presumably interact with most NRs. However, in screening of multiple LXXLL motifs from different steroid receptor coactivators, I learned that this is the tightest interacting. Further, in comparison to LRH, SF-1, and glucocorticoid receptor, AR shows the tightest binding to the TRAP100 motif. So, while it is a general transcription factor, TRAP100 is arguably the most relevant non-aromatic NR box. Given it's relative tight affinity and direct role in recruiting RNA pol II, TRAP100 is likely the key adaptor molecule for androgen activation of transcription.

Because receptor loses ligand during the course of the SPR experiment, it is necessary to keep the ligand in the running buffer at levels much higher than the K_D. This precludes simultaneous assay of the three different ligand states. So, comparison of ligand states was done over three separate experiments using the same exact coregulator peptide stocks and sample

preparation protocols. For each experiment, the different liganded receptors were immobilized at comparable densities to flow cells 2, 3, and 4, enabling triplicate measure. Dilutions of coregulator motifs were prepared by serial transfer of 180µL into 120µL in standard 96-well plates, resulting in 0.6x concentration steps. A fifteen point concentration series generated from a high concentration of 25µM was used in order to adequately sample the isotherm.

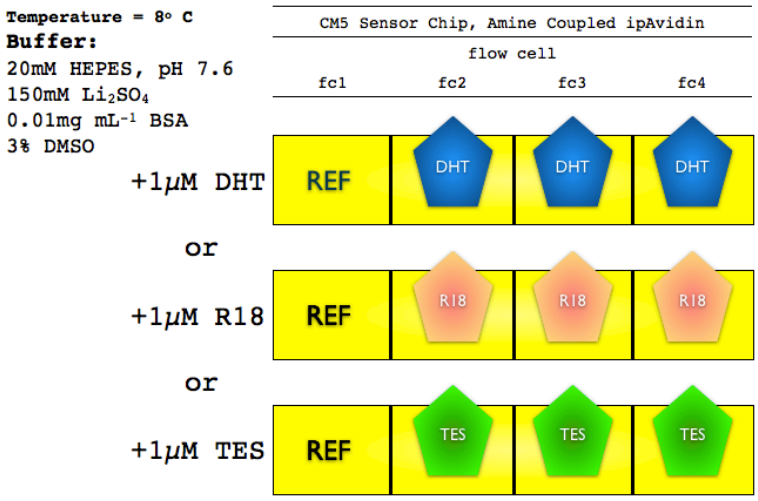


Figure 1. Experimental setup for ligand comparison experiments.

A control injection of 200nM NTD was included with each sample injection to confirm the stable activity of the immobilized receptors.

All raw traces used in the analysis appear in Figures 2 through 4. Data was processed and parameterized in Matlab. The main quantities of interest were the concentration that gave half occupancy ($K_{1/2}$) and the lifetime of formed complexes ($t_{1/2}$). While a number of strategies were considered for the $K_{1/2}$ determination, ultimately expectancy maximization of the absorption energy distribution was used.[12] Whether fit with the Langmuir model or the Hill equation, the trend reported is consistent. For determination of the half life of different complexes, the dissociation curves were fit using a model that explicitly treats mass transport. TRAP 100 was determined to dissociate too rapidly for estimation of dissociation kinetics.

DHT :

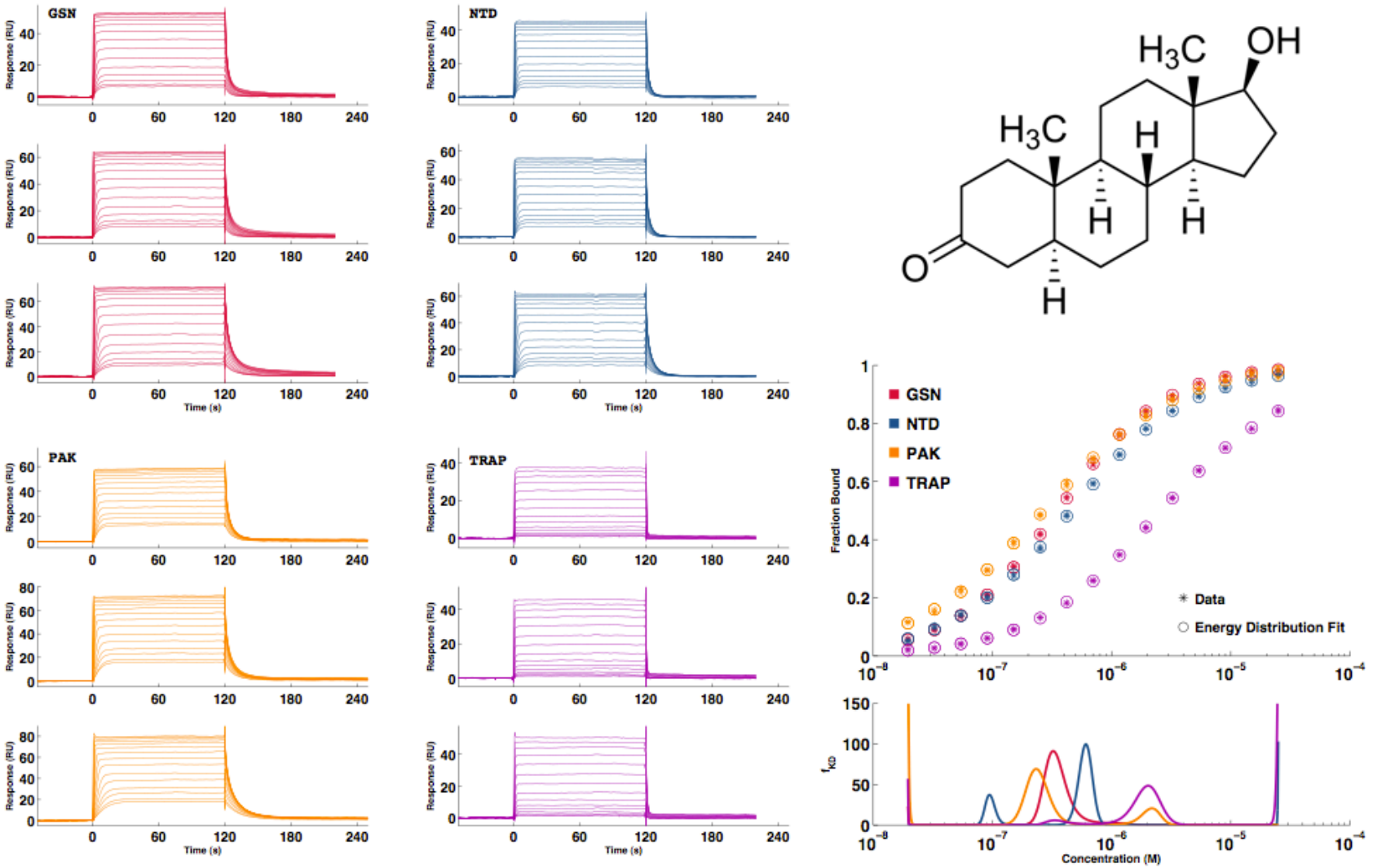
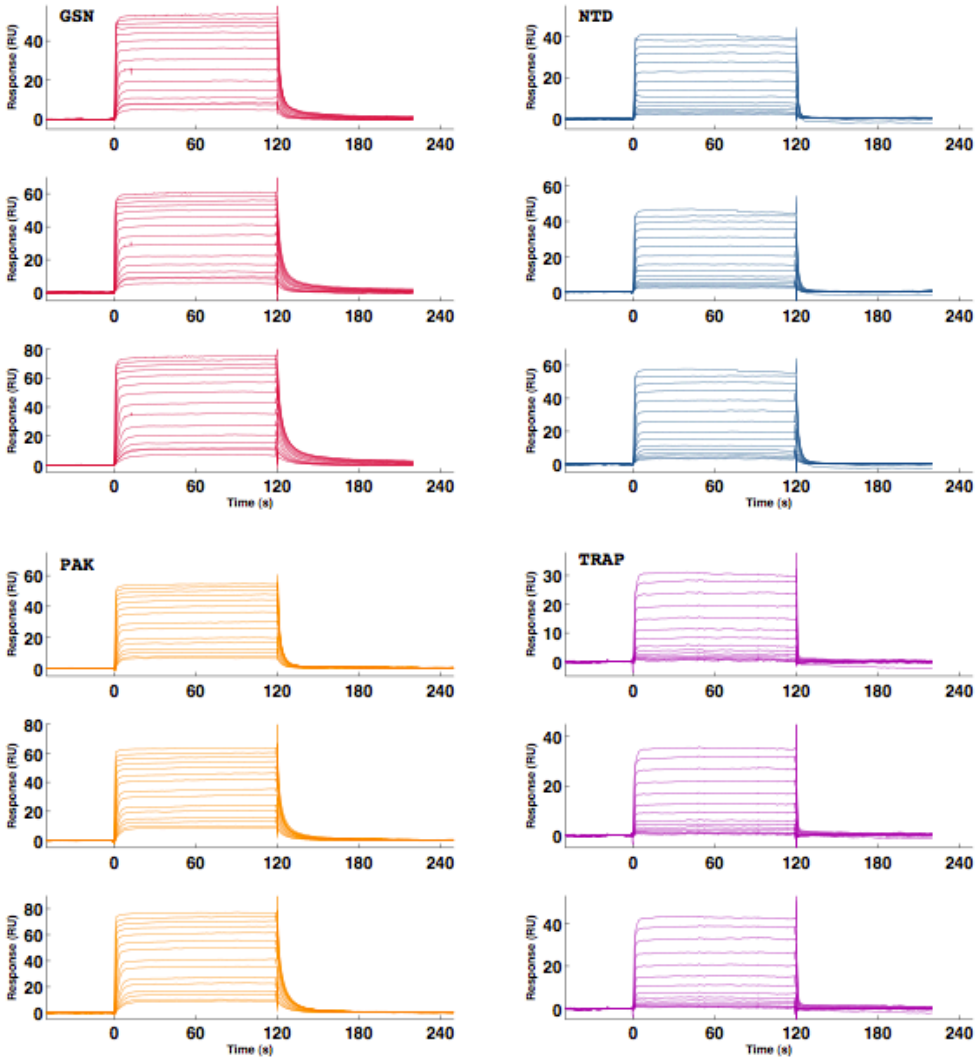
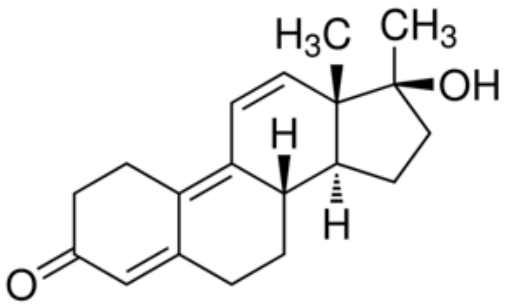


Figure 2. Raw traces and isotherms for DHT liganded AR.

R1881:





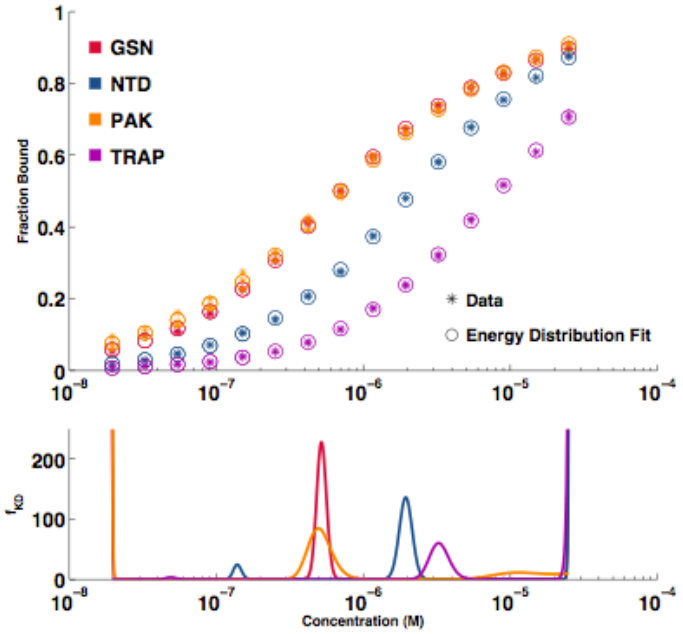
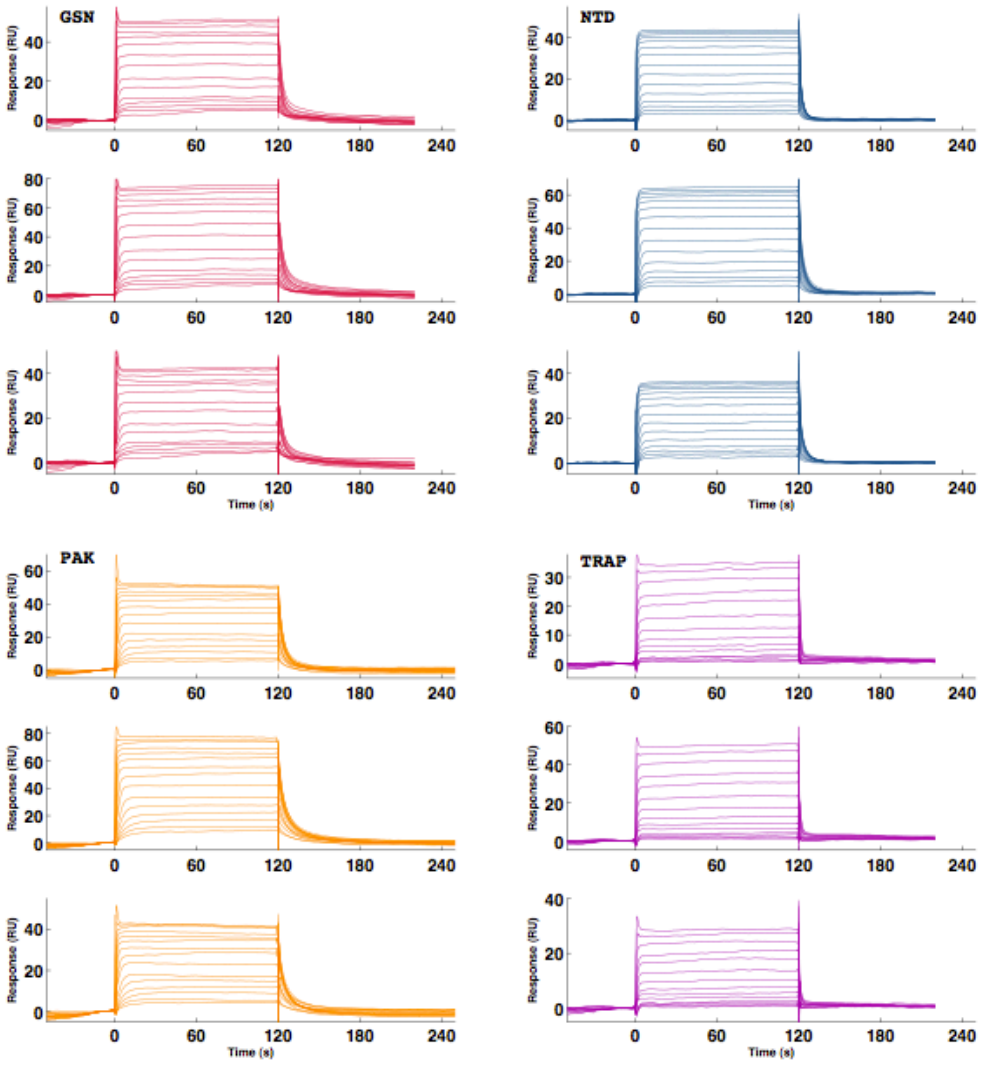
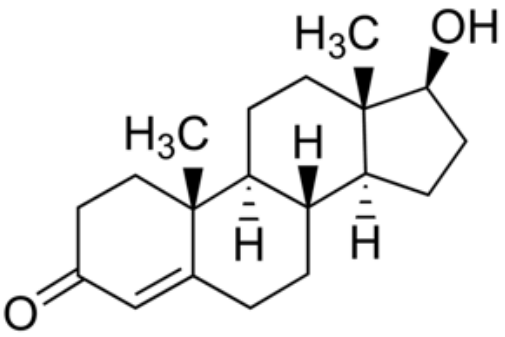


Figure 3. Raw traces and isotherms for R1881 liganded AR.

TES :





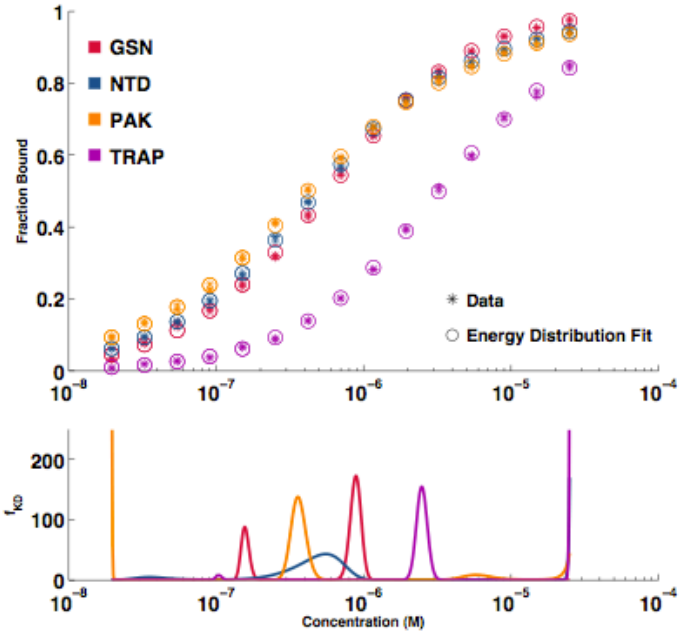
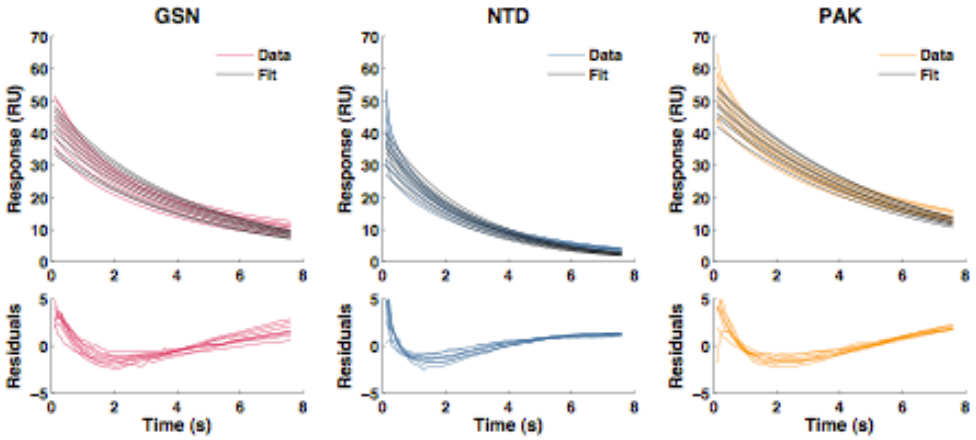
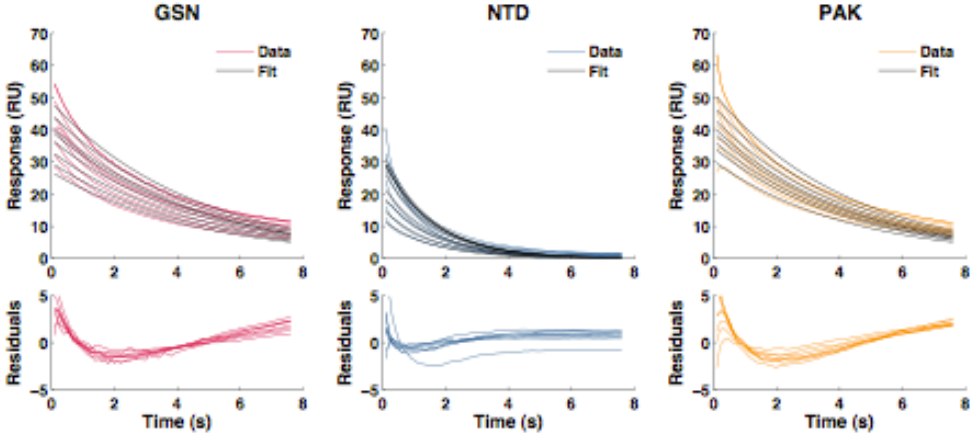


Figure 4. Raw traces and isotherms for TES liganded AR.

DHT:



R1881:



TES:

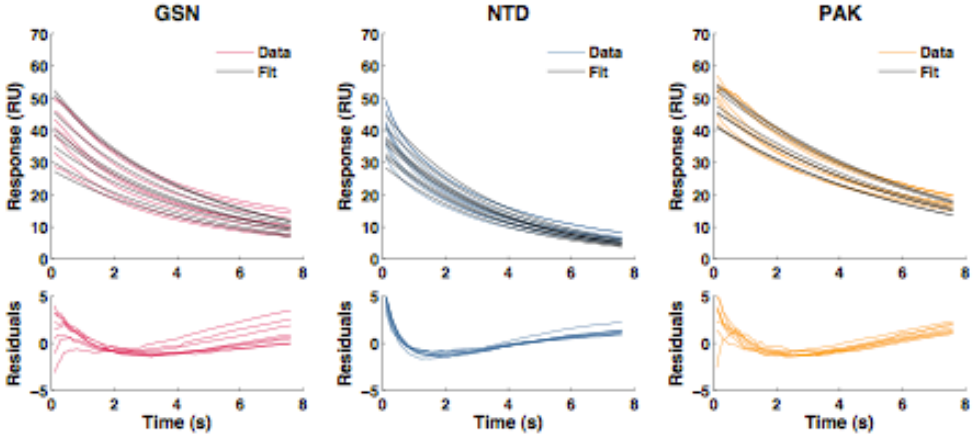


Figure 5. Dissociation fits for GSN, NTD, and PAK to the DHT, R1881, and TES liganded androgen receptor.

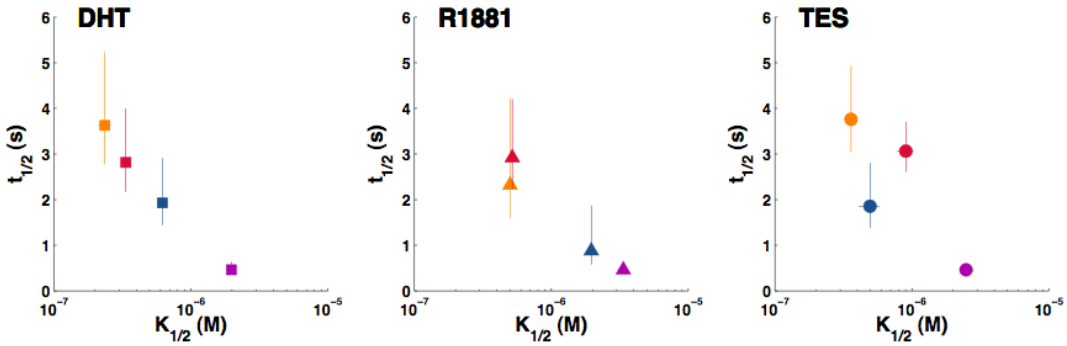
Results:

The binding of different coregulator motifs to the androgen receptor exhibits non-ideal behavior, irrespective of ligand. I consider my inability to formulate a physically meaningful model that accurately describes the kinetics of coregulator binding as my greatest failing in graduate school. However, in the absence of a complete model, the data does allow for parameterization of both the concentration and time dependencies of complex formation.

Considering the steady state data, there is evidence of heterogeneity in the isotherms. In contrast to published studies using fluorescently labeled coregulator motifs that report cooperative binding,[13] SPR measurement reveals an anti-cooperativity in the isotherms. This deviation from the Langmuir model is taken as multiple bound states with different affinities and a weak non-specific interaction. It is also worth noting that prior to inclusion of BSA in the running buffer or the use of LoBind 96-well plates, cooperative binding was also observed by SPR. This is attributed to the sticky / hydrophobic coregulator motifs adsorbing to the walls of the sample tubes. This uniform subtraction of some small amount of sample more significantly affects the absolute concentration of low concentration samples in the isotherm, resulting in an underrepresentation of response and the appearance of cooperativity. Concentration dependence was determined by individual fitting of each flow cell using the AED approach. While multiple peaks do appear in these distributions, only the major binding event was considered and the mean and standard deviation were calculated based on data collected for the three flow cells.

The kinetics of dissociation were fit using the first eight seconds following sample injection. Dissociation of the TRAP coregulator was deemed too rapid for fitting, and the presented $t_{1/2}$ is the experimental limit of a dissociation rate at 1.5 s^{-1}. This is the first

measurement of kinetics for coregulator binding to the androgen receptor and the complexes are surprisingly short-lived. The model used explicitly treats mass transport, but there is still evidence of systematicity in the residuals, see Figure 5. One possible explanation for the apparent structure in the residuals is the existence of multiple bound conformations that dissociate at different rates. Considering this, it should be noted that fitting sums of exponentials significantly improves the fit. But as the data also shows evidence of mass transport, it would be in bad taste to discount this effect. And, unfortunately, there is no sensible model for the simultaneous treatment of multiple bound states and mass transport. So, again in the interest of parsimony, the single bound state model is used. While the error in fitting with this model is considerable, the off rate is well determined.



Ligand:	DHT		R1881		TES	
GSN:	335 ± 29 nM	2.8 ± 0.6 s	518 ± 22 nM	2.9 ± 0.7	903 ± 120 nM	3.1 ± 0.4 s
NTD:	620 ± 17 nM	1.9 ± 0.5 s	1.96 ± 0.058 μM	0.9 ± 0.3 s	495 ± 85 nM	1.8 ± 0.5 s
PAK:	235 ± 24 nM	3.6 ± 0.85	500 ± 7 nM	2.3 ± 0.7 s	360 ± 20 nM	3.8 ± 0.7 s
TRAP:	1.97 ± 0.1 μM	< 0.4 s	3.35 ± 0.24 μM	< 0.4 s	2.48 ± 0.075 μM	< 0.4 s
	$<K_{1/2}> \pm 1\ \sigma$	$<t_{1/2}> \pm 1\ \sigma$	$<K_{1/2}> \pm 1\ \sigma$	$<t_{1/2}> \pm 1\ \sigma$	$<K_{1/2}> \pm 1\ \sigma$	$<t_{1/2}> \pm 1\ \sigma$

Figure 6. Energetic and kinetic parameters for binding of the four coregulators to AR LBD under three different ligand contexts. Affinity ($K_{1/2}$) and kinetics ($t_{1/2}$) are presented as the mean ± 1 standard deviation from measurements of the 3 flow cells.

It is unsatisfying to accept these inadequate models, however, the purpose of these experiments was to compare the effect of ligand on coregulator recognition. Therefore, it was necessary to reduce the complexity of the binding traces to these single parameters that allow direct comparison between DHT, R1881, and TES. It is this aim of comparison that makes this exercise stomachable.

Comparison of Ligand Effects on Coregulator Binding:

For the purpose of this discussion, the DHT bound receptor will serve as the reference state. Rank ordering of affinities places PAK < GSN < NTD < TRAP. There is significant correlation between the concentration and time dependencies for this set of interactants with the DHT bound receptor. Further, in this two dimensional space, the four interactions are well separated with respect to each other.

Against the R1881 liganded receptor, all interactions are significantly red-shifted in both concentration and time dependencies. The relative rank ordering of affinities for the coregulators is the same as for the DHT liganded receptor. However, there is no longer any energetic discrimination between PAK and GSN against the R1881 liganded receptor. The most striking discrepancy is the > 3 fold weakening of interaction with the NTD. In it's shifted position, it is no longer well separated from the interaction with TRAP. From a kinetic perspective, GSN is equivalent between DHT and R1881, while PAK dissociates more rapidly. Also, the weakening of NTD is accompanied by significant acceleration to the dissociation kinetics.

For the TES liganded receptor, there is a shift in the rank ordered affinities, PAK < NTD < GSN < TRAP. The measured affinity for NTD is significantly tighter for the TES liganded receptor than for the DHT liganded receptor. All other interaction energies are red-shifted

with respect to the DHT liganded receptor. Interestingly, the complexes formed by GSN and PAK with the TES liganded receptor are longer lived than when bound to the DHT receptor.

It is interesting to consider the properties of an "ideal" ligand. If the signaling pathway is one with a deterministic outcome, the ideal ligand could possess an extremely tight energy of interaction and activation of the receptor would maximize energies of downstream binding events. For a receptor that varies signaling output depending on cellular context, the essential quality of the ligand could be the capacity to aid in the discrimination of interacting partners. In this case, the biophysical space that is accessible to the activated form of the receptor is of importance.

For these three ligands, the energies of direct interaction are comparable. Considering the affinity and kinetic space used in coregulator recognition, the natural ligands DHT and TES seem to bestow a biophysical sensitivity to the activated androgen receptor with unique parameters describing the different interactions. Activation by the synthetic ligand R1881 seems to compromise this ability and the receptor is less aware of the identity of its interaction partners. For an already promiscuous surface, any compromise in energetic discrimination would only further muddy the landscape of interaction potentials.

Considering only the NTD, it is remarkable how much control the identity of ligand has over this interaction. The NTD is the ever-present chief competitor for access to the coregulator binding surface. While it is unclear how to model this competition, the relative energies between the three ligand states inform the requisite concentrations of a secondary coregulator in order to displace the NTD. The TES liganded receptor would be the most difficult to access, followed closely by the DHT liganded receptor. In contrast, the R1881 liganded receptor could easily be accessed with relatively low levels of a secondary coregulator.

In fact, if we accept that the activity of transcriptional activation is mediated by a shared set of general steroid receptor coactivators, this activity will be mediated by recruitment of the weak binding LXXLL motifs. So, while R1881 decreases the affinity for all coregulators including the LXXLL of TRAP, the significant hit in affinity for the NTD effectively increases the ability of LXXLLs to interact by weakening the major competing interaction. Along this same line, we find potential explanation for weaker transcriptional activation by testosterone. Because the NTD binds most tightly to the TES liganded receptor, the coregulator binding surface is more auto-inhibited, and competition by LXXLL bearing coregulators requires higher concentrations.

This section is a very simple story. All that happened was the measurement of four coregulators interacting with three different ligand states of the androgen receptor. This has been done before. But I have done it better. And I have gotten different answers. The fact that all prior published measures of AR coregulator binding interactions report cooperativity is troubling. The appearance of anti-cooperativity in the isotherms obtained by SPR measurement indicates the presence of heterogeneity in the structures of the complexes. This is further substantiated by the complex kinetics of dissociation. Perhaps the greatest mechanistic insight into androgen receptor function provided by SPR measurement is the rapid rate of dissociation. The function of the receptor is to bring a huge swarm of multifunctional enzymes for initiation of transcription. By having fast kinetics of dissociation, the receptor is able to interact with more coregulators. I often anthropomorphize the receptor as a glad handing politician. Its job is not to form long lasting relationships, rather it functions to bring as much machinery as possible to specific genomic regions.

In order to investigate the structural mechanism by which HBP ligands affect coregulator binding energetics, the following section describes a few crystal structures.

Overview of the Data Sets:

With over 60 crystal structures of the AR LBD in the PDB, it is important to question the merits of further conventional structural studies. In comparison of these structures, one is hard pressed to identify static structural consequences of different HBP or AF2 ligands. In fact, I was quite reluctant to add a structural arm to this project. Even for the novel coregulator motifs of PAK and GSN, I reasoned that crystallography would provide little mechanistic insight into coregulator binding kinetics or the energetic coupling between the HBP and the coregulator binding surface.

My opinion changed after meeting James Fraser. He had successfully used ambient temperature Xray diffraction studies to gain mechanistic insight into another oft studied molecule, cyclophilin A.[14] All of the Xray structures of nuclear receptor ligand binding domains had been obtained under cryo conditions (T ~100K). In a systematic study comparing ambient and cryo datasets, Fraser et al showed that cryo-cooling not only reduced harmonic vibrations but also removed a number of alternate conformations.[15]

Thus began my structural studies of the androgen receptor. In total, I collected 178 ambient temperature datasets of the receptor. These datasets include different HBP and coregulator binding surface ligands, in addition to a number of mutants. Appendix 1 provides an overview of all collected data and the merged datasets will be made available following publication.

This massive dataset enabled a systems approach to the structural biology of the androgen receptor. Considering only the three HBP ligand states of the androgen receptor, I collected 27, 29, and 11 datasets of the DHT, TES, and R1881 liganded wildtype receptor, respectively. In refining these different datasets, I created a single master structure of the

receptor, only modifying the ligand and then refining all structures. Following two rounds of conventional refinement, a simulated annealing refinement was run before Ringer was used to probe the electron density at all sidechain chi angles.[16]

Ringer outputs were then ported into Matlab and pairwise correlation coefficients were calculated for the 472 sidechain dihedrals across all 67 structures. Analysis of these results is still in progress. But the results themselves are visually compelling and representative plots are shown in Figures 7 through 12 on the following pages. When these images are flipped through, it becomes apparent that there are certain chi angles that are under the control of ligand. This novel approach to protein structure gives unbiased and strong evidence for changes to absolute conformation and shifts in the structural ensemble. Were this comparison of structures to be done on single refined models, it would be very difficult to make conclusions about the orientation of the chi3 angle on a solvent exposed arginine. However, this phenomenon is made clear by the heat map for ARG726. There is a remarkable amount of heterogeneity within the individual ligand structure sets. And the relative heterogeneity seems to be a function of ligand. For instance, there are a number of positions that exhibit low correlation within DHT and R1881, but high correlation across the TES set of structures. This is also a potential mechanism of coupling ligand identity to structure.

A number of the most striking heat maps are not surprisingly found for the residues that directly contact the ligand. However, there are numerous distant positions that appear to cluster by ligand, indicating that the allosteric processing of HBP ligand is far-stretching across the entire structure of the receptor. Again, these results would carry little significance in single model comparisons. The clear outlier in many of these plots is the R1881 set of structures.

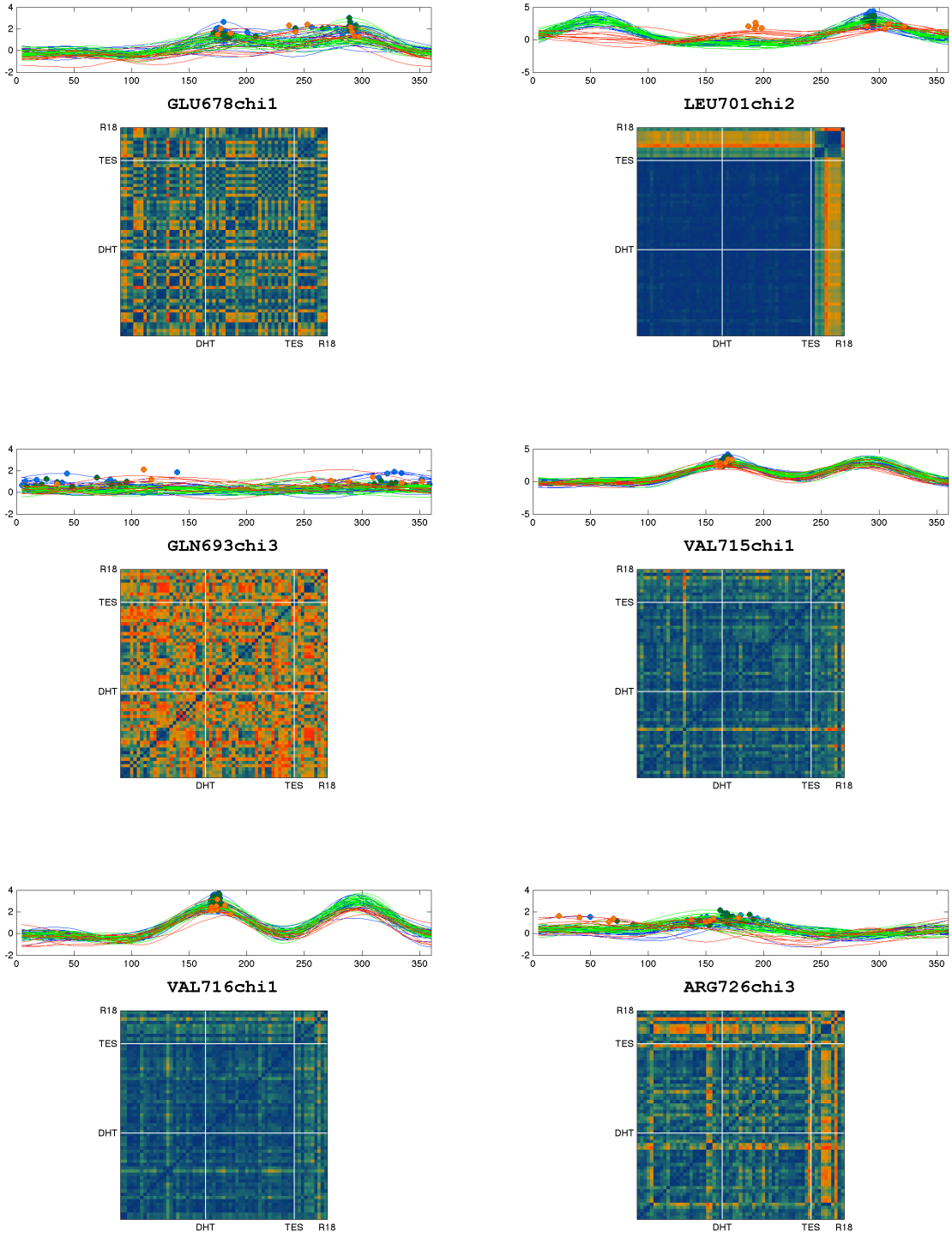


Figure 7. Selected ringer plots and pairwise correlation heatmaps for the DHT, TES, and R1881 liganded WT androgen receptor.

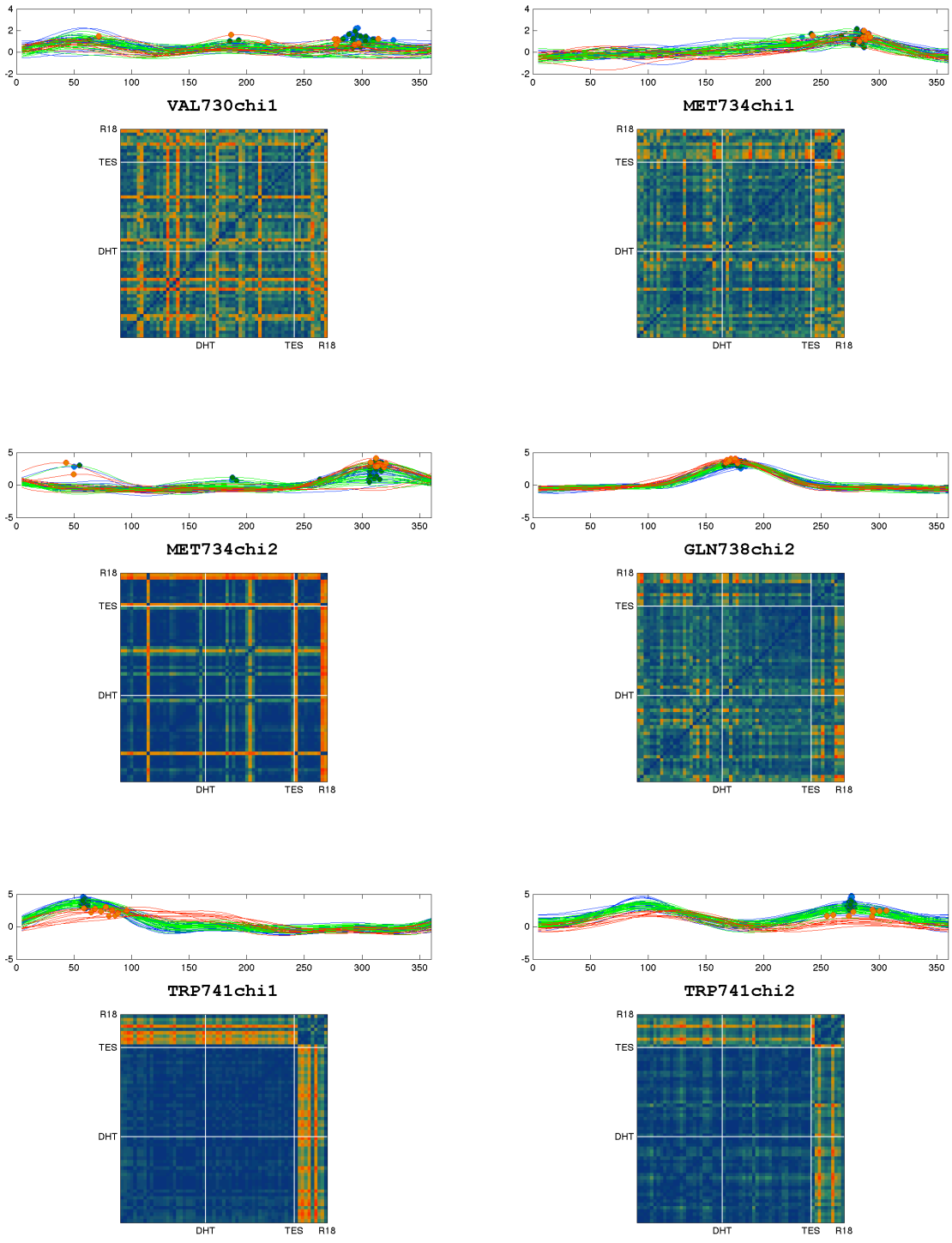


Figure 8. Selected ringer plots and pairwise correlation heatmaps for the DHT, TES, and R1881 liganded WT androgen receptor.

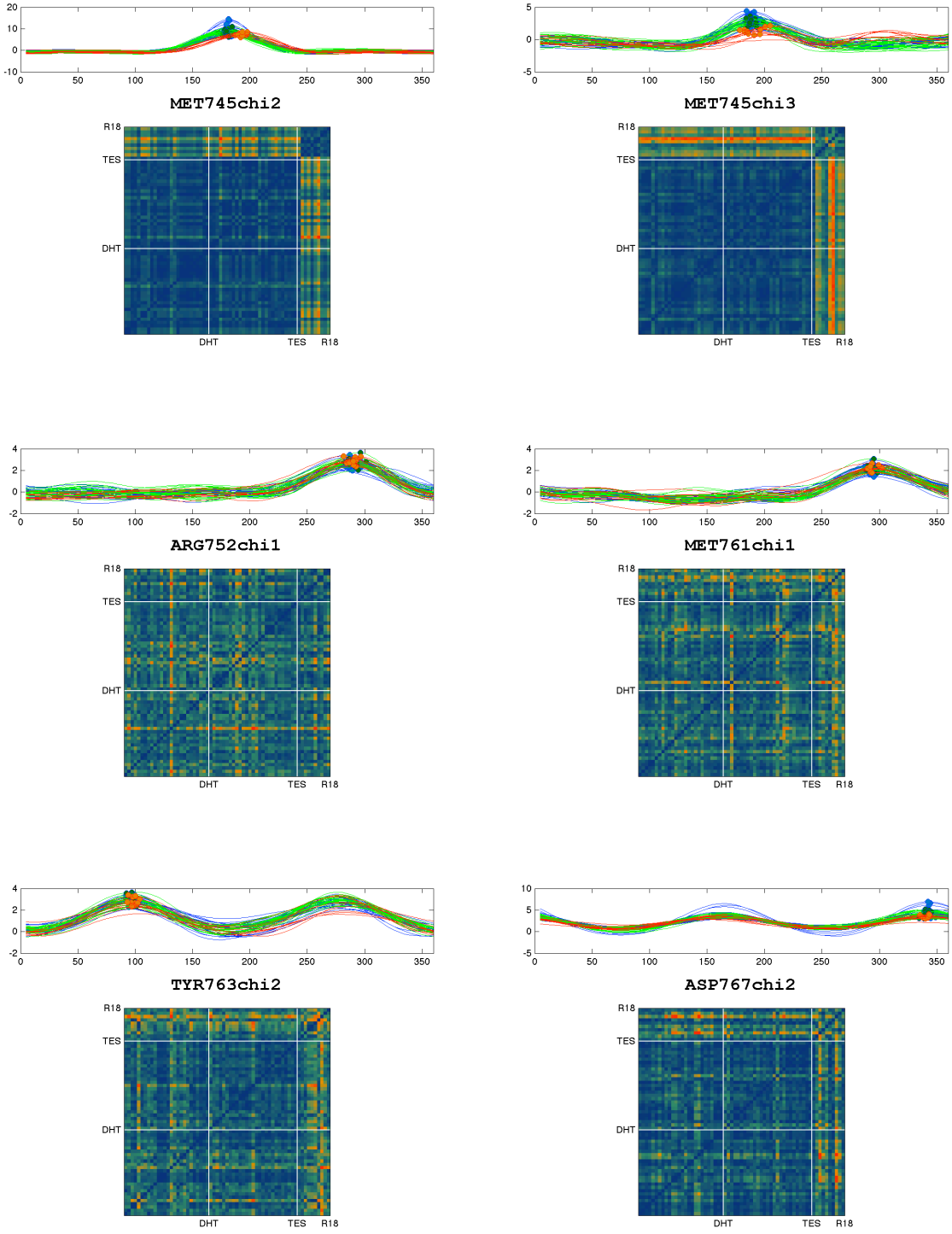


Figure 9. Selected ringer plots and pairwise correlation heatmaps for the DHT, TES, and R1881 liganded WT androgen receptor.

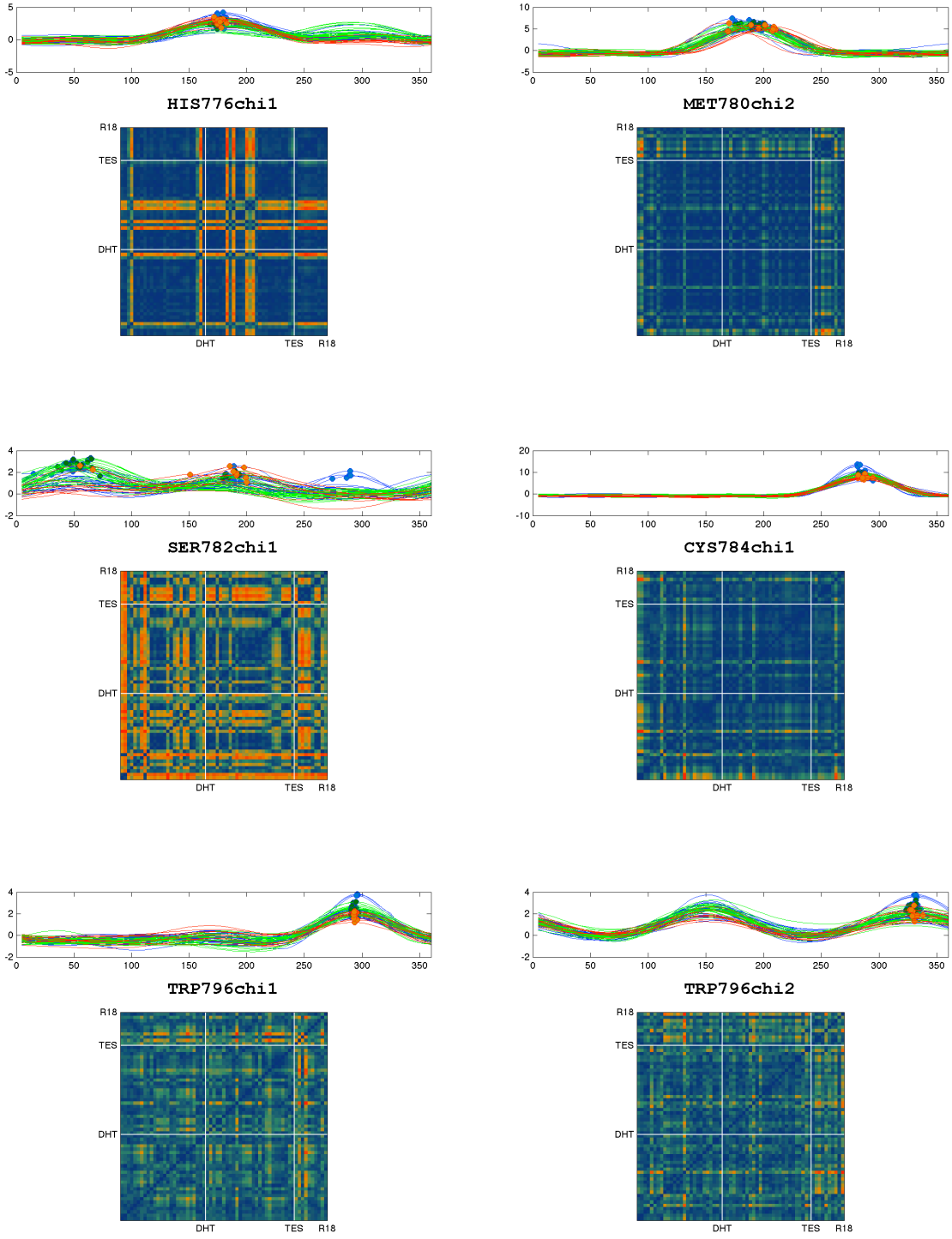


Figure 10. Selected ringer plots and pairwise correlation heatmaps for the DHT, TES, and R1881 liganded WT androgen receptor.

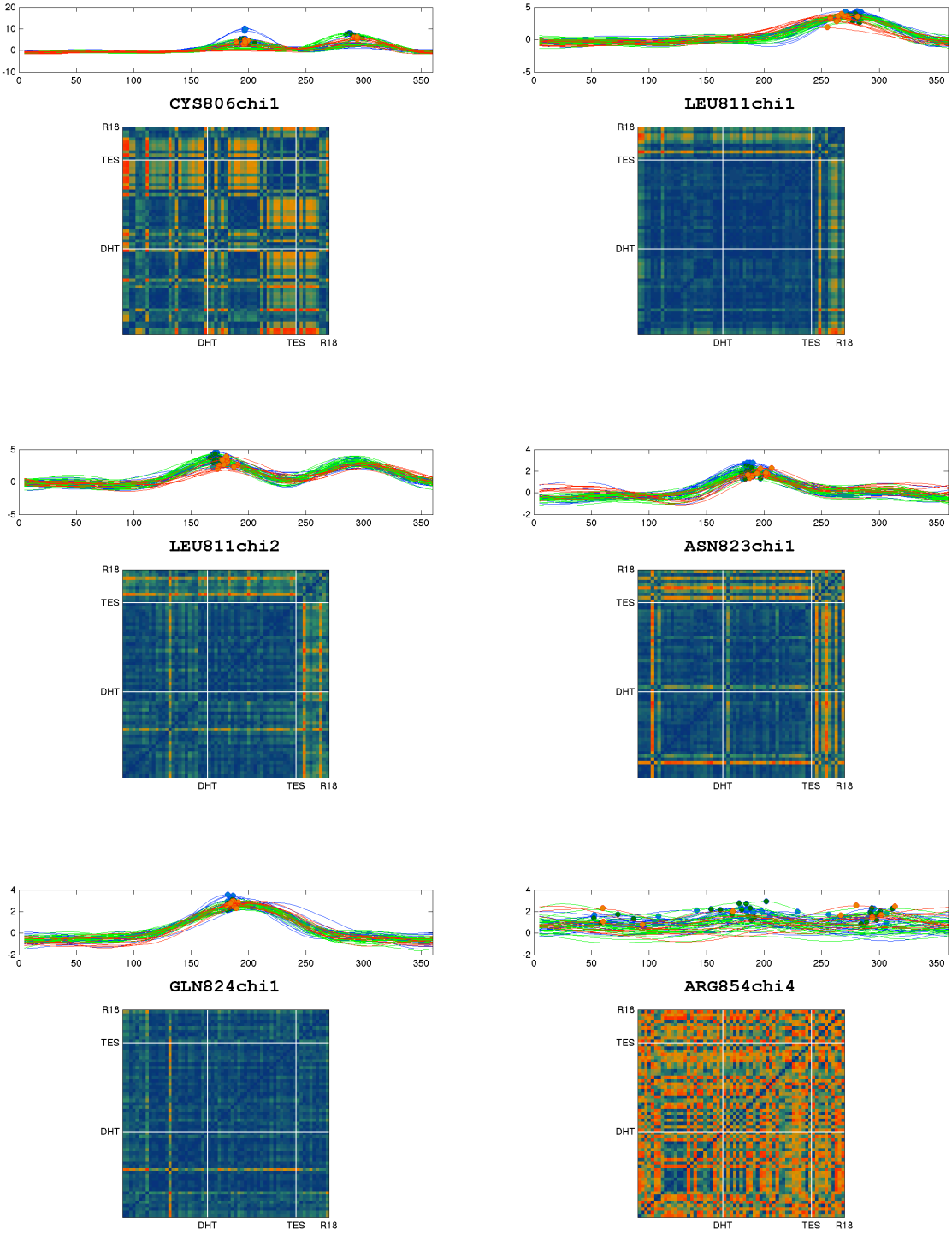


Figure 11. Selected ringer plots and pairwise correlation heatmaps for the DHT, TES, and R1881 liganded WT androgen receptor.

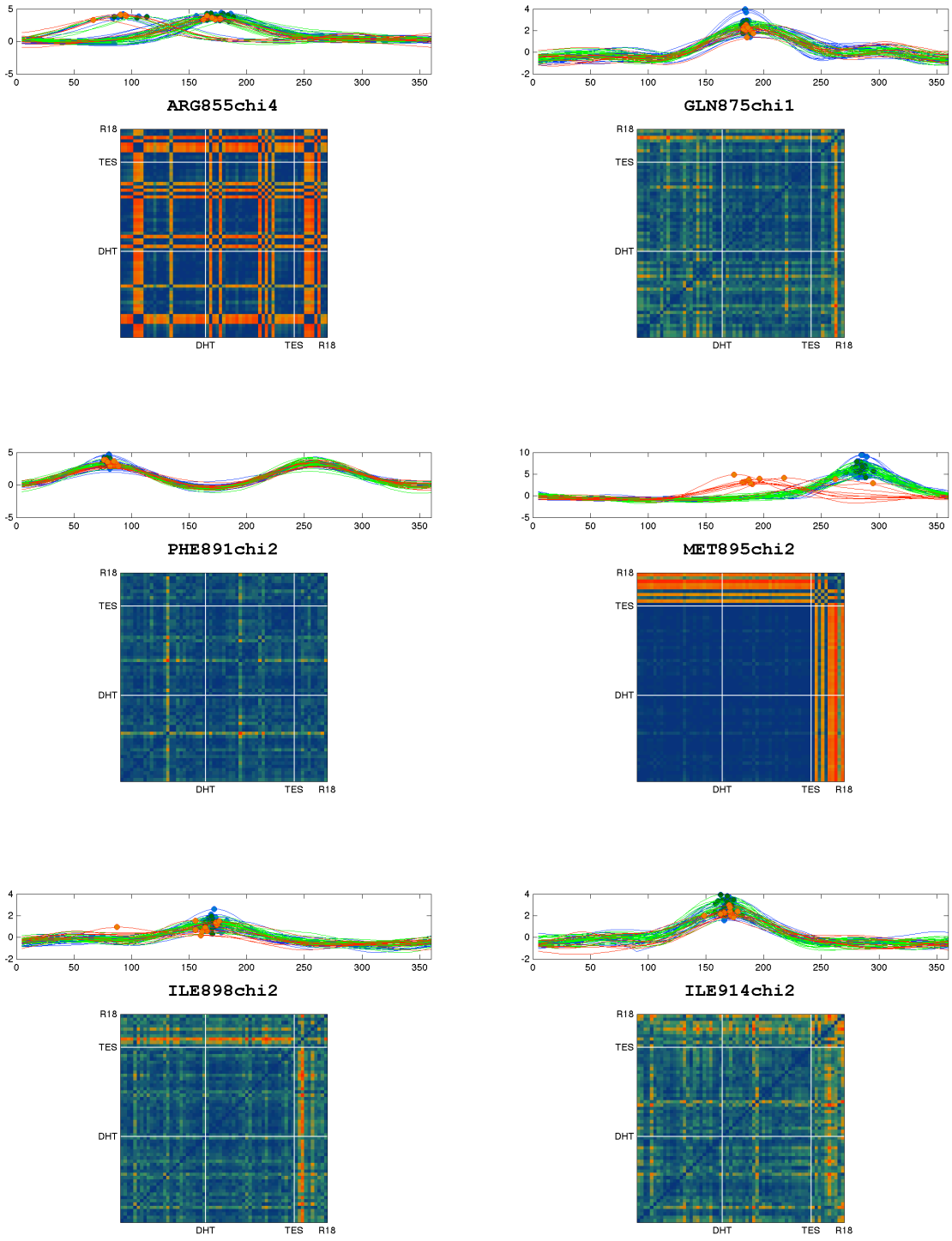


Figure 12. Selected ringer plots and pairwise correlation heatmaps for the DHT, TES, and R1881 liganded WT androgen receptor.

Weak NTD Binding to the R1881 Liganded Receptor:

The most significant functional conclusion from the SPR measurement of ligand dependent coregulator binding was the weak interaction between the NTD and the R1881 liganded receptor. This > 3-fold change in $K_{1/2}$ suggests that the R18 liganded receptor is incapable of making some of the contacts that allow for tight binding of this coregulator. Using the same protocol for generation of the Ringer heat maps, structures of the three ligand states bound to the NTD coregulator were used to explore the structural differences, see Table 2. While there is slight variation in the resolution of these six structures, the 0.37 Å difference between the high and low resolution structure is still small enough to allow direct comparison. Refinement of single conformer identical protein structure models with different ligands against the different datasets progressed nicely with similar refinement statistics across this panel of structures.

In order to compare the contact surfaces (CS) for the different ligands, areaimol[17] within the CCP4[18] program suite was run on the six different models. For the three no peptide (nop) structures, the HBP ligand contact areas are equivalent, with the largest ∂CS ~ 6 Å^2 between DHT and R1881. For the NTD bound structures, the coregulator makes equivalent contacts irrespective of ligand, with the largest ∂CS ~ 4 Å^2 between DHT and TES. Structure of the

Structure ID (see appendix 1)	Resolution	R/R$_{free}$	Contact Surface (HBP)	Contact Surface (AF2)
DHT_nop_22	1.47 Å	0.1553/0.1833	223.2 Å^2	n/a
DHT_ntd_132	1.63 Å	0.1523/0.1885	231.0 Å^2	474.9 Å^2
R18_nop_179	1.84Å	0.1818/0.2279	217.0 Å^2	n/a
R18_ntd_153	1.70 Å	0.1653/0.2057	206.0 Å^2	478.3 Å^2
TES_nop_140	1.70 Å	0.1591/0.1923	222.4 Å^2	n/a
TES_ntd_165	1.82 Å	0.1887/0.2259	217.4 Å^2	478.6 Å^2

Table 2. Structures used in the comparison of NTD bound states.

coregulator binding surface (AF2) is shown in Figure 13, models are colored blue, orange, green for DHT, R1881, and TES, respectively. Because binding of NTD to the different ligand states of the androgen receptor buries equivalent surface area at the coregulator binding surface, the discrepancy in binding energetics is likely not due to HBP ligand controlling how coregulators are engaged. If we turn our attention to the change in HBP CS between the nop and ntd structures within a given ligand, a clue presents itself. For the DHT liganded receptor, binding of the NTD actually increases the CS at the HBP by ~8 Å^2. For the TES liganded receptor, we learn that NTD binding actually decreases the CS at the HBP by ~5 Å^2. However, for the R1881 liganded receptor, there is an ~11 Å^2 reduction in the CS at the HBP.

This clue motivated closer inspection and manual building of the residues contacting the ligand. In Figure 13, electron density is shown for the HBP ligand and TRP741 for the DHT, TES,

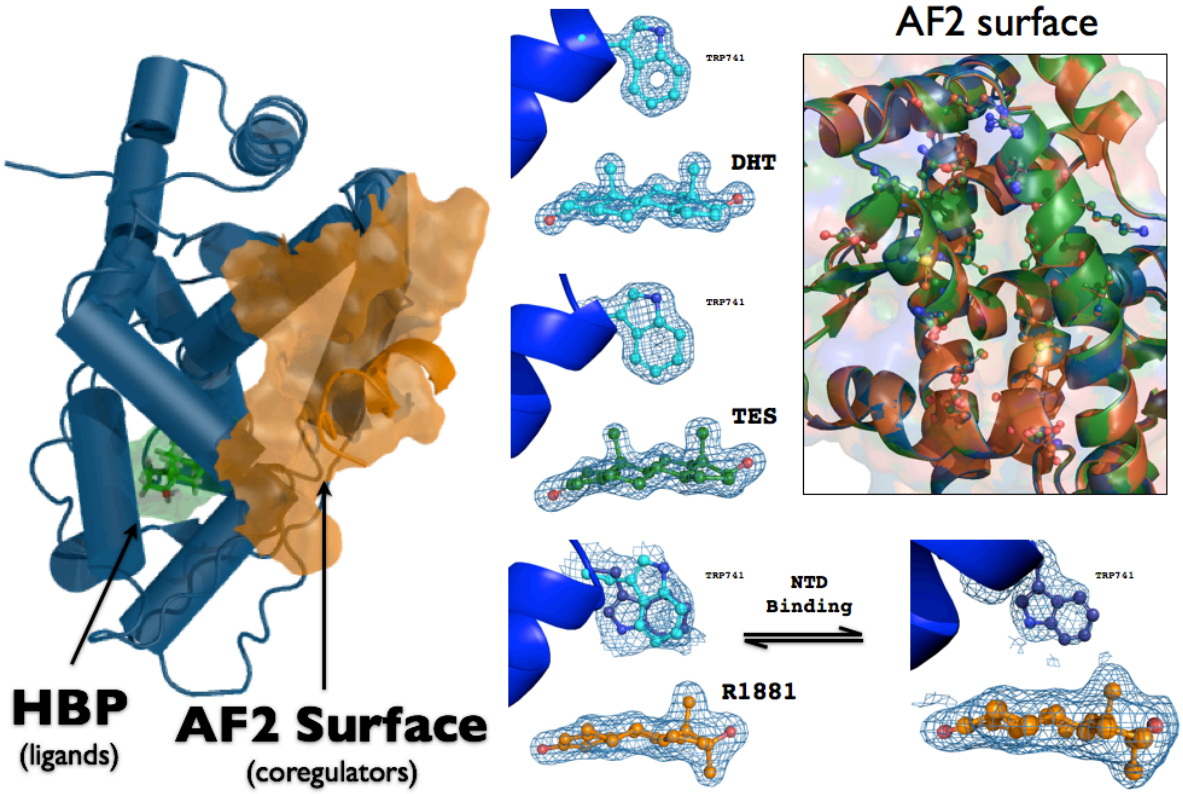


Figure 13. Coregulator binding remodels the HBP.

and R1881 liganded structures. The major difference between the natural ligands (DHT/TES) and the synthetic ligand (R1881) is the methylene group at C19. Returning to the steroid as a key analogy, this ridge is missing in R1881, and the tumblers of the receptor are engaged in a different manner. While the CS analysis was performed with single conformer models, the peculiar density at TRP741 for the R1881 liganded structure necessitated manual building of alternate conformers. With the missing methylene for R1881, TRP741 exhibits polystery and the structure of the HBP is more fluid.

Comparison of electron density in the NTD bound models of the DHT and TES liganded receptor revealed no gross changes in the steroid binding modes (not shown). In contrast, the NTD bound structure of the R1881 liganded receptor reveals that NTD binding collapses the structural ensemble to a single conformer at TRP741 (Figures 13 & 14). This conformer is distinct from the conformer that is found in the DHT and TES liganded structures. In addition to providing a potential mechanism for the weakened affinity of the NTD for the R1881 liganded receptor, this surprising result reveals the impact of coregulator binding on receptor structure.

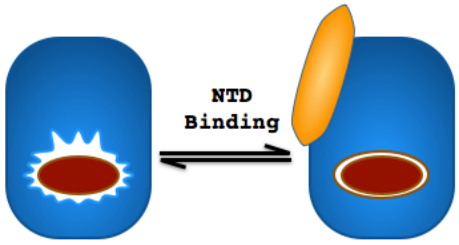


Figure 14. NTD binding collapses the structural ensemble of the HBP.

The study of nuclear receptor function has been primarily HBP ligand centric. Coregulator binding was the function that was activated by the HBP ligand. But this comparative study of ligands ultimately revealed that coregulator induced conformational changes play the key role in the structural explanation for weakened affinity to the R1881 liganded receptor.

Coregulator Binding Affects Receptor Structure:

Prior to this work, our understanding of coregulator binding was strictly from a steady-state perspective. This is evidenced in the language of the literature, where papers describe the accommodation of coregulators by the receptor.[19,20] While the binding data in these works is questionable, it is the absence of kinetic insight that allowed our neglect of deeper structural consequences. It is the kinetic aspect of the SPR measurement that mandates an appreciation of how coregulator functions to change the structure of the receptor. Therefore, I draw distinction to the coregulators as the protagonists of this story, and it is their actions on the receptor that give rise to changes in the half-life of different bound complexes.

Structure ID (see appendix 1)	Coregulator	Resolution	R/R_{free}
DHT_src14_16	$Src1_4$	1.96 Å	17.6/21.5
DHT_grip_115	$Src2_3$	1.82 Å	18.1/21.2
DHT_trap_102	TRAP100	1.80 Å	18.2/20.6
DHT_gsn_110	Gelsolin	1.81 Å	15.4/20.0
DHT_ntd_21	NTD	1.57 Å	12.2/16.9
DHT_pak_20	PAK6	1.38 Å	11.6/14.6

Table 3. Structures used in the comparison of coregulators bound to the DHT liganded receptor.

Loosely, there are two kinetic classes of coregulator interactions with the androgen receptor. Coregulators that engage the receptor with leucines (LXXLL) dissociate too rapidly for measurement, placing the upper limit of the half-life at < 0.4 seconds. The AR-specific coregulators carry phenylalanines at the +1 and +5 positions, and dissociate much more slowly with distinct half-lives ranging from ~2 to ~4 seconds.

Again, using ambient temperature xray diffraction, a number of different coregulator bound structures were obtained for the DHT liganded receptor. In Figure 15, SPR measurement and B-factor representations of the structures in Table 3, make plain the superficial contacts made by the general transcription factors that prevent long lived bound states of the androgen receptor.

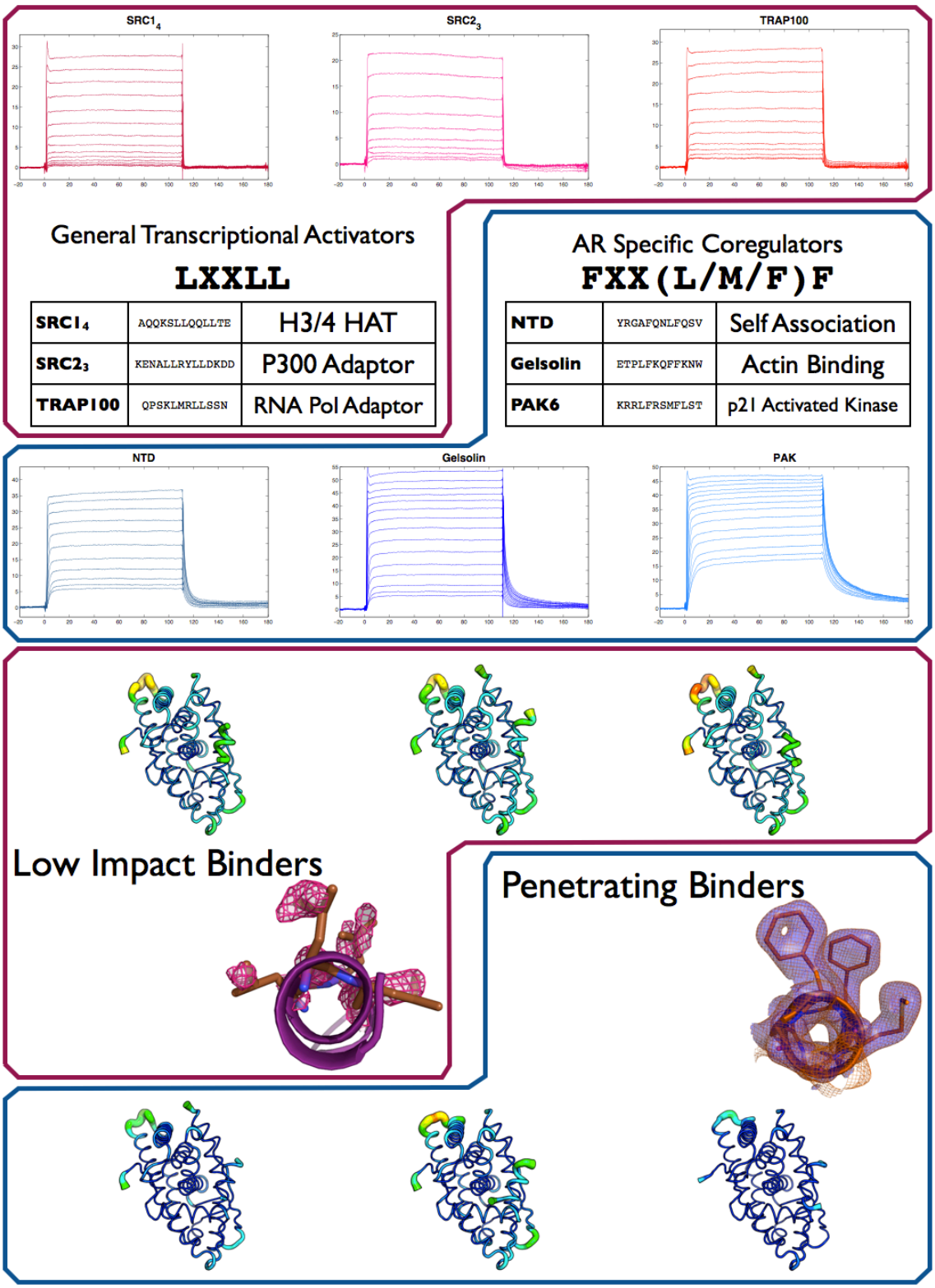


Figure 15. General transcription factors with LXXLL motifs dissociate rapidly, exhibit high temperature factors, and do not remodel the structure of the receptor. The AR specific coregulators dissociate slowly, bind stably, and elicit structural changes throughout the receptor.

Slow Dissociating Coregulators Elicit Deep Structural Changes:

In comparison of static structural changes at the contact surface for the different coregulator bound structures, the differences are quite subtle. Despite the high B-factors of the fast dissociating coregulator motifs, they still force the coregulator binding surface open and interact with the same residues as the slow dissociating motifs. The B-factors on the receptor side are also elevated by interaction with the fast dissociating coregulators, but the conventional modes of analysis have rarely considered this aspect of the complexes. In extreme cases, fast dissociating complexes are sometimes assigned triple digit B-factors,[13] but the structural changes are analyzed with utter disregard for this evidence of highly mobile interactions. In certain structures, I too observe elevated B-factors (70<B<90) in the bound coregulator motif for high resolution structures of fast dissociating complexes. This can be conflated with occupancy issues, but I assert >90% occupancy based on the conditions that the crystals are grown. For all structures with bound coregulators, the crystallization solution

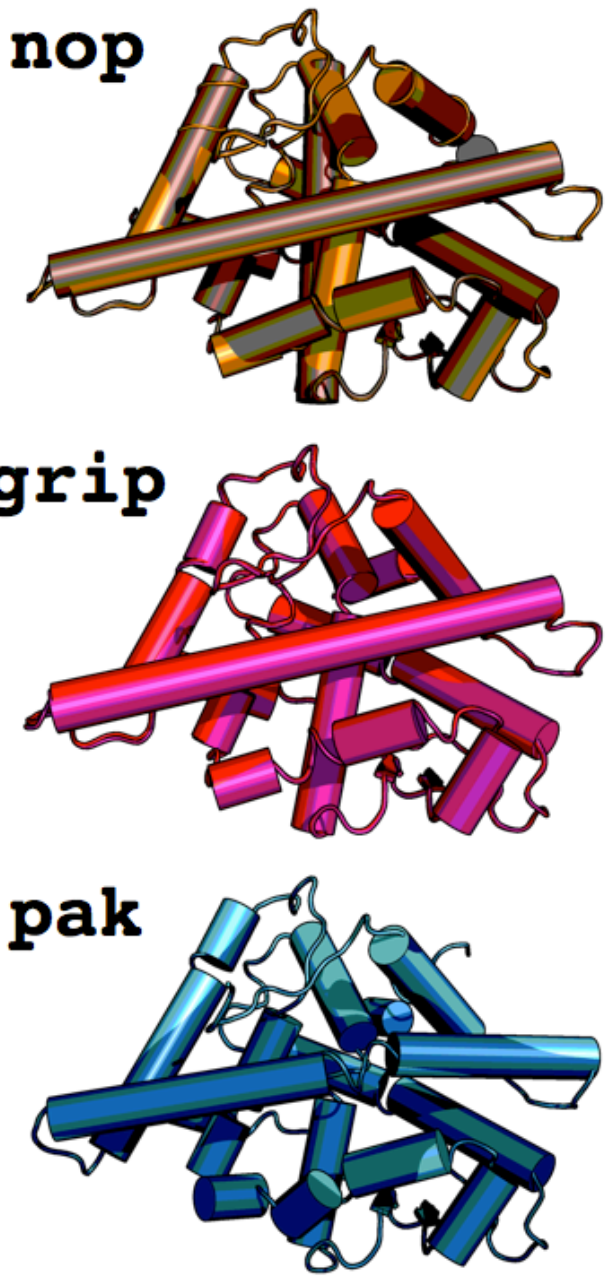


Figure 16. Coregulator identity impacts Helix 10 on the opposite face of the androgen receptor LBD.

contains the coregulator motif at > 100X the $K_{1/2}$. Rather, I interpret these high B-factor interactions as fundamental to how they engage the receptor. These general transcription factors are only using the receptor to bury their hydrophobic surfaces, and their mobile binding modes give a lower energetic barrier to dissociation.

With an eye towards appreciating multiple conformations of bound coregulators, it seems that the fundamental difference that differentiates the kinetics of these interactions is not what surface is engaged, but changes to how the surface is engaged. For the fast interacting coregulators, the interactions do not penetrate beyond the first shell. In contrast, the slow dissociating coregulators firmly grasp the coregulator binding surface and structural changes propagate throughout the structure of the receptor. In Figure 16, a view of the surface opposite the coregulator binding surface is presented for two models of NOP, GRIP (or $SRC2_3$), and PAK. For the GRIP structures, the back side of the molecule is indistinguishable from the NOP structures. The long helix that is in the foreground is H10/11. The PAK models reveal a significant bend to this structural element. While this does not inform the path by which the structural change is transmitted by the coregulator, it dramatically reveals how penetrating the binding mode is for the slowest dissociating coregulator.

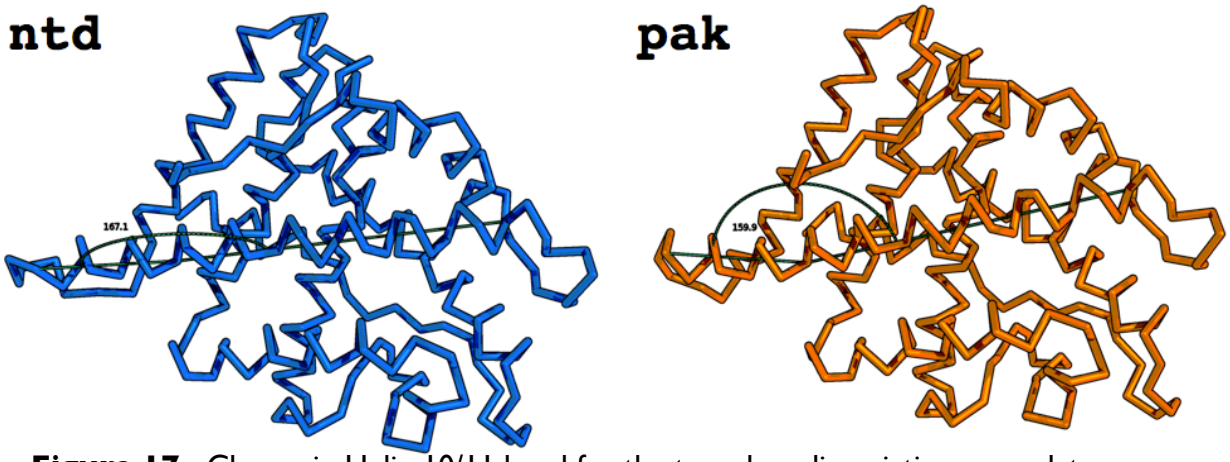


Figure 17. Change in Helix 10/11 bend for the two slow dissociating coregulators NTD and PAK.

For the extreme case of comparing structural changes for one of the fastest dissociating coregulators with the slowest (>8.5 fold difference in $t_{1/2}$), coregulator identity controls the conformation of H10/11. But is this a general mechanism by which coregulators perturb the receptor structure along this helix, increasing the barrier for dissociation? Figure 17 compares the H10/11 bend between the NTD and PAK bound models. For NTD, which forms complexes of moderate lifetime ($t_{1/2}$ ~ 1s), H10/11 shows a bend of 167°, or 13° relative to straight. The PAK bound model ($t_{1/2}$ ~ 3.5s) shows a bend of 22° relative to straight. This 8° difference in helical bend represents a major structural change between the two complexes. Interestingly, the C-terminal end of this helix (residues 863-884) show a preponderance of disease mutations. Future work should explore the effect of disease mutations on the bend of this helix and the kinetic discrimination of coregulator binding motifs.

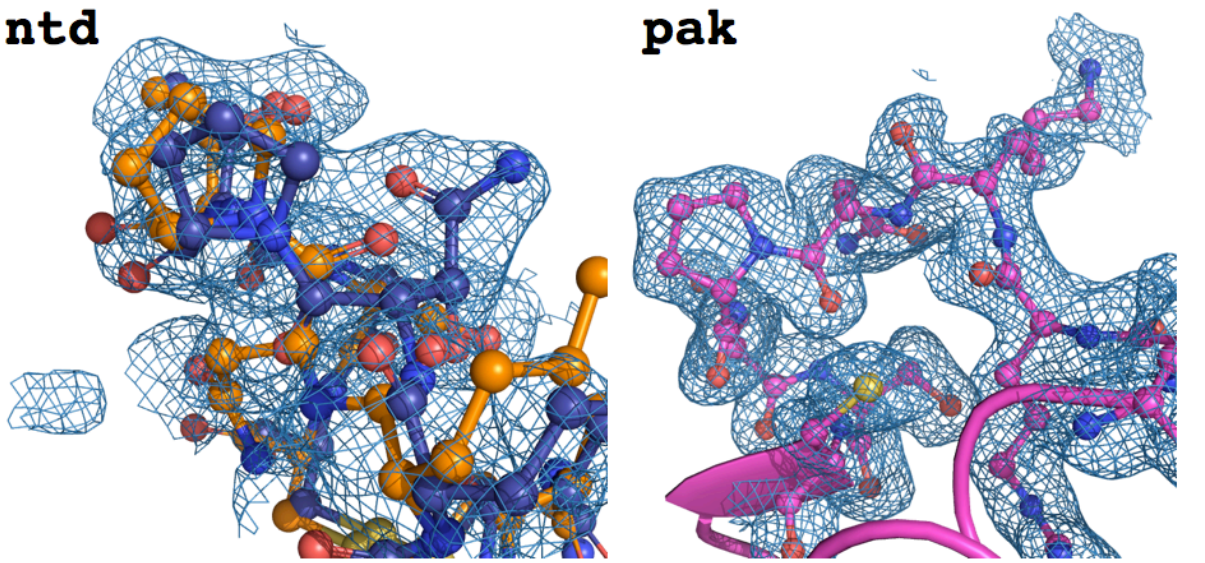


Figure 18. Structural heterogeneity of Loop 9/10. Electron density for L9/10 in the NTD bound model indicates the presence of both cis and trans PRO849 conformers. PAK bound model exclusively samples the trans PRO. Inter-species sequence alignment of AR reveals conservation of this loop sequence, suggesting a conserved mechanism for reading of coregulator identity.

In Figure 17, the helical bend provides a simple quantitation of the differences in structural perturbations by the NTD and PAK coregulators. At the N-terminal region of H10/11, there is an even larger change in the polysteric properties of the loop bridging H9 and H10. In Figure 18, electron density of this region is shown for the two models. For the NTD bound model, there is clear evidence of structural heterogeneity, and possible co-existence of cis and trans prolines. For the PAK bound model, the density indicates a single conformer. This region is well conserved between species, and the mechanistic implications are explored in the following section.

An Ensemble Perspective of Structure:

For the highest resolution ambient temperature structures that were obtained, conventional refinement inadequately describes the data. In contrast to past cryo structures of the receptor, numerous positions show evidence of polystery. While it is possible to manually build in the requisite multi-conformer models, this process carries an inherent modeling bias. As static structural changes have, in the past, failed to capture the structural consequences of coregulator binding, the polystery needed to be captured in an unbiased way.

Recent advances in crystallographic refinement have opened the door for quantitative comparisons of structural ensembles. As implemented in Phenix,[21] ensemble refinement provides an ideal platform for the selection of alternative conformations in the absence of map or modeling bias.[22] In contrast to conventional refinement, ensemble refinement couples molecular dynamics simulations to crystallographic refinement. And instead of giving a single structure, an ensemble model consisting of many equally weighted models is produced.

Twenty-four parameter combinations were explored in order to find the optimal input parameters for ensemble refinement of a high resolution cryo dataset of the AR LBD. In order to draw comparisons, this same parameter set was then used for each of five ambient datasets. For the cryo dataset, ensemble refinement resulted in roughly equivalent refinement statistics (∂R = -0.0027, ∂R_{free} = +0.0030). In contrast, ensemble refinement of the ambient temperature datasets improved refinement statistics for all structures (mean ∂R = -0.025, mean ∂R_{free} = -0.011).

Motivation for Dihedral Angle Processing

The resultant ensemble models are rich in data content and novel comparison between structures can proceed with analytics that are usually reserved for NMR structures or molecular dynamics simulations.

While a metric such as the root mean squared fluctuation will give information on local flexibility, when reduced to individual residues, calculated rmsf values might be inflated for rigid positions that are flanked by regions of relative disorder. This loss of per residue resolution prompted the consideration of dihedral angle distributions for structural comparison. Further, dihedral space has been used to identify allosteric networks.[23]

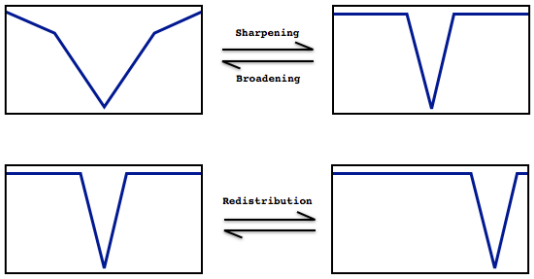


Figure 19. The two types of structural perturbations at the dihedral level. Sharpening and broadening can be quantitated out by order parameter processing. Redistribution of dihedral sampling can be quantitated by calculation of pairwise distance correlation coefficients between ensemble dihedral distributions.

To quantify the breadth of individual dihedral angle distributions, order parameters were calculated[24] using a bootstrap with replacement strategy. To allow position specific comparison between ensemble models, dihedral order parameters were summed within each residue carrying the bootstrap determined uncertainty. Changes in order parameter were then taken as the difference in order parameter vectors for each ensemble model.

Correlation coefficients of ensemble dihedral distributions report on the redistribution of conformers. Again, using a bootstrap with replacement strategy, correlation coefficients allowed dihedral level comparison between ensemble models. In conventional crystallographic structural terms, the order parameter processing will capture changes in flexibility, while the correlation coefficient processing reflects structural changes. Appendix 3 details the Matlab code for these analytics.

Dataset (Appendix 1)	input stats	R, R_{free}	Resol-ution	Ensemble Size	Loop State
CRYO_MV_1pt4	13.75, 16.37	13.48, 16.67	1.41Å	100	reduced
DHT_nop_22	15.36, 17.28	12.26, 15.97	1.47Å	84	reduced
DHT_nop_48	15.47, 17.78	11.85, 17.03	1.64Å	167	disulfided
DHT_ntd_21	14.79, 17.72	12.30, 16.37	1.57Å	56	disulfided
DHT_ntd_132	14.86, 18.32	12.40, 16.83	1.63Å	50	reduced
DHT_pak_20	14.57, 16.82	12.52, 16.31	1.38Å	125	reduced

Table 4. Ensemble Refinement Statistics for the Structural Comparisons of cryo vs. ambient vs. NTD vs. PAK. NTD and nop ambient datasets were often found with a disulfide between CYS844 and CYS852 at the base of L9. In contrast, across 6 different datasets of the PAK bound receptor, this disulfide was always reduced. For comparison between structures, only the reduced states are considered.

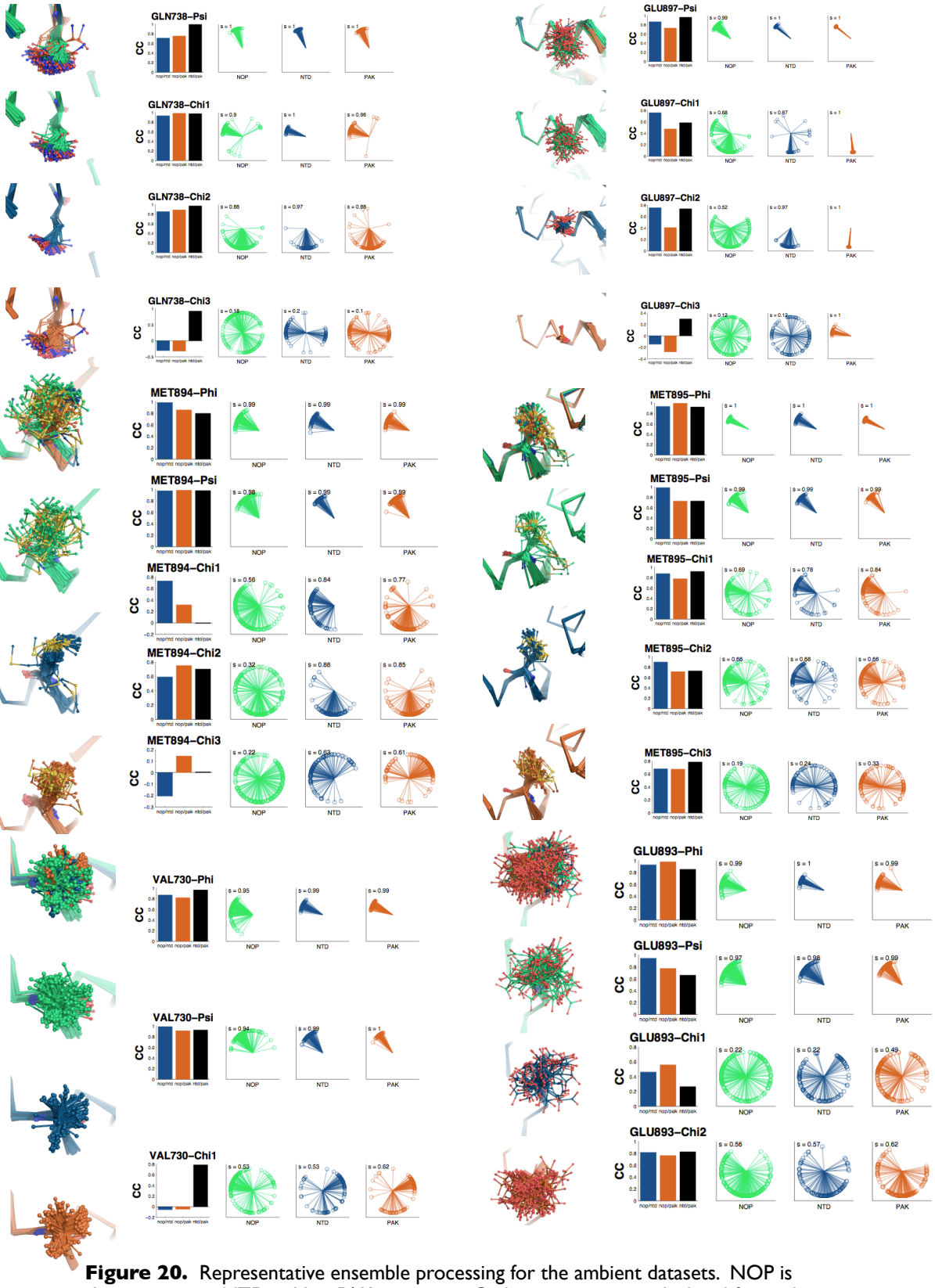


Figure 20. Representative ensemble processing for the ambient datasets. NOP is shown in green, NTD in blue, PAK in orange. Order parameter is calculated for each dihedral, and correlation coefficients (CC) are computed between the three models.

Ambient Temperature Crystallography Reveals the Structural Plasticity of the AR:

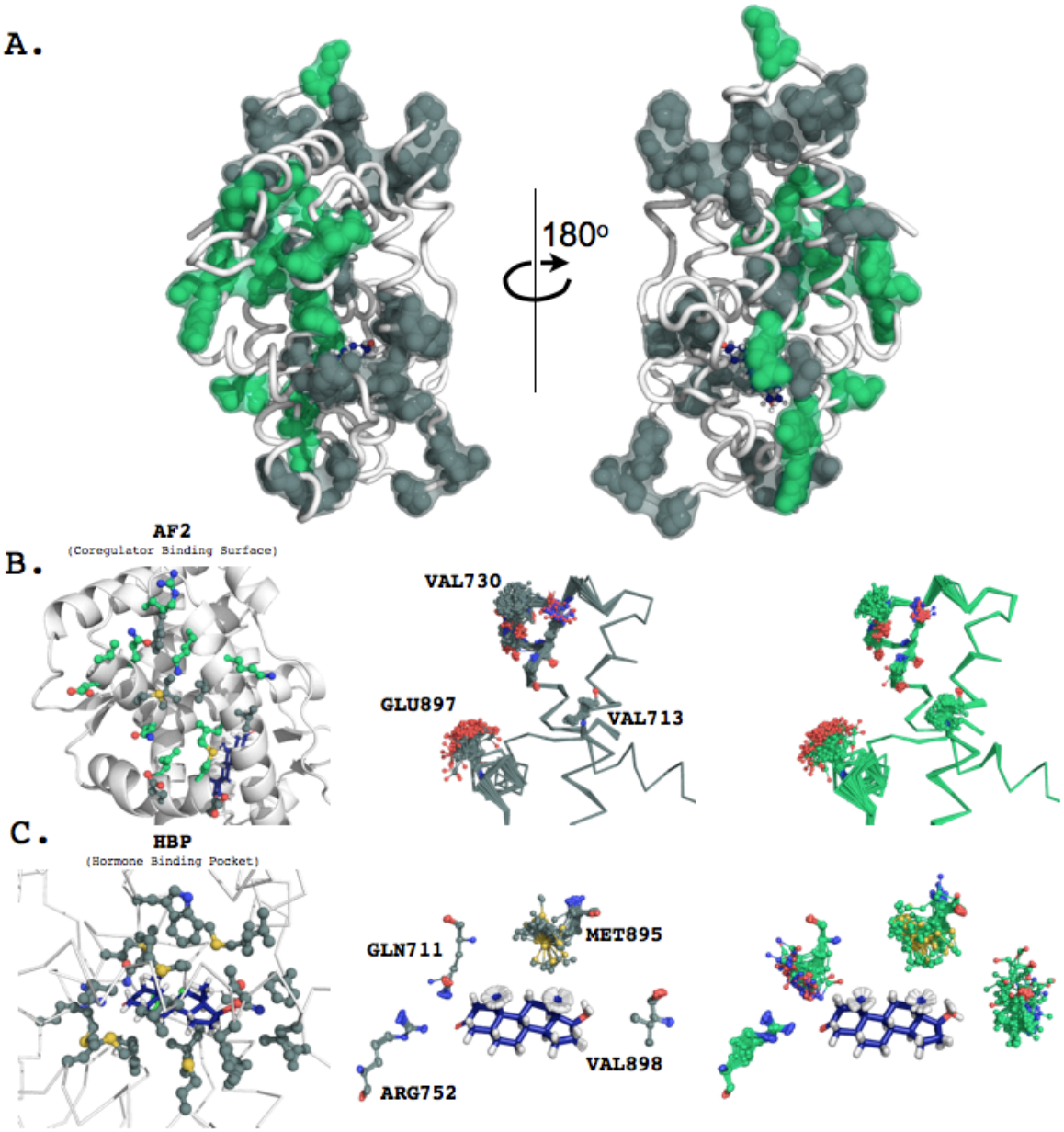


Figure 21. Comparison of the cryo and the ambient temperature ensemble models.

Differences in order parameters between the cryo dataset and the ambient temperature dataset were determined by subtraction of the per residue order parameter vectors ($\partial s = s_{i,RT} - s_{i,CRYO}$). The resultant ∂s distribution is approximately normal with a mean less than zero ($\mu = -0.12$, $\sigma = 0.39$), indicating that many positions in the AR LBD sample a more sharply defined

energy landscape under cryo conditions than at ambient temperature. However, the ambient temperature associated broadening of dihedral angle distributions is not global and many positions have higher order parameters at ambient temperature. As a simple filter for significance, residues with ∂s values outside of one standard deviation from the mean are mapped to the structure in Figure 21 A. Considering the residues that show sharper dihedral distributions in the cryo dataset (gray spheres), there are two major clusters that bridge secondary structural elements: 1. The C-terminal end of helix 5 connects via helices 1 and 2 to the N-terminal end of helix 12; 2. Helices 8, 9, and 10 comprise another island of relative ordering in the cryo dataset. For the residues with sharper distributions in the ambient temperature dataset (green spheres), the primary cluster stretches from the top of the AF2 site

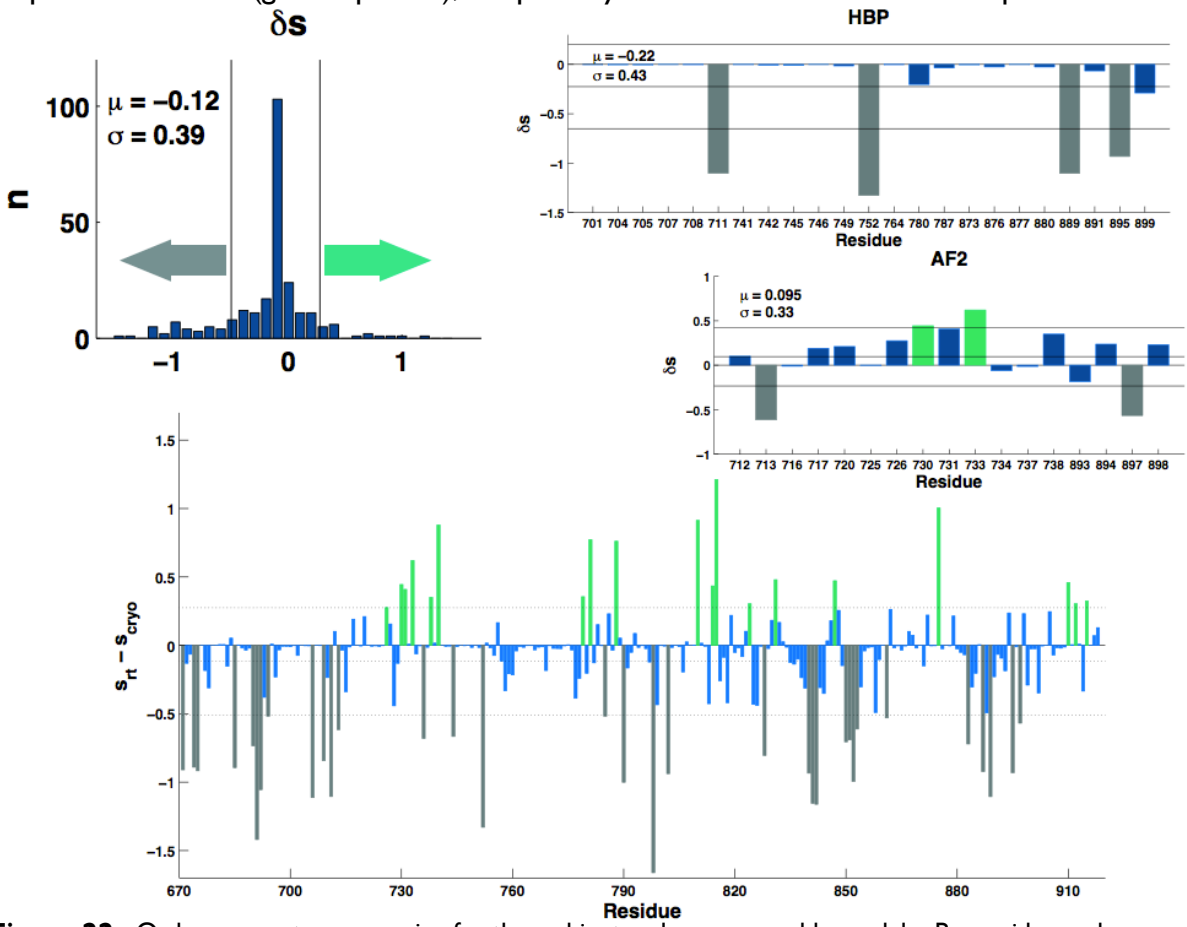


Figure 22. Order parameter processing for the ambient and cryo ensemble models. Per residue order parameters were calculated by s_{rt} - s_{cryo}. The global ∂s distributions is shown, in addition to the focused ∂s distributions for the HBP and the coregulator binding surface.

(helix 3) through helices 4, 8, and 9 to seemingly stabilize the C-terminal side of helix 12 relative to the cryo dataset.

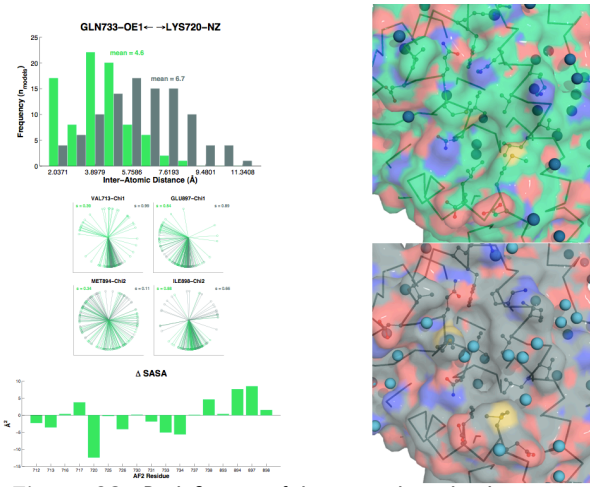


Figure 23. Redefinition of the coregulator binding surface at ambient temperature. Cryo in grey, ambient in green. Change in per residue solvent accessible surface area (SASA) was calculated by $SASA_{rt}$ - $SASA_{cryo}$.

The Coregulator Binding Surface is More Occluded Under Ambient Temperature

To explore local differences at the coregulator binding surface, the ∂s distribution was recalculated for the 17 residues that comprise the AF2 site. While the N-terminal side of helix 12 is characterized by a sharper ensemble distribution under cryo, the majority of residues that comprise the AF2 site appear more ordered under ambient temperature ($\mu = 0.095$, $\sigma = 0.33$). In the left hand panel of Figure 21 B, AF2 residues with sharper distributions at ambient temperature are colored green, while those with broader distributions are colored grey. At the bottom of the AF2 site, GLU897 and VAL713 are quite restricted in the cryo dataset. In contrast, the upper region of the AF2 site samples a tighter distribution in the ambient temperature dataset by a sharpening of the backbone dihedral distributions between residues 726 through 733. In fact, comparison of the solvent accessible surface area for the AF2 site residues across all models in the ensembles reveals a 10% change in the solvent accessible surface area. Under cryo conditions, the AF2 site exposes an additional 8.3 $Å^2$ of solvent accessible surface area relative to the ambient temperature dataset. This surprising result is attributed to increased solvation under cryo as waters interfere with weak hydrogen bonds that seal the upper half of the AF2 surface in the ambient temperature dataset.

H12 and the Hormone Binding Pocket Exhibit A Broadened Ensemble Under Ambient Temperature

The 23 residues that form the buried HBP were also considered in a focused ∂s distribution. In contrast to the relative sharpening of the AF2 site, nearly all positions in the HBP exhibit ensemble broadening in the ambient dataset (μ = -0.22, σ = 0.43). In Figure 21 C, the ensemble models for the HBP are shown. While the majority of HBP residues show only slight broadening in the ambient dataset, four positions are significantly broadened: GLN711, ARG752, VAL889, and MET895. These significant contributors to the overall negative distribution are shown in ball and stick in the middle and right panel. Interestingly, mutations at these four positions (Q711E, R752Q, V889M, and M895T) have been reported in cases of androgen insensitivity syndrome. It is interesting to note the broad distribution in χ angles for MET895 in the cryo dataset; under ambient temperature, this broadening extends to the backbone order parameters affecting the overall positioning of helix 5. VAL889 also exhibits a broadening in backbone space as the entire loop leading into helix 12 samples a more diverse ensemble in the ambient dataset. In contrast, backbone order parameters for ARG752 and GLN711 are only slightly broadened in the ambient dataset. However, the χ angle distributions at these positions become much broader under ambient temperature. GLN711, which is largely static in the cryo dataset, completely abandons its position under ambient temperature with broadened distributions at all sidechain dihedrals. ARG752 also shows broadening across all sidechain dihedrals, but in contrast to GLN711, the functional role of ARG752 in coordination of the carbonyl oxygen on dihydrotestosterone is maintained.

In summary of the comparison between the ensemble models derived from the cryo dataset and the ambient temperature dataset, the principal functional sites of the AR LBD are

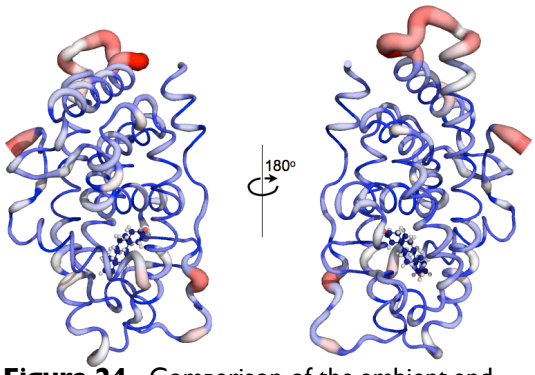


Figure 24. Comparison of the ambient and cryo ensemble models by per residue correlation coefficient. Thick and red is low correlation, while thin and blue is high correlation.

redefined. While cryo conditions lead to ensemble broadening and increased solvent accessibility at the AF2 site, the ambient temperature dataset reveals a surprising fluidity to the hydrophobic core of the AR LBD. Perhaps herein lies an important lesson in the temperature dependence of hydrophobic structural elements. The freezing that accompanies cryo data collection can trap waters at hydrophobic surfaces and disguise occluded sites as accessible. The freezing process also has the potential to remove some of the inherent fluidity in greasy hydrophobic cores.

SPR Measurement of Coregulator Interactions With the AR

In order to explore the energetics and kinetics of interactions at the AF2 site, surface plasmon resonance (SPR) was used to monitor the association of coregulator fragments to the AR LBD. By site specific biotinylation of the AR LBD, the classical configuration was inverted and rather than flowing purified receptor across immobilized peptides, stable surfaces of AR LBD were assayed for interaction with injected coregulator samples. This approach provides the advantages of minimal protein consumption and homogeneity of the injected samples. Coregulator interactions are relatively weak, and accurate determination of affinities requires concentrations in the tens of micromolar. While this concentration is certainly accessible for LBDs, trace contaminants from recombinantly expressed proteins (bacterial chaperones) become relevant at these concentrations and confound measurement. In the SPR experiment, sample homogeneity is of the utmost importance, and, simply, it is far easier to achieve pure peptide samples than pure protein samples. Under optimal buffer conditions at 8C, the

receptor surfaces proved stable for days allowing repeated measurements and monitoring of surface activity.

In this study, the two coregulator fragments are NTD: YRGAFQNLFQSV and PAK: KRRLFRSMFLST. For each measurement of AR interaction with NTD or PAK, individual peptide binding isotherms were collected from low to high concentration, alternating between coregulators. Isotherm replicates are obtained by scaling responses between different density surfaces.

In Figure 25 A, representative sensorgrams for the interaction of the NTD (blue) and PAK6 (orange) fragments with the same low density surface of AR LBD are shown. In qualitative comparison of these ~3-fold dilution series, both coregulator fragments are shown to approach saturation at a similar rate but do so with markedly different kinetics.

NTD and PAK Have Similar Interaction Energies

In total, four independent experiments were used to determine the affinity of NTD or PAK for AR. Duplicate fifteen point isotherms (*) are plotted with fits (o). While affinities are

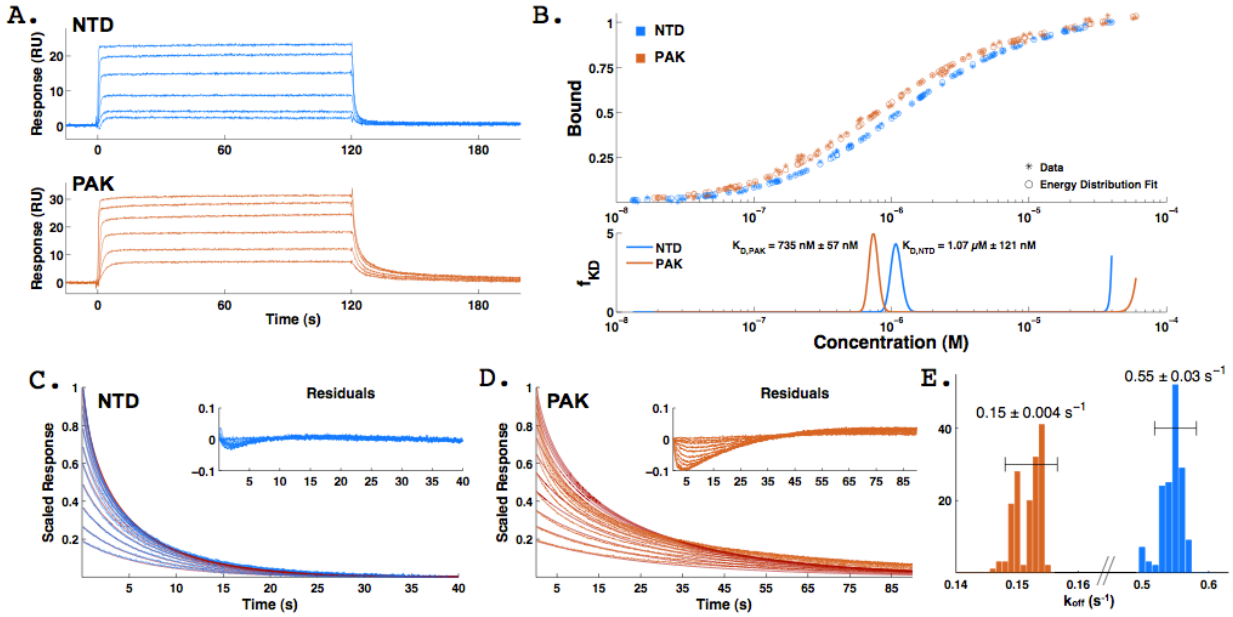


Figure 25. SPR measurement of NTD and PAK interaction with the DHT liganded receptor.

well determined with a standard langmuir binding model, the use of adsorption energy distributions gives significantly lower error and provides a more accurate picture of these interactions. The associated energy distributions are presented in the lower panel as the scaled mean of re-centered distributions. In this style, error in dissociation constants can be taken as the half width at half maximum. The largest uncertainty in this experiment is the absolute concentration of the peptides.

Were an interaction to proceed with a single free energy change, the associated energy distribution would be a delta function. However, molecular recognition and association do not necessarily follow a single two-state reaction coordinate. For a single molecular recognition event, there is decidedly a single free energy change. However, when measured in bulk, structural heterogeneity in the unbound molecules will manifest as a distribution of free energies of interaction.

While it is tempting to ascribe differences in breadth of affinity distributions to structural heterogeneity in the bound and unbound states, these differences are most likely reflective of experimental error.

PAK Dissociates More Slowly Than the NTD

The most appreciable qualitative difference between these interactions is the slow dissociation for PAK relative to NTD. In determination of dissociation kinetics, 54 traces from three different experiments were considered. Using a model that explicitly treats the effects of mass transport in determining dissociation kinetics, global k_{on} / k_{off} were fit while allowing k_{tr} and s_{max} to float for the individual experiments. In order to calculate error in parameter estimation, the data was iteratively fit (n = 200) with random weighting of each trace's contribution to the sum of squared error calculation. Representative traces from the

dissociation phase for NTD and PAK are plotted in Figure 25 C & D, respectively (note the difference in x-axis scaling). Fits from global parameter estimation are plotted in red.

Dissociation of the NTD is well described by a single off rate, as evidenced by only slight deviation from uniformity in the residuals (Figure 25C, inset). In contrast, the dissociation of PAK is significantly slower and shows marked systematic error in the residuals (Figure 25D, inset). This ~10% deviation of the fit indicates an inadequacy of a single dissociation model and suggests heterogeneous dissociation kinetics. Inclusion of multiple chemical off rates in the dissociation model improves the fit residuals but there is loss of precision in parameter estimation. For this reason, and in the interest of parsimony, the fit k_{off} values from the single dissociation model are suitable for comparison between NTD and PAK. The distribution of fit k_{off} values are plotted in Figure 25 E (NTD in blue, PAK in orange) on a discontinuous x-axis. With a k_{off} of 0.55 s^{-1}, the NTD bound AR LBD has an average half-life of ~1.2 seconds. PAK dissociates with a k_{off} of 0.15 s^{-1}, giving a half-life of ~4.6 seconds. By extrapolation of these off rates to the affinity constants determined by fitting of the steady state, the NTD interaction proceeds with a k_{on} of 4.8 x 10^5 $M^{-1}s^{-1}$, while PAK also shows a significantly slower k_{on} of 1.9 x 10^5 $M^{-1}s^{-1}$. This almost four-fold difference in the rate of dissociation is remarkable in light of the comparable energetics. Further, given the similar composition of these two fragments and the small size, it is surprising that such kinetic discrimination is possible at the AF2 site.

The use of SPR to study these interactions was motivated by comparison, not in depth mechanistic insight. SPR measurement of AF2 site recognition of coregulator fragments reveals weak binding in the micromolar range. The affinities of interaction with NTD or PAK are comparable, with PAK binding ~1.4-fold more tightly. The similar thermodynamics of interaction belie a significant discrepancy in the kinetics of dissociation. By comparison of steady-state

binding, we learn that NTD and PAK have similar affinities for the AR LBD. While there is little reason to extend the *in vitro* biophysical measurement of the interaction between an isolated LBD and a 12-mer to the complex environment of the cell, the kinetic differences intrigue and warrant further scrutiny from this reductionist perspective.

The crystallographic structures of NTD and PAK bound AR LBD were determined using ambient temperature xray diffraction. Refinement of these structures was conducted in the same manner as the apo datasets before ensemble refinement was used to generate a series of models. While the comparison of the cryo and ambient datasets redefines the structural plasticity of the AR LBD, this lateral comparison reveals perturbations to the ensemble as a consequence of coregulator binding. It is important to note that the three structures under consideration were all crystallized at pH 8 in 500mM Li_2SO_4, which facilitates direct structural comparison under similar conditions to the SPR assay buffer (pH 7.6, 150mM Li_2SO4).

Coregulator Binding Shapes the AF2 Surface

For both bound structures, the coregulator fragments are bound at 100% occupancy, and the core residues of interaction (**F**QN**LF** / **F**RS**MF**) show correspondingly low b-factors in conventional refinement (μ_{NTD} = 22.8, μ_{PAK} = 15.4). In the left panel of Figure 27 A, the AF2 surface of the APO AR LBD is highlighted with electron density shown for key LBD residues involved in coregulator recognition. For both MET734, MET894, and GLU897 electron density indicates little energetic preference in sidechain dihedral distributions. These broad distributions are plotted radially to the right of the structure. In the middle and right panel, the bound AF2 site is shown

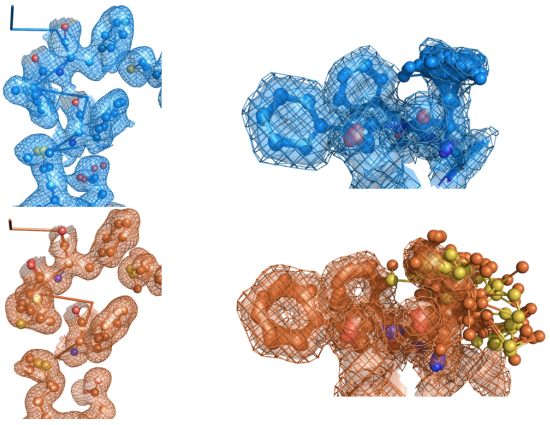


Figure 26. Ensemble binding modes of the coregulators. NTD in blue, PAK in orange.

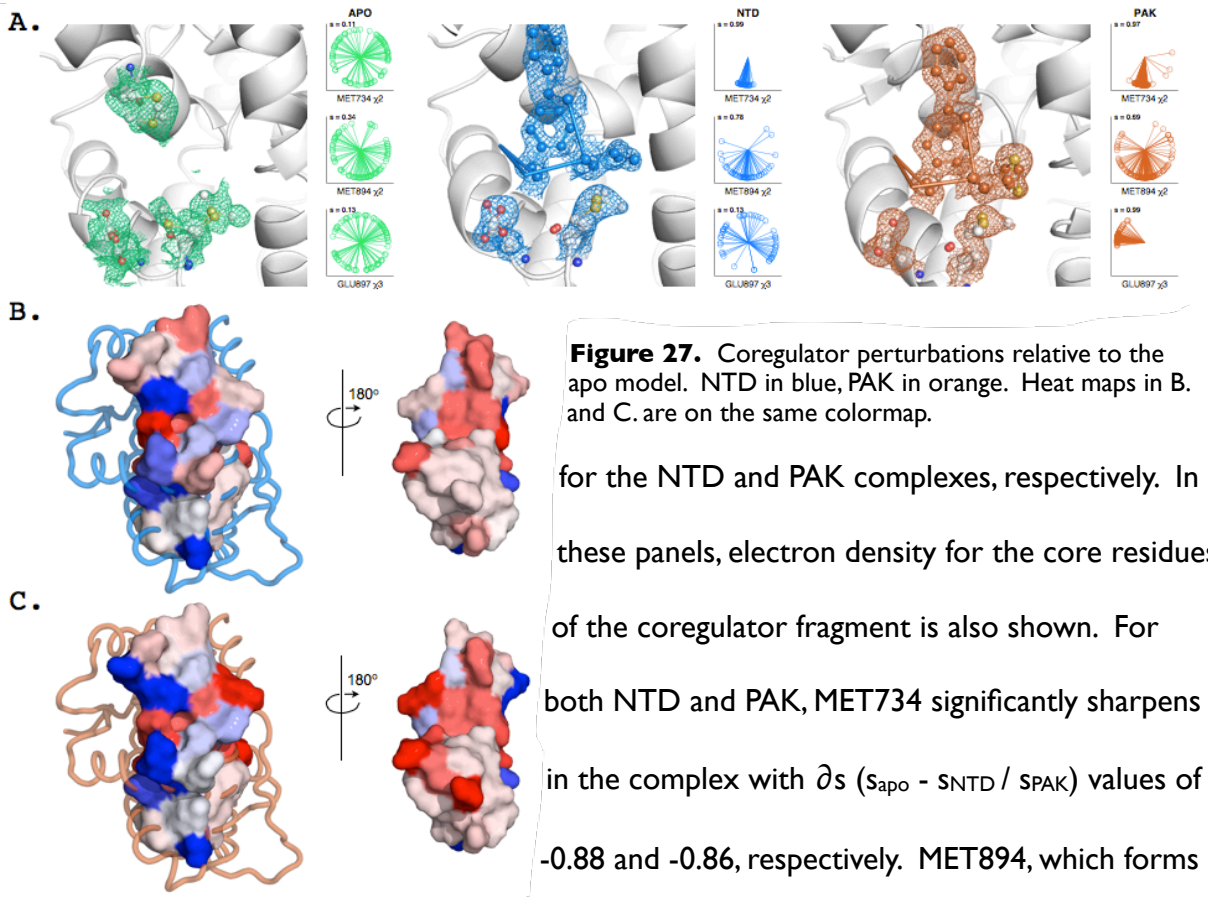


Figure 27. Coregulator perturbations relative to the apo model. NTD in blue, PAK in orange. Heat maps in B. and C. are on the same colormap.

for the NTD and PAK complexes, respectively. In these panels, electron density for the core residues of the coregulator fragment is also shown. For both NTD and PAK, MET734 significantly sharpens in the complex with ∂s (s_{apo} - s_{NTD} / s_{PAK}) values of -0.88 and -0.86, respectively. MET894, which forms the bottom of the hydrophobic pocket that accommodates the +4 position in the coregulator motif, also significantly sharpens (∂s_{NTD} = -0.44, ∂s_{PAK} = -0.35). In the PAK structure, this pocket is filled by the methionine residue on the coregulator with density indicative of alternative conformations. In contrast, the leucine of NTD at this position is docked with a single conformation. The ensemble order parameter processing captures this with a comparatively sharper distribution at MET894 for the NTD bound LBD. The charge clamp residue GLU897 undergoes sharpening at X1 and X2 dihedral distributions for both NTD and PAK, but only PAK sharpens the X3 position in firmly anchoring the N-terminus of the helical coregulator fragment. In order to calculate the amount of buried surface area, all models in the ensemble were considered. On the ensemble scale, both complexes reveal a dynamic bound state outside of the core residues. On average, the NTD buries 504 ± 43 Å^2, while the PAK

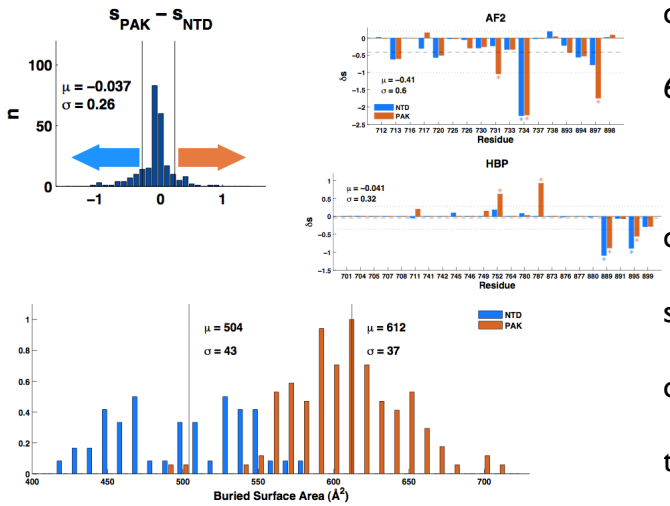


Figure 28. The global ∂s distributions were calculated as s_{PAK} - s_{NTD}. Focused ∂s distributions were calculated relative to the nop ambient model, NTD in blue, PAK in orange. Contact area at the coregulator binding surface was calculated for the ensemble models.

coregulator extends the buried surface to 612 ± 37 Å^2.

These local perturbations to the coregulator binding site are somewhat straightforward and can be captured with conventional refinement. It is established that MET734 undergoes a significant conformational change upon coregulator binding.[25] Further, there is evidence that coregulators can engage the N-terminal charge clamp (GLU897) to differing degrees. However, ensemble refinement also provides a quantitative metric to investigate perturbations to the entirety of the receptor's structure.

By comparison of order parameters, residues that exhibit either ensemble sharpening or broadening in the coregulator bound state are identified. In considering ensemble broadening, this reflects a region of the structure that has a well defined energetic minimum in the unbound state; in broadening, coregulator binding promotes energetic degeneracy in local structural states. Ensemble sharpening represents the inverse phenomenon where a region of the structure has little energetic preference in the unbound state, but complex formation reduces the structural sampling.

Shared Perturbations to the Ensemble

At the backbone level, ~53% (n = 133) of residues in the AR LBD respond to coregulator binding, either NTD or PAK, with sharpening of their Φ- and Ψ-dihedral distributions. Spatially, these sharpened positions cluster on the surface of the molecule with a

prominent grouping at the coregulator binding surface. Nearly all positions along helix 12 undergo sharpening with significant ordering of the loop connecting helix 10/11 to helix 12. Interestingly, the local sharpening effect seems to propagate outwards to distant surfaces along helices 7, 9, and the N-terminal half of helices 10/11, which also exhibit sharpening. At the sidechain level, sharpened positions (~48%, n = 112) are, again, principally at the surface of the AR LBD. However, sharpening of X-dihedral distributions of the protein appear to form more contiguous clusters through the interior of the protein, especially at the AF2 side of the HBP and on the upper side of helix 5.

In the absence of entropic compensation, one might expect to exclusively observe ensemble sharpening as a consequence of coregulator binding to the receptor. Surprisingly, there are numerous positions in the AR LBD that respond to coregulator binding by broadening in both backbone (14%, n = 35) and sidechain (18%, n = 45) dihedral distributions. In contrast to the principally surface positions that respond to coregulator binding with ensemble sharpening, the majority of broadened residues are found on the interior of the protein. The major cluster of broadening is found on the backside of the HBP, coupling the bottom of helix 5 to the β-sheet and helices 7 and 10/11. A secondary cluster of broadening is found along the middle of helix 8.

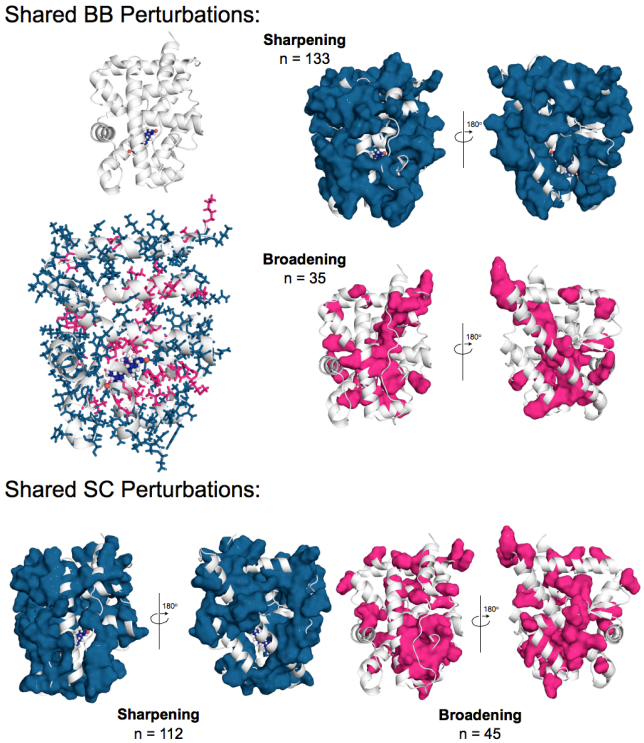


Figure 29. Shared perturbations to the ensemble. Positions that exhibit sharpening are colored blue, while broadened positions are in magenta.

In the coregulator bound state,

roughly half of the AR LBD responds with sharpening in the dihedral distributions of the ensemble. Intuitively, there is a clearly visible cluster of sharpening at the coregulator binding surface. While the extent of propagation to distant surfaces of the receptor is striking, this is consistent with an overall structural stabilization upon coregulator binding. At the coregulator binding surface, slight backbone broadening is observed at the periphery, while GLN738 and ILE898 exhibit broadening in sidechain dihedral distributions. The extent of broadening through the interior of the AR LBD upon coregulator binding is striking. If the APO receptor represents the ideal packing of the sidechains that comprise the core of the AR LBD, the coregulator bound state sacrifices this stability in engaging the residues of a coregulator fragment. Local eversion of hydrophobic sidechains upon engaging the coregulator destabilizes the second shell, whereby broadening distributions propagate along the bottom of helix 5 and the top of helix 8. While the cluster along helix 8 is flanked by regions of sharpening, the broadening on the hormone side of helix 5 extends to a large cluster of residues at the A-ring pocket of the HBP.

By comparison of shared perturbations to order parameter, coregulator responsive regions of the structure are identified. This position specific analysis of shared perturbations provides a basis for testing the hypothesis of coregulator specific structural changes. For the null hypothesis of NTD and PAK eliciting the same changes to the receptor, with the slow kinetics of PAK dissociation as a product of moving the receptor further along the same reaction coordinate, these shared perturbations should be of greater magnitude in the PAK bound ensemble. Interestingly, for both the broadened and sharpened sets of shared perturbations, there are insignificant differences in the magnitude of NTD or PAK changes relative to APO (Backbone: $\mu_{NTD,sharp} \sim -0.06$, $\mu_{NTD,broad} \sim 0.03$; $\mu_{PAK,sharp} \sim -0.05$, $\mu_{PAK,broad} \sim 0.02$; Sidechain: $\mu_{NTD,sharp} \sim -0.22$, $\mu_{NTD,broad} \sim 0.17$; $\mu_{PAK,sharp} \sim -0.22$, $\mu_{PAK,broad} \sim 0.16$).

Coregulator Specific Perturbations to the Ensemble

At the ensemble level, differential perturbations to the coregulator binding surface and the hormone binding pocket are primarily in magnitude. The change in order parameter with respect to the APO structure is mapped to these two surfaces in Figure 27 B and C for the NTD and PAK ensembles, respectively. Here, the LBD is oriented with the coregulator binding surface in the foreground with the overall structure shown in ribbon (NTD in blue, PAK in orange). In the 180° rotated view, the ribbon is hidden for clarity. The surfaces are colored with the same heat map for NTD and PAK. The differential activation of the coregulator binding surface is evidenced in the relative sharpening and broadening at the periphery. The PAK bound ensemble is more strongly sharpened along helix 4, while the NTD bound ensemble is sharper along the helix 3 side of the coregulator binding surface. At the AF2 side of the hormone binding pocket, NTD sharpens the ensemble distributions for VAL889 and MET895 to a greater extent than PAK. In the 180° rotated view, the hormone binding pocket surface is in the foreground. While the summed order parameters for the majority of residues in the hormone binding pocket (17 / 23) are unaffected by coregulator binding, ARG752 is broadened in both complexes ($\partial s_{NTD} \sim 0.2$, $\partial s_{PAK} \sim 0.6$). And the distribution of MET787 appears strongly sensitive to coregulator identity as NTD does not perturb the order parameter at this position but PAK very strongly broadens with a ∂s of ~ 0.9. At the principal functional surfaces of the AR LBD, there are coregulator specific perturbations to the ensemble.

Identification of Distant Changes to the AR LBD

Residue specific differences in order parameter between the PAK and NTD structure ($\partial s = s_{PAK} - s_{ntd}$) were calculated. The mean of this distribution is -0.04 with a standard deviation of 0.29 indicating little difference in the overall breadth / sharpness of the dihedral

distributions for these ensemble structures. Figure 30 A, maps the regions of significant differences in order parameter between NTD and PAK. Residues with sharper distributions in the NTD bound complex are shown as blue spheres, while those exhibiting sharper distributions in the PAK bound complex are colored orange. In this structure wide comparison of order parameters, the coregulator specific perturbations to the APO ensemble are readily identifiable at both the coregulator binding surface and the hormone binding pocket.

While other significant differences in order parameter are distributed across the structure of the AR LBD, clusters of coregulator specific ensemble sharpening are apparent. Interestingly, clusters of sharpening for one coregulator bound ensemble are often found adjacent to sharpened clusters for the other coregulator bound ensemble. This spatial relationship suggests that for the sharpening of specific regions there is often an accompanying broadening at flanking structural regions. For example, there is a region of NTD sharpening that originates at the N-terminal side of helix 12 and extends across helices 3 and 1 to ARG752 at

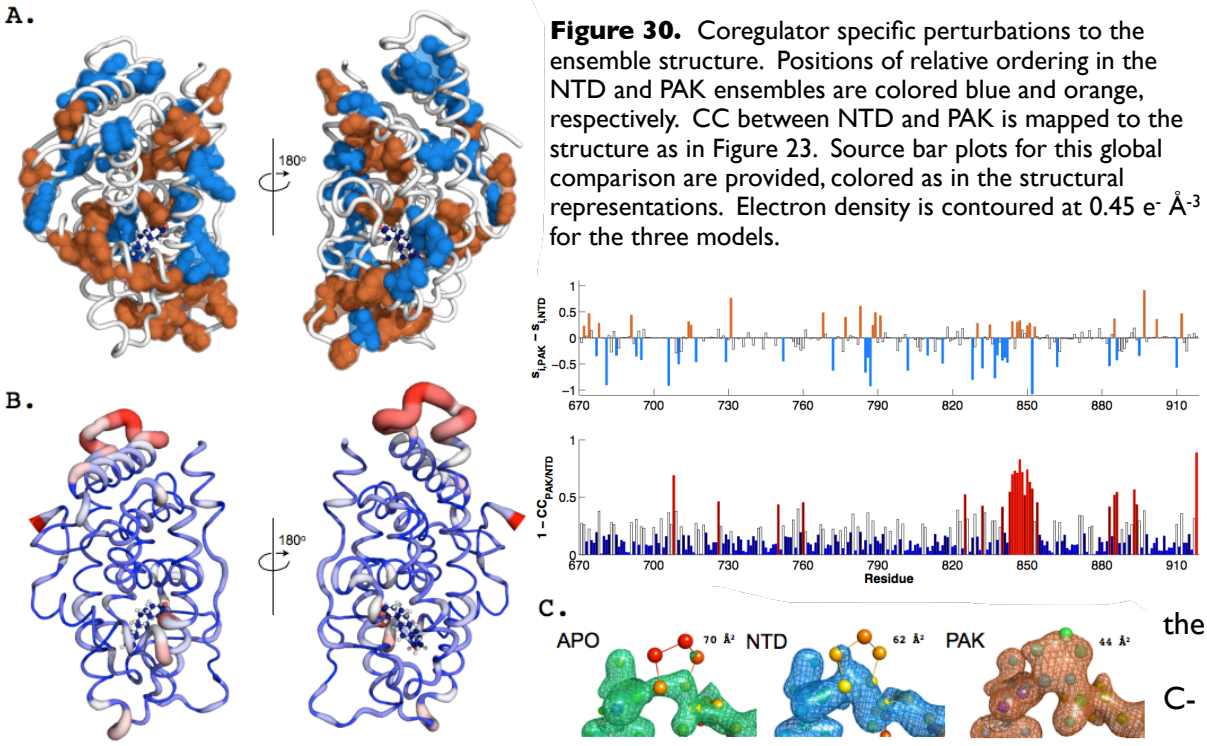


Figure 30. Coregulator specific perturbations to the ensemble structure. Positions of relative ordering in the NTD and PAK ensembles are colored blue and orange, respectively. CC between NTD and PAK is mapped to the structure as in Figure 23. Source bar plots for this global comparison are provided, colored as in the structural representations. Electron density is contoured at 0.45 e^- $Å^{-3}$ for the three models.

the C-

terminal side of helix 5; on all four sides of this narrow band, there are positions of significant sharpening for PAK. This result attests to fluidity in hydrophobic packing. If the energetics of a specific region become well defined, neighboring regions broaden in compensation. One of the more conspicuous instances of this broadening / sharpening tradeoff occurs at the N-terminal end of helix 8, the C-terminal end of helix 9, and the N-terminal end of helix 10. This large cluster of sharpened distributions in the NTD bound ensemble is oddly discontinuous across the loop bridging helices 9 and 10. In fact, this sharpening at the base of the loop seems to promote broadening, as the PAK bound ensemble is significantly sharper in the loop.

The position specific change in order parameter informs the relative definition of local structural energetics. This metric elucidates the regions of the structure that respond to coregulator binding with either a sharpening or broadening in the energy landscape, irrespective of coordinate. However, the change in order parameter does not capture a redefinition of the minimum in a given energy surface. For instance, the change in order parameter metric will not report on the extreme case of two out of phase dihedral distributions with order parameters of unity. For this reason, pairwise correlation coefficients of dihedral distributions were also considered in comparison of the NTD and PAK bound ensembles. Figure 30 B maps the per residue correlation coefficient to the putty cartoon structure of AR LBD. Positions of high correlation are colored blue while positions of low correlation are colored red. At initial inspection, there is considerable overlap between the set of residues identified with the order parameter comparison and the set identified by correlation coefficient. However, in many regions that cluster by change in order parameter, the cluster is extended by flanking regions of low correlation. The loop bridging helices 9 and 10 again makes itself conspicuous with low correlation across all positions in the loop. Considering the cluster at the base of the loop,

positions of low correlation extend both N-terminally along helix 9 to the juncture with helix 4 at the top of the coregulator binding surface and C-terminally down the back of the molecule along helix 10.

Loop 9 is Ordered in the PAK Bound LBD

L9 is a region of the AR LBD structure that is historically difficult to model. Across deposited cryo datasets, models of L9 are often unbuilt or sufficiently high in B-factor (> 100 $Å^2$) to render their position moot. The ability to visualize this density was an unexpected boon of collecting diffraction at ambient temperature. The additional 3 Å in the **a** axis of the unit cell creates enough of a buffer such that neighboring symmetry related molecules do not encroach on the density of the loop. In fact, within 5 Å of L9 in the high resolution cryo dataset, there are 60 protein atoms from neighboring molecules. In contrast, only 19 protein atoms from neighboring molecules are within 5 Å of L9 in the ambient datasets. While the proximity of neighboring molecules can make model building difficult, the extra room also provides a more native vision of this region of AR LBD structure.

In Figure 30 C, the electron density ($2F_o$ - F_c) for the C-terminal portion of L9 is shown for the APO, NTD-bound, and PAK-bound LBD structures. In order to compare maps across the structures, electron density is equivalently contoured at 0.45 e $Å^{-3}$ for the three maps. In the APO and NTD-bound structures, the majority of the backbone is trace-able through density, however, many sidechains lack density and model building is challenging. Of particular interest is the complete absence of sidechain density for PRO849, which could be a product of either structural heterogeneity or true disorder. In both the APO and NTD-bound structures, this is clearly reflected in the high B-factor of the γ-carbon. In contrast, this region of L9 is very well defined in the PAK-bound structure with complete density for PRO849. After correcting

for potential resolution differences in B-factor, the change in B-factor relative to APO for PAK bound is ~23 $Å^2$, and for NTD is ~12 $Å^2$. This conventional refinement approach to compare structural changes in a region that is highlighted by metrics of ensemble comparison reveals a significant sharpening of the distribution in the PAK bound complex.

L9 Composition

Structurally, L9 appears sensitive to the identity of bound coregulator. It is remarkable that such a distant and disordered region of the AR LBD is coupled to the coregulator binding surface and a single well defined conformer is observable in the PAK bound state. Returning to the analogy of the coregulator binding surface as the active site for AR LBD, the molecule has seemingly evolved a mechanism for informing structure of coregulator identity / composition. If we consider proline isomerization at this loop, certain coregulators may have the potential to create a long lived induced fit species by freezing out either cis or trans conformers of PRO849. That this proposed mechanism for induced fit / molecular memory is over 20 Å distant from the active site is truly amazing. From an evolutionary perspective, it is not difficult to see the advantages of controlling an active site with a distant and enduring switch. For one, the active site is left unchanged and accessible for future interactions with distant energetic steering that can bestow fidelity on the promiscuous coregulator binding surface.

But the proposition of proline isomerization is far from the only mechanism at work in this region (Figure 32 A). The sequence of L9 is conserved across all vertebrate ARs and is a hotspot for post-translational modifications. Beginning with the threonine adjacent to PRO849, PIM-1 kinase has been shown to phosphorylate this position[26]. The phosphorylation of THR850 precedes the ubiquitination of both LYS845 and LYS847, which potentiate AR transcriptional activity[27]. From the perspective of AR as a substrate, the presentation of this loop, whether

disordered or fixed in conformation, will likely have downstream effects on the rates of post translational modification, and ultimately the transcriptional activity of the receptor. Further, the regulatory role of this loop is substantiated by a known metastatic prostate cancer mutation that introduces a glycine at ARG846.[28] This introduction of flexibility would likely haywire any regulatory role and diminish coregulator specific control over the conformational sampling of L9.

However, the coupling between the conformational distribution of L9 and the coregulator binding site is weak evidence for mysterious forces acting at a distance. Even if switching at this loop was between well defined, distinct conformations, solely structural evidence would constitute unidirectional communication, which is not possible / highly unlikely in the thermodynamic basis of allostery. To be blunt, these distant structural changes, as presented, could be weakly ascribed to resolution differences or perhaps even more simply credited to crystallographic artifacts. With that said, across 27 ambient temperature datasets of the APO AR LBD and eight ambient temperature datasets of the NTD bound AR LBD we observed no instances of L9 ordering akin to the 6 ambient temperature datasets of the PAK bound AR LBD.

As alluded to earlier, allostery is not a unidirectional phenomenon. The coupling of these sites is directly testable by mutation to the loop. Whether by monitoring the effects of mutation on the structure of the AR LBD or by alteration of coregulator binding energetics, the burden of proof necessitates demonstration of communication from L9 to the coregulator binding surface.

Let us return to our representative coregulator fragments. The AR LBD recognizes NTD and PAK with similar energies of interaction. However, both association and dissociation

of PAK proceed at a much slower rate. These slowed kinetics are indicative of greater structural rearrangements and higher kinetic barriers of activation. From the structural analyses, we learn that NTD and PAK both perturb the structural ensemble of AR LBD at distant regions throughout the structure. With that said, the most striking difference across the structures is the profound ordering of L9 in the PAK bound ensemble. In contrast, the relative ordering of L9 is indistinguishable between maps of the APO or the NTD-bound AR LBD. Given this information, we would expect little change to NTD binding energetics upon mutation to L9. On the other hand, PAK binding energetics would be expected to be sensitive to mutation at L9.

At the extremes of the structural hypothesis, there are two relevant mutations. A constitutively trans mimetic at the position of PRO849 would likely predispose the AR LBD to interaction with PAK. Preclusion of loop ordering by introducing disorder would negatively impact the energetics of interaction with PAK. For either mutation, little to no effect would be expected for the energetics of interaction with the AR LBD.

In order to probe the role of backbone flexibility at PRO849, a valine was introduced at this position (P849V). To eradicate the potential for ordering, the extreme case of an all GLY-SER L9 (GS-L9) was created. These mutations were introduced into the N-terminal avi-tag construct and purified under the same conditions as the WT protein.

For accurate comparison of binding energetics, SPR measurements were made concurrently between surfaces of WT and mutant AR LBDs captured at comparable densities on different flow cells of the same chip. In testing for NTD and PAK, serial dilutions of each were alternated and collected in quadruplicate. To be clear, steady state responses for the WT and mutant AR LBDs were recorded based upon the measured response to injection of the

same peptide sample. In control experiments establishing the precision of the SPR measurement of coregulator affinity for the WT protein, there is appreciable flowcell to flowcell variation. For instance, in duplicate isotherms for PAK binding to WT AR LBD immobilized at similar densities on the three flowcells of the SPR chip (sextuplicate measure), the largest discrepancy in fit dissociation constant for the individual isotherms is 18 nM. Steady state responses for the WT and mutant AR LBDs were recorded based upon the measured response to injection of the same peptide sample. To establish the energetic effects of mutation to L9 on the binding of coregulators, dose response experiments were first considered individually using the same concentration samples for the interaction with the WT and the mutant AR LBD surfaces (n = 4). Each individual recording of NTD or PAK isotherms yields a measured affinity for the WT and the mutant receptor. Energetic differences in molecular recognition are reported with respect to WT.

In Figure 32 B, merged isotherms for NTD interaction with WT, P849V, and GS-L9 AR LBD are plotted. As in the comparison of NTD and PAK binding against the WT AR LBD, fits to steady state responses were obtained by expectancy maximization of the adsorption energy distribution. In the right hand panel, the NTD adsorption energy distributions for the P849V and GS-L9 are plotted relative to the WT AR LBD. Dissociation constants for the interactions, taken from the local maximum in the energy distribution, range from 1.10 μM to 1.14 μM. This 40 nM

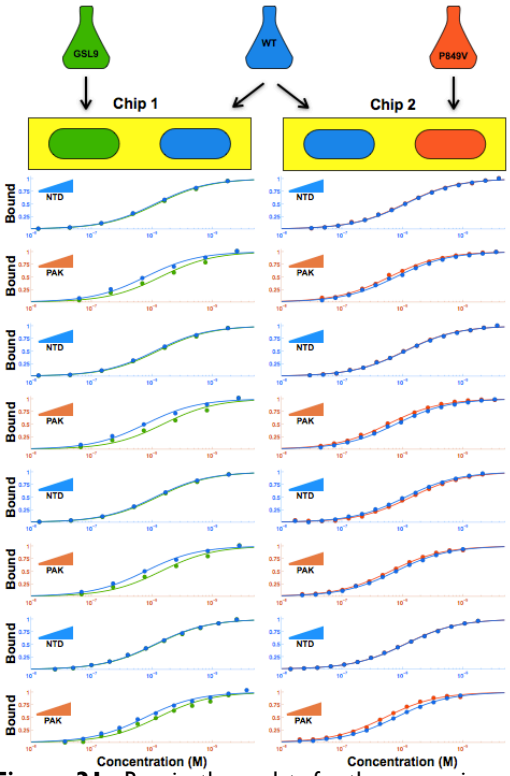


Figure 31. Raw isotherm data for the comparison of L9 mutants in the recognition properties of NTD and PAK.From top to bottom, across the two chips, this is the order in which the isotherms were collected.

difference in dissociation constant is statistically insignificant and within experimental error, proving that the energetics for AR LBD recognition of NTD are unaffected by mutation to L9.

Figure 32 C shows the isotherms for PAK interaction with the same SPR surfaces that were used in the measurement of NTD binding energetics. In contrast to NTD binding, both mutations shift the equilibrium response along the concentration axis. As in Figure 32 B,

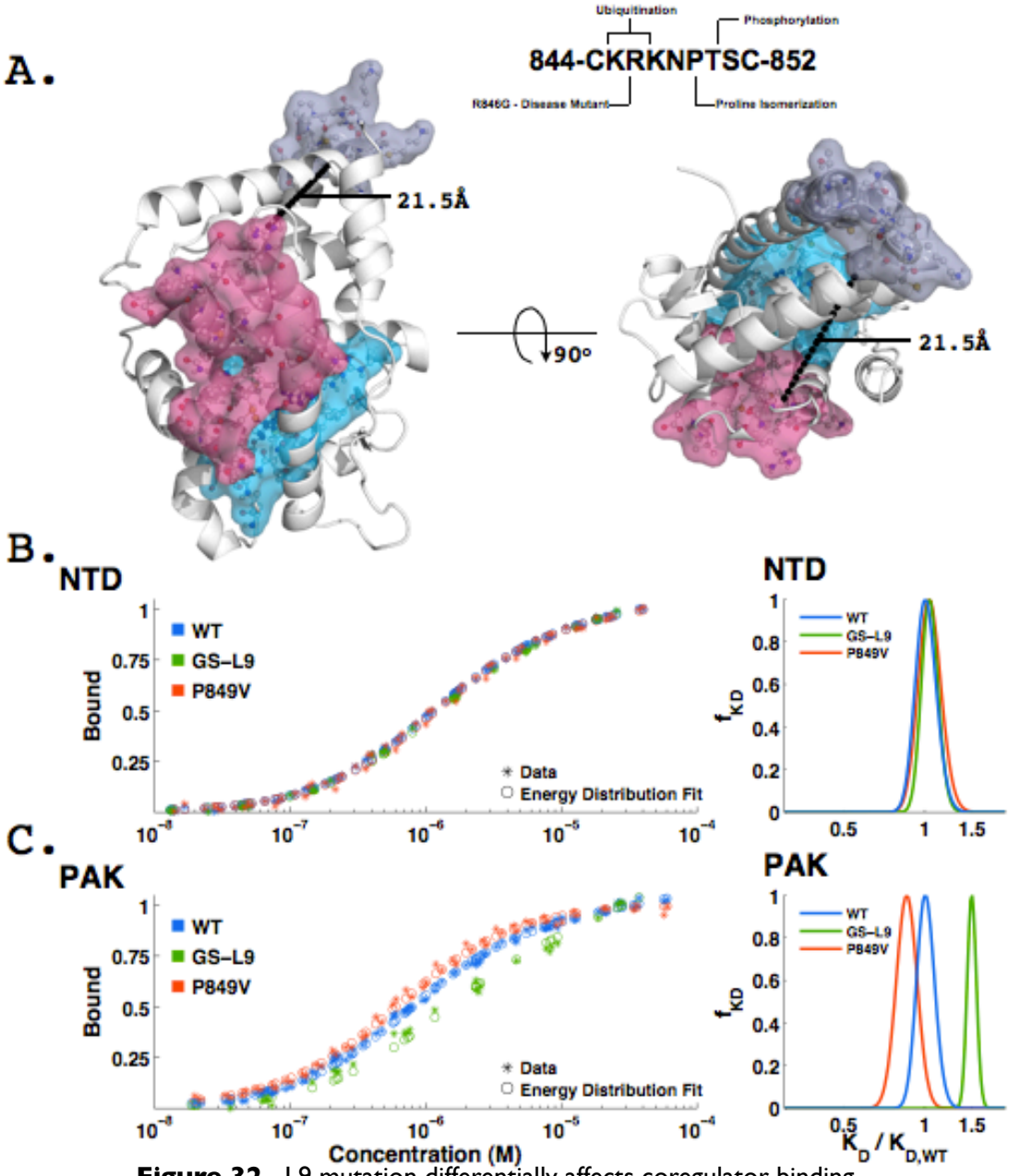


Figure 32. L9 mutation differentially affects coregulator binding.

associated adsorption energy distributions are plotted for the L9 mutants relative to WT in the right hand panel. The dissociation constants for these interactions are 790 nM, 660 nM, and 848 nM for the WT, P849V, and GS-L9 AR LBDs respectively. While this marks an incredibly subtle range in the affinity of interaction with PAK, there is statistical significance to the difference in interaction energy for the P849V and GS-L9 mutants ($z = 4.25$). Interestingly, the dissociation kinetics of PAK are unaffected by mutation to L9, suggesting that the slow kinetics of PAK interaction stem from structural rearrangements outside of L9.

These energetic differences in interaction energy for PAK binding are very slight. But we must remember the distance at which this is taking place. At a minimum distance of 21.5 Å, L9 is on the opposite face of the AR LBD, and it is remarkable that there is any energetic effect on coregulator binding. Returning to our ensemble view of AR LBD structure, introduction of order to L9 (P849V) instills a predilection for interaction with PAK. In contrast, abrogation of order to L9 (GS-L9) leads to weaker binding of PAK. The all important control for these comparisons is the binding of NTD to the same proteins. While mutation to L9 has a minor effect on the energetics of PAK binding ($\Delta\Delta G_{PAK} \sim 195$ cal mol^{-1} K^{-1}), NTD binding proceeds with indistinguishable energetics to WT ($\Delta\Delta G_{NTD} \sim 10$ cal mol^{-1} K^{-1}). It is in this ~19-fold difference in the effect of L9 mutation on coregulator binding that significance is found.

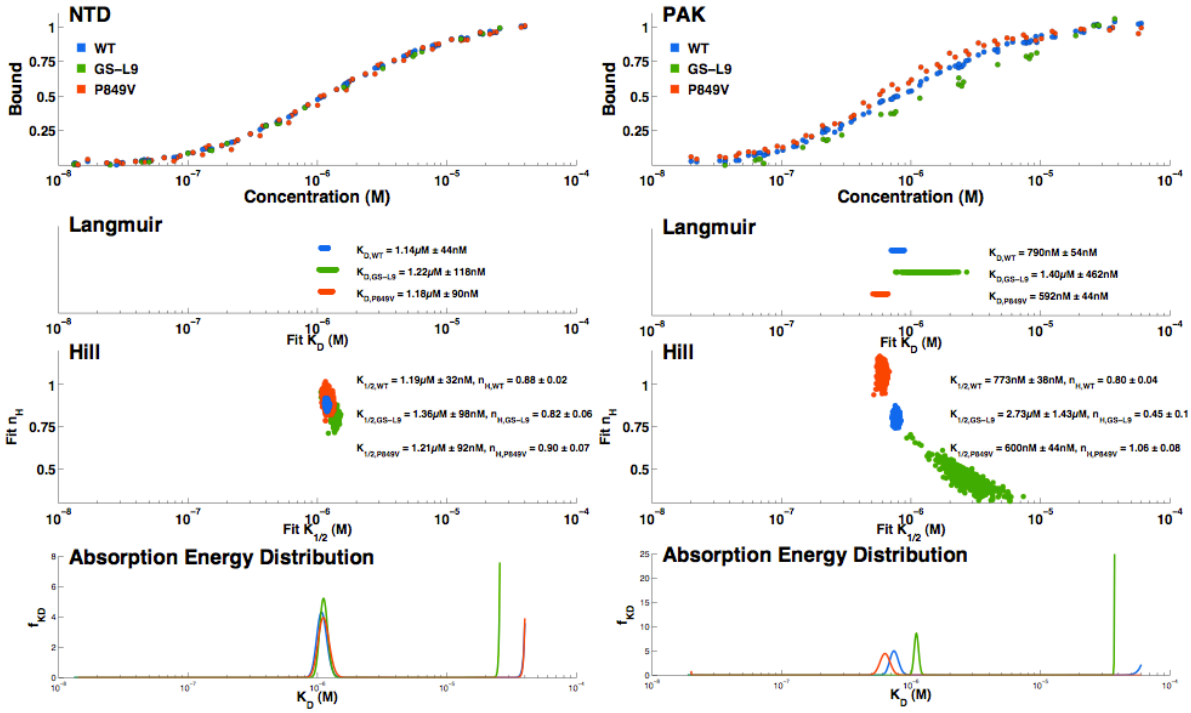


Figure 33. Alternative fitting strategies for comparison of L9 effects on coregulator recognition properties. The AED method yields the most conservative estimate.

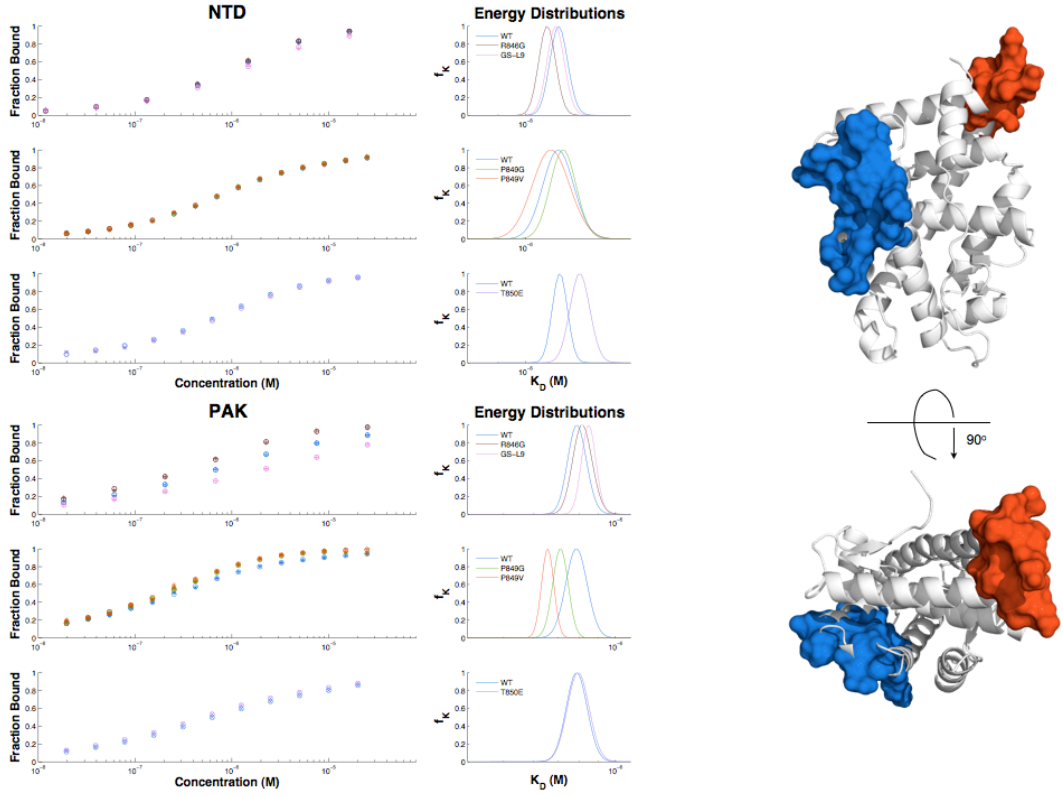


Figure 34. Effects of other mutations on binding of NTD and PAK.

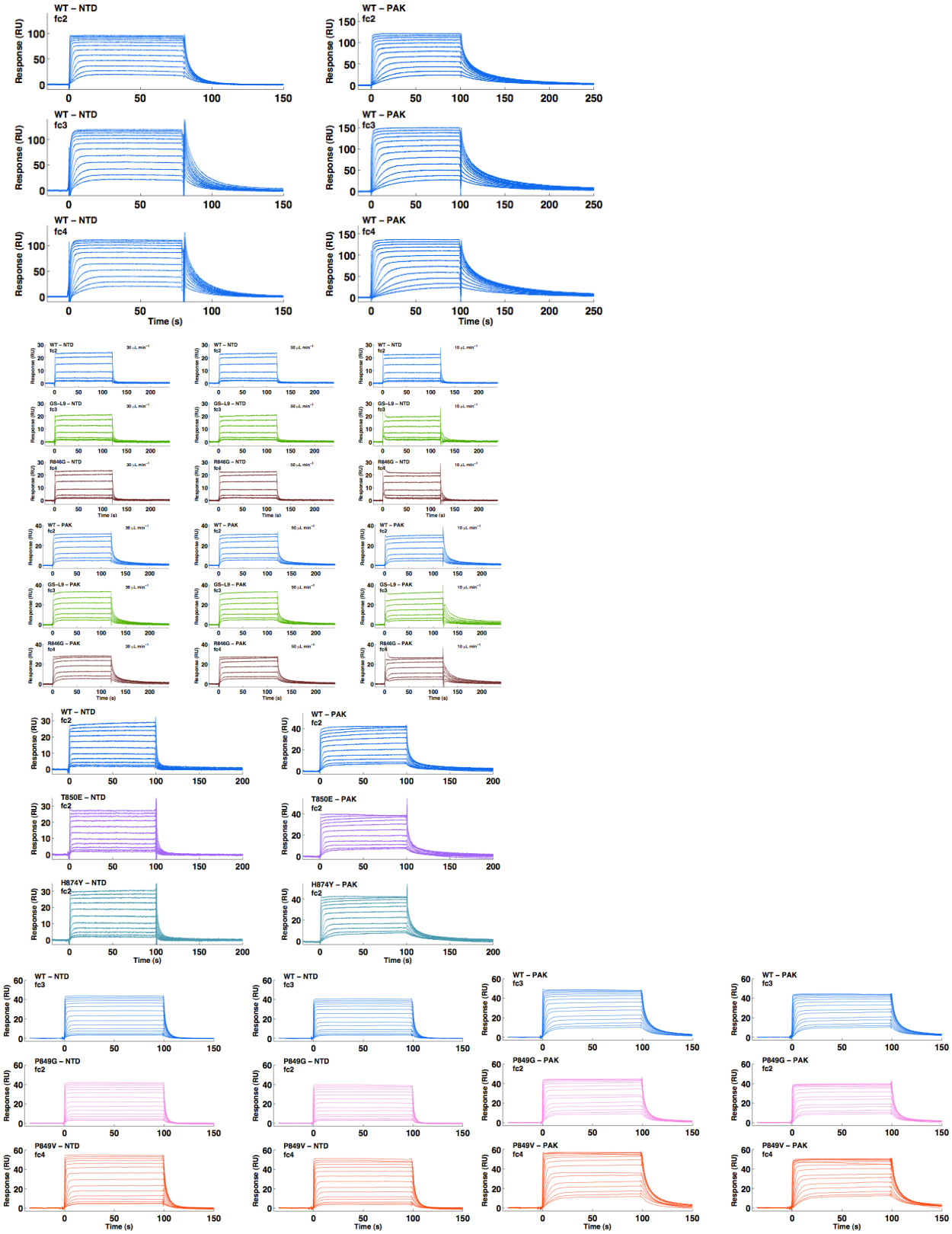


Figure 35. Representative raw SPR traces considered in this chapter.

References:

1. Ozers, M. S. *et al.* The androgen receptor T877A mutant recruits LXXLL and FXXLF peptides differently than wild-type androgen receptor in a time-resolved fluorescence resonance energy transfer assay. *Biochemistry* **46**, 683–695 (2007).
2. Hur, E. *et al.* Recognition and accommodation at the androgen receptor coactivator binding interface. *PLoS Biol.* **2**, E274 (2004).
3. Nishimura, K. *et al.* Modulation of Androgen Receptor Transactivation by Gelsolin A Newly Identified Androgen Receptor Coregulator. *Cancer Research* **63**, 4888–4894 (2003).
4. Sun, H. Q. Gelsolin, a Multifunctional Actin Regulatory Protein. *Journal of Biological Chemistry* **274**, 33179–33182 (1999).
5. van de Wijngaart, D., Dubbink, H. & Molier, M. Inhibition of androgen receptor functions by gelsolin FxxFF peptide delivered by transfection, cell-penetrating peptides, and lentiviral infection - van de Wijngaart - 2010 - The Prostate - Wiley Online Library. *The …* (2011).
6. Schaufele, F. *et al.* The structural basis of androgen receptor activation: intramolecular and intermolecular amino–carboxy interactions. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9802–9807 (2005).
7. De Angelis, R. W., Yang, Q., Miura, M. T. & Bain, D. L. Dissection of Androgen Receptor–Promoter Interactions: Steroid Receptors Partition Their Interaction Energetics in Parallel with Their Phylogenetic Divergence. *J Mol Biol* 1–13 (2013).doi:10.1016/j.jmb.2013.07.033
8. Lee, S. R. *et al.* AR and ER interaction with a p21-activated kinase (PAK6). *Molecular Endocrinology* **16**, 85–99 (2002).
9. Gartel, A. L. & Radhakrishnan, S. K. Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer Research* **65**, 3980–3985 (2005).
10. Liu, T. *et al.* p21-activated Kinase 6 (PAK6) Inhibits Prostate Cancer Growth via Phosphorylation of Androgen Receptor and Tumorigenic E3 Ligase Murine Double Minute-2 (Mdm2). *Journal of Biological Chemistry* **288**, 3359–3369 (2013).
11. Belakavadi, M. & Fondell, J. Role of the mediator complex in nuclear hormone receptor signaling. (2006).
12. Sandblad, P., Arnell, R., Samuelsson, J. & Fornstedt, T. Approach for reliable evaluation of drug proteins interactions using surface plasmon resonance technology. *Anal. Chem.* **81**, 3551–3559 (2009).
13. Zhou, X. E. *et al.* Identification of SRC3/AIB1 as a Preferred Coactivator for Hormone-activated Androgen Receptor. *Journal of Biological Chemistry* **285**, 9161–9171 (2010).
14. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
15. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16247–16252 (2011).
16. Lang, P. T. *et al.* Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Science* **19**, 1420–1431 (2010).

17. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**, 379–IN4 (1971).
18. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242 (2011).
19. Estebanez-Perpina, E. The Molecular Mechanisms of Coactivator Utilization in Ligand-dependent Transactivation by the Androgen Receptor. *Journal of Biological Chemistry* **280**, 8060–8068 (2004).
20. Darimont, B. D. *et al.* Structure and specificity of nuclear receptor-coactivator interactions. *Genes & Development* **12**, 3343–3356 (1998).
21. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst (2010). D66, 213-221 [doi:10.1107/S0907444909052925]* 1–9 (2010).doi:10.1107/S0907444909052925
22. Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *Elife* **1**, e00311 (2012).
23. McClendon, C. L., Friedland, G., Mobley, D. L., Amirkhani, H. & Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **5**, 2486–2502 (2009).
24. Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751 (1992).
25. Askew, E. B., Gampe, R. T., Stanley, T. B., Faggart, J. L. & Wilson, E. M. Modulation of Androgen Receptor Activation Function 2 by Testosterone and Dihydrotestosterone. *Journal of Biological Chemistry* **282**, 25801–25816 (2007).
26. Linn, D. E. *et al.* Differential Regulation of Androgen Receptor by PIM-1 Kinases via Phosphorylation-dependent Recruitment of Distinct Ubiquitin E3 Ligases. *Journal of Biological Chemistry* **287**, 22959–22968 (2012).
27. Dirac, A. M. G. & Bernards, R. The Deubiquitinating Enzyme USP26 is a Regulator of Androgen Receptor Signaling. *Molecular Cancer Research* 1–11 (2010).doi: 10.1158/1541-7786.MCR-09-0424
28. Marcelli, M. *et al.* Androgen receptor mutations in prostate cancer. *Cancer Research* **60**, 944–949 (2000).

Conclusion:

Allosteric controls of molecular recognition

Allostery is a consistent theme throughout this manuscript. And, while Appendices 4 and 5 do not touch upon promiscuity, they are excellent examples of switch-like allosteric controls regulating biological function. For the case of the promiscuous nuclear receptor, simple switches would not allow function in the complex milieu of potential interactants.

Nuclear receptors are very powerful signaling molecules. From translation to proteosomal degradation, they are in a state of continual interaction with small molecules, other proteins, and DNA sites. With literally hundreds of potential interaction partners, the nuclear receptor walks a fine line between normal function and disease states. Prior to this work, my general view of this process was a strict concentration dependent competition between potential interactants.

We are taught that biology works on a logarithmic scale, and subtle changes of less than ~5-fold should be discounted. Much of the work in this manuscript describes small differences in affinity or kinetics, and many of my peers have questioned the importance of such subtle differences. These next two sections take different approaches to highlight an appreciation of subtle changes to the energetics of molecular recognition. And while the premises of the arguments can be construed as contradictory, they are independent thought experiments and should be treated as such.

On subtle changes in interaction energy:

For a given signal source, (ie- circulated steroids) a concentration gradient is established. At steady state, the concentration of unbound signal molecules approaches a linear gradient. From an affinity / $K_{1/2}$ perspective, we can define a point along that gradient where the probability of receptor binding is 50%. Depending on how steep the gradient is, subtle changes to affinity could have profound effects on the location of this event horizon at which there is a

50% probability of binding. Consider a 10nM source of steroid entering a cell, and place a sink at the center of the nucleus. With time, a linear gradient is established across the radial distance (r) of the cell. At a distance of r/2, the concentration is 5nM, and were that steroid testosterone, all androgen receptors outside that distance would have a 50% chance of becoming activated. Now if that steroid were dihydrotestosterone, the event horizon is located at r/10, and a much larger population of the androgen receptors within the cell would have the 50% activation potential. While it is a more difficult thought experiment to envision concentration gradients of intracellular proteins and genomic DNA sites, this event horizon for binding argument extends to the numerous coregulators and response elements. I argue that the relevance of a change in interaction energy will largely depend on the steepness of a given signal's established concentration gradient.

On the kinetics of competition:

At a concentration of 1mM, the mean intermolecular distance is 118.43Å. For a single receptor, that means that whenever a ligand approaches within 100Å, the receptor is effectively bathed in a 1mM solution of that ligand. This realization does not break the competition model, rather it highlights how, from the perspective of a single molecule, the half-life of different potential complexes might be the most relevant physical parameter. Consider coregulator binding, if a receptor finds itself with two potential ligands within this 100Å interaction distance, the effective concentration of both species is so high above the μM $K_{1/2}$'s that the selection of interactant would truly be stochastic. So, the receptor basically alternates between binding ligand A and ligand B. But if these two ligands dissociate at different rates, say $t_{1/2,A} = 1.5 \times t_{1/2,B}$, the time dependent occupancy of the receptor shifts towards ligand A. Now let's open the system and allow the molecules to diffuse away. Every time the receptor interacts with ligand

A, ligand B gets 50% more time to diffuse away, and the local environment can swing towards primarily ligand A. Further, it is the concentration independence of dissociation kinetics that allow extrapolation to a cellular context where comprehension of absolute concentrations is impossible.

Structural consequences of function:

In both DNA recognition and coregulator binding we have learned that nuclear receptors discriminate interactions via structural changes at distant sites. These structural changes modulate fundamental biophysical aspects of the interactions, specifically, the cooperativity and kinetics of DNA recognition, and the presentation of functional sites in coregulator binding, and potentially the steering of future interactions.

It is difficult to imagine the ancestral nuclear receptor (NR_o). But for simplicity, let's assume it interacted with one ligand, one DNA sequence, and one master transcriptional activator. As size and complexity of the genome increased, the first potential off-target interactions would be the appearance of DNA sequences similar to the original. To avoid binding to the wrong DNA sequences, DNA mediated dimerization becomes advantageous. As dimerization begins its evolution, structural coupling of this surface to DNA sequence enhances the NRs discrimination and the potential to regulate diverse genomic response elements presents itself. Now, with numerous sites that can be regulated, the absolute number of the sole transcriptional activator becomes limiting, and gene duplication of this molecule is an advantage. As mutations are introduced in the duplicated activators, the coregulator binding surface is reshaped by compensatory mutation and protein-protein promiscuity is introduced.

The constant check on this evolution is the requirement that signaling remains controlled and specific. So, the receptor must remain cognizant of the environment, and further

layers of regulation are added in the form of post-translational modifications. I argue that this function is so exquisitely controlled that, in interactions with hundreds of different partners, it would be impossible to structurally encode binary switches that could differentiate all possible interactants. Instead, we find the natural solution of multiple structural spectra. Different regions of the receptor exist in multiple states, defined by both amplitude and phase.

Structure can and does control function. In this work, we learn how different, but seemingly equivalent, binding functions can lead to temporal structural remodeling. And the cycle repeats...

Appendix 1:

Ambient temperature datasets of the androgen receptor ligand binding domain

Between April, 2011 and May, 2013 I collected 178 ambient temperature datasets of the androgen receptor ligand binding domain under various ligand contexts and mutations. The construct used in crystallization was (SG)-PRO671-GLN919. For production of the different hormone bound states, AR was expressed in 2X LB with 25μM supplementation of either dihydrotestosterone (DHT), testosterone(TES), or metribolone (R1881). To ensure, homogeneity, following TEV-cleavage of the N-terminal 6xHIS tag, proteins were further purified by cation exchange using a 0 - 0.6M Li_2SO_4 gradient, eluting at ~0.1 M Li_2SO_4. Protein was diluted to ~25μM before being snap-frozen in 10% glycerol, and stored at -80C. For crystallization, individual tubes were thawed and concentrated ~5x before setting drops. Unless otherwise specified, coregulator fragments were co-crystallized by setting drops with peptide at a 10:1 molar excess of AR LBD.

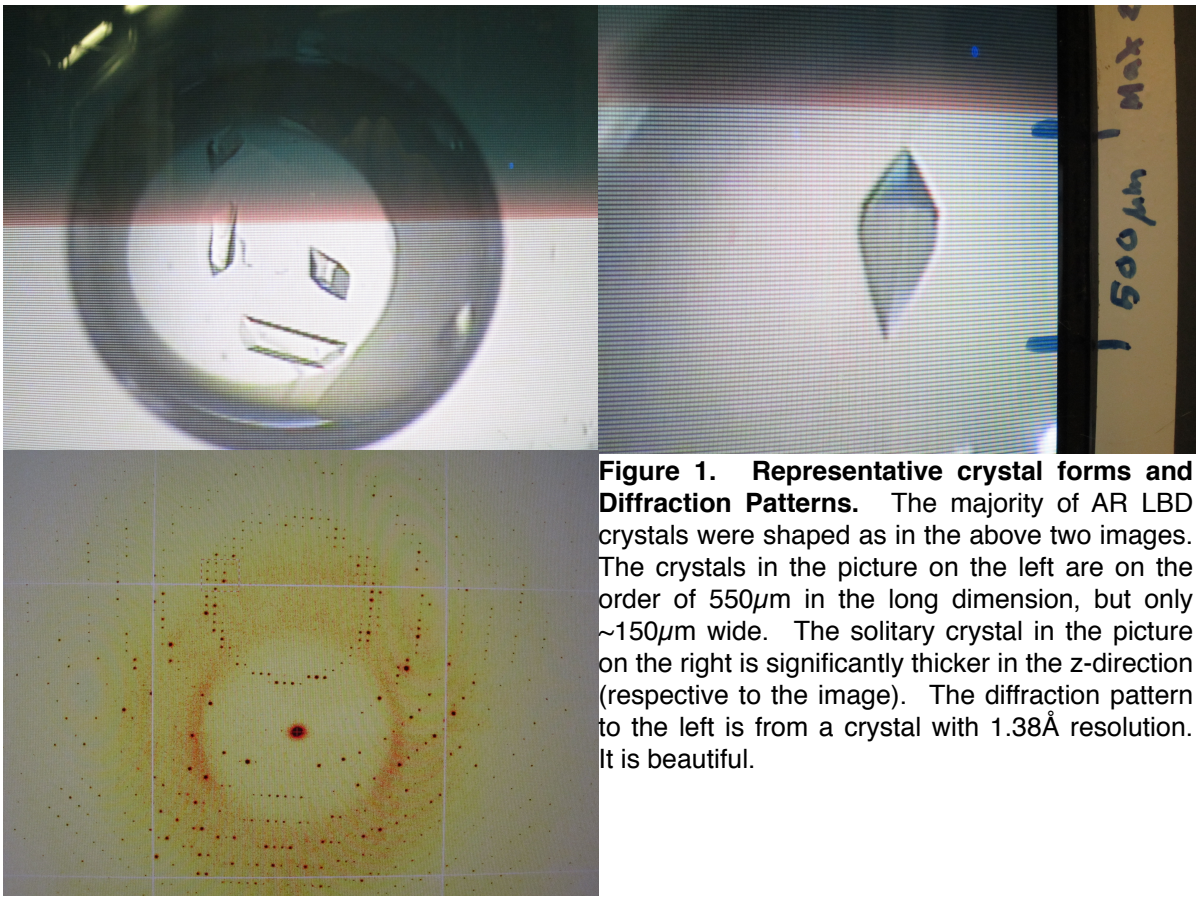


Figure 1. Representative crystal forms and Diffraction Patterns. The majority of AR LBD crystals were shaped as in the above two images. The crystals in the picture on the left are on the order of 550μm in the long dimension, but only ~150μm wide. The solitary crystal in the picture on the right is significantly thicker in the z-direction (respective to the image). The diffraction pattern to the left is from a crystal with 1.38Å resolution. It is beautiful.

All crystals were grown under a custom lithium sulfate / pH grid. Creating this 96-well block was arguably the most intelligent thing I did in graduate school. The Li_2SO_4 concentration range of this grid is 0.2M to 1.8M, and the pH ranges from 7 to 10. Prior to large scale crystal growth, a mosquito screen of this 96-point grid was used to identify lead conditions. For ease of handling, the majority of drops were grown in microbatch plates under Al's Oil (1:1 paraffin:silicon). Protein concentration was optimized towards large crystals.

Crystallization plates were driven to the advanced light source, and all data was collected at beamline 8.3.1. On a typical overnight (12am - 8am) shift, we could loop and collect data on ~25 crystals. The ambient temperature crystallography was surprisingly fast, and, as all crystals were P_{212121}, data collection proceeded without any test shots using 0.5 second exposure with 1° spacing over 104 frames. Xray damage was minimized by using the aluminum foil attenuator, and flux was decreased by setting the energy of the beam to 13,000 keV.

Coregulator	Sequence
ara70	TSEKFKLLFQSY
cbp	DGTDFGSLFDLE
dax	QGSILYSLLTSS
gripa	KENALLRYLLDKDD
gsn	ETPLFKQFFKNW
MtoL	KRRLFRSLFLST
n3	ASSSWHTLFTAE
ntd	YRGAFQNLFQSV
pak	KRRLFRSMFLST
pakLs	KRRLLRSLLLST
pgc	AEEPSLLKKLLLAPANTQ
src14	AQQKSLLQQLLTE
src31	SKGHKKLLQLLTCSS
trap	QPSKLMRLLSSN

Table 1. Sequences of coregulator fragments.

Crystals were oriented such that diffraction occurred through the thickest part. Additionally, when possible, the crystal was translated to minimize radiation damage at the beam spot.

All ambient temperature datasets were processed using xia2 with the following command:

xia2 -3dai -zero_dose -spacegroup p212121

The following table presents scaling statistics for the high resolution (outershell) of the merged datasets. These datasets will be made available to the UCSF community through the Macromolecular Structure Group. Associated

log (*filename*_aimless.log) and reflection (*filename*.mtz) files of the wild-type receptor are named by "ligand"_"nop/coreg"_"%d", where the integer is from internal tracking. Mutant structures without coregulators are recorded by "A(residue number)B"_"%d". For some coregulator bound structures, the weak interactants are occasionally found at < 1.0 occupancy.

Ambient Temperature Datasets (T = 275-278 K):

Table 2. List of ambient temperature datasets.

*: the PAK peptide was soaked into this crystal by addition of 1μL of 5mM peptide to a 2μL drop ~10 minutes prior to data collection.

filename	Ligands		High Resolution Shell				P_{212121}		
	HBP	AF2	Resolution (Å)	Completeness	R_{merge}	<I> / σ(I)	a (Å)	b (Å)	c (Å)
DHT_nop_111	DHT	apo	1.46	99.5	0.660	2.2	58.46	66.97	68.38
DHT_nop_22	DHT	apo	1.47	99.7	0.551	2.5	58.60	67.01	68.43
DHT_nop_112	DHT	apo	1.54	95.5	0.466	2.0	58.51	67.01	68.40
C844S_90	DHT	apo	1.58	99.8	0.691	2.4	58.29	67.34	73.49
C844S_92	DHT	apo	1.58	99.8	0.594	2.4	57.94	67.06	73.38
DHT_nop_114	DHT	apo	1.61	99.8	0.686	2.1	58.05	67.41	73.42
DHT_nop_113	DHT	apo	1.64	99.9	0.642	2.4	57.88	67.07	73.22
DHT_nop_48	DHT	apo	1.64	100	0.620	2.2	57.29	66.79	73.57
C844S_91	DHT	apo	1.67	99.8	0.705	2.4	58.31	67.25	73.40
C844S_93	DHT	apo	1.74	100	0.685	2.3	57.91	66.95	73.38
R846G_96	DHT	apo	1.75	99.9	0.698	1.6	58.03	66.75	73.33
T850E_169	DHT	apo	1.75	99.8	0.666	2.0	57.58	66.55	71.91
DHT_nop_67	DHT	apo	1.79	99.4	0.600	2.3	58.05	67.17	73.03
DHT_nop_83	DHT	apo	1.80	99.5	0.612	2.3	57.97	67.22	73.03
R846G_05	DHT	apo	1.81	99.6	0.791	2.1	59.86	67.76	73.49
DHT_nop_43	DHT	apo	1.83	99.9	0.724	1.7	57.53	66.58	72.38
DHT_nop_44	DHT	apo	1.84	99.9	0.634	2.1	57.54	66.52	72.2
DHT_nop_53	DHT	apo	1.86	99.9	0.628	2.2	57.35	66.64	71.58
R846G_94	DHT	apo	1.86	99.8	0.659	2.1	57.57	66.64	71.23
R846G_95	DHT	apo	1.86	99.8	0.707	2.0	58.00	66.69	73.29
T850E_178	DHT	apo	1.86	90.8	0.678	2.0	56.10	66.28	71.06
DHT_nop_45	DHT	apo	1.87	99.8	0.834	1.8	57.51	66.64	72.59
DHT_nop_84	DHT	apo	1.87	99.8	0.636	2.2	57.88	66.91	73.27
DHT_nop_89	DHT	apo	1.87	97.3	0.700	1.5	57.64	66.43	71.99
DHT_nop_13	DHT	apo	1.88	99.4	0.692	2.3	58.48	67.04	73.65
DHT_nop_42	DHT	apo	1.91	97.0	0.585	2.3	57.56	66.56	72.01
DHT_nop_55	DHT	apo	1.91	99.8	0.728	2.0	57.48	66.47	71.80
H874Y_187	DHT	apo	1.91	99.3	0.556	2.3	57.18	66.50	71.34
DHT_nop_40	DHT	apo	1.93	99.4	0.639	2.3	57.42	66.55	72.06
DHT_nop_126	DHT	apo	1.94	99.7	0.602	2.4	59.24	67.33	73.38
T850E_171	DHT	apo	1.94	96.6	0.670	2.2	57.45	66.70	69.39

filename	Ligands		High Resolution Shell				P_{212121}		
	HBP	AF2	Resolution (Å)	Completeness	R_{merge}	<I> / σ(I)	a (Å)	b (Å)	c (Å)
T850E_193	DHT	apo	1.95	99.5	0.656	2.5	57.37	66.62	71.51
DHT_nop_56	DHT	apo	1.97	99.9	0.688	2.0	57.35	66.76	71.82
DHT_nop_54	DHT	apo	1.98	99.7	0.529	2.3	57.55	66.92	72.25
DHT_nop_52	DHT	apo	1.99	99.7	0.534	2.5	57.48	66.86	71.87
T850E_174	DHT	apo	1.99	99.8	0.543	2.3	56.90	66.64	71.43
DHT_nop_41	DHT	apo	2.00	99.6	0.674	2.3	57.46	67.11	73.57
P849V_186	DHT	apo	2.00	94.5	0.584	2.8	57.41	66.63	71.21
DHT_nop_47	DHT	apo	2.04	99.6	0.597	2.5	57.48	66.81	73.28
DHT_nop_14	DHT	apo	2.05	99.9	0.621	2.5	59.08	67.16	73.55
DHT_nop_50	DHT	apo	2.07	99.9	0.54	2.6	57.12	66.67	71.65
T850E_189	DHT	apo	2.09	100	0.574	2.9	57.52	66.79	72.96
H874Y_183	DHT	apo	2.1	97.6	0.553	2.1	57.52	66.66	73.48
P849V_191	DHT	apo	2.11	88.5	0.632	2.2	57.43	66.72	71.38
R846G_98	DHT	apo	2.12	99.5	0.887	2.4	58.66	66.96	72.98
H874Y_176	DHT	apo	2.13	99.5	0.597	2.0	57.31	66.79	71.71
T850E_168	DHT	apo	2.19	94.2	0.623	2.8	57.08	67.11	70.11
H874Y_192	DHT	apo	2.19	99.6	0.598	2.0	55.77	66.73	71.18
P849V_194	DHT	apo	2.19	99.6	0.530	2.3	57.52	66.60	71.36
P849V_182	DHT	apo	2.20	88.9	0.582	2.3	57.21	66.60	71.17
V730M	DHT	apo	2.22	92.3	0.621	2.5	57.23	66.78	71.61
P849V_184	DHT	apo	2.25	100	0.591	2.4	57.35	66.53	71.40
T850E_181	DHT	apo	2.29	39.3	0.371	2.2	57.02	66.68	71.47
coPIM_170	DHT	apo	2.33	90.7	0.598	2.2	57.11	66.78	70.72
DHT_nop_46	DHT	apo	2.36	99.3	0.590	2.2	57.61	66.78	72.11
DHT_ara_34	DHT	ara70	2.29	100	0.512	2.5	57.25	66.84	72.11
DHT_ara_35	DHT	ara70	2.35	99.4	0.551	2.4	57.13	66.96	71.32
DHT_ara_33	DHT	ara70	2.86	99.8	0.765	1.8	57.36	66.89	71.79
DHT_cbp_37	DHT	cbp	2.00	99.9	0.502	2.3	57.05	66.86	71.53
DHT_cbp_36	DHT	cbp	2.20	98.4	0.609	2.0	57.19	67.11	72.06
DHT_cbp_38	DHT	cbp	2.23	98.9	0.625	2.2	57.30	67.12	72.09
DHT_dax_103	DHT	dax	1.72	96.8	0.733	2.3	58.49	67.33	68.48
DHT_dax_108	DHT	dax	1.77	99.4	0.770	1.9	58.44	67.42	68.66
DHT_dax_105	DHT	dax	1.78	99.1	0.641	2.2	58.19	67.34	73.14
DHT_dax_101	DHT	dax	1.85	99.6	0.947	2.4	58.31	67.28	68.34
DHT_grip_115	DHT	gripa	1.82	99.9	0.625	2.2	59.02	68.25	73.53

filename	Ligands		High Resolution Shell				P_{212121}		
	HBP	AF2	Resolution (Å)	Completeness	R_{merge}	<I> / σ(I)	a (Å)	b (Å)	c (Å)
DHT_grip_151	DHT	gripa	1.93	93.1	0.637	2.2	57.75	67.84	73.71
DHT_grip_107	DHT	gripa	2.06	99.8	0.503	2.7	55.21	67.49	71.29
DHT_grip_119	DHT	gripa	2.07	99.6	0.672	2.2	55.93	67.47	70.90
DHT_grip_39	DHT	gripa	2.18	99.8	0.499	2.4	55.72	67.08	70.75
DHT_gsn_110	DHT	gsn	1.81	100	0.707	2.1	57.72	67.36	73.53
DHT_gsn_64	DHT	gsn	1.89	99.5	0.640	2.4	56.47	66.14	73.37
DHT_gsn_65	DHT	gsn	1.89	99.6	0.679	2.2	55.81	66.61	73.01
DHT_gsn_152	DHT	gsn	2.02	99.3	0.561	2.2	58.17	67.91	73.53
DHT_gsn_75	DHT	gsn	2.04	99.7	0.596	2.4	57.33	67.70	69.11
DHT_gsn_66	DHT	gsn	2.08	96.2	0.918	1.8	53.32	66.38	70.68
DHT_gsn_78	DHT	gsn	2.32	99.5	0.647	2.0	57.08	67.45	69.16
DHT_n3_12	DHT	n3	1.93	99.4	0.692	2.3	59.56	67.78	73.52
DHT_ntd_63	DHT	ntd	1.53	51.3	0.616	2.3	54.01	65.88	70.41
DHT_ntd_21	DHT	ntd	1.57	99.9	0.697	3.0	56.76	66.78	71.22
DHT_ntd_132	DHT	ntd	1.63	82.2	0.577	2.1	58.50	67.03	68.36
DHT_ntd_32	DHT	ntd	1.74	100	0.681	2.2	56.83	66.88	71.48
DHT_ntd_29	DHT	ntd	1.78	99.9	0.624	2.1	56.92	66.73	71.35
DHT_ntd_26	DHT	ntd	1.80	100	0.730	1.9	56.94	66.80	71.29
DHT_ntd_30	DHT	ntd	1.85	99.8	0.680	2.1	56.88	66.71	71.37
DHT_ntd_27	DHT	ntd	1.93	99.8	0.606	2.2	56.92	66.78	71.21
DHT_ntd_31	DHT	ntd	2.04	99.9	0.629	2.5	56.78	66.79	71.25
DHT_pak_20	DHT	pak	1.38	90.6	0.599	1.9	58.22	67.33	68.13
DHT_pak_25	DHT	pak	1.38	74.3	0.532	2.1	58.22	67.31	68.15
DHT_pak_19	DHT	pak	1.39	63.4	0.568	2.2	58.15	67.33	68.08
DHT_pak_24	DHT	pak	1.39	93.1	0.583	2.0	58.21	67.34	68.21
DHT_pak_23	DHT	pak	1.41	71.2	0.621	2.1	58.12	67.33	68.06
DHT_pak_127	DHT	pak	1.48	62.6	0.545	2.1	58.30	67.52	68.17
DHT_pakLs_185	DHT	pakLs	2.27	99.8	0.609	2.1	56.74	67.13	71.14
DHT_pakLs_173	DHT	pakLs	2.58	84.6	0.893	2.6	58.14	68.07	73.50
DHT_pgc_51	DHT	pgc	2.44	51.6	0.418	1.4	56.27	67.02	71.16
DHT_src14_16	DHT	src14	1.96	99.6	0.559	2.4	57.64	66.77	71.31
DHT_src14_18	DHT	src14	2.17	99.1	0.575	2.1	57.67	66.96	71.38
DHT_src14_150	DHT	src14	2.22	99.7	0.541	2.3	58.23	66.49	71.49
DHT_src31_134	DHT	src31	2.09	99.8	0.547	2.6	56.83	66.82	71.06
DHT_src31_133	DHT	src31	2.18	98.6	0.688	2.1	56.96	66.89	71.19

filename	Ligands		High Resolution Shell				P_{212121}		
	HBP	AF2	Resolution (Å)	Completeness	R_{merge}	<I> / σ(I)	a (Å)	b (Å)	c (Å)
DHT_trap_102	DHT	trap	1.80	99.8	0.816	1.9	58.80	67.40	68.90
DHT_trap_03	DHT	trap	1.89	99.7	0.614	2.3	59.46	67.86	73.33
DHT_trap_04	DHT	trap	1.89	98.7	0.854	2.1	58.90	67.64	73.25
DHT_trap_71	DHT	trap	2.00	99.6	0.589	2.4	57.83	66.73	71.58
DHT_trap_68	DHT	trap	2.08	99.9	0.581	2.5	57.80	66.77	71.12
DHT_trap_104	DHT	trap	2.08	99.6	0.597	2.3	57.89	66.56	71.35
DHT_trap_100	DHT	trap	2.10	99.9	0.636	3.3	58.93	67.90	71.80
DHT_trap_70	DHT	trap	2.19	99.7	0.550	2.6	57.82	66.73	71.50
DHT_trap_106	DHT	trap	2.19	99.8	0.665	2.2	58.66	67.38	73.00
DHT_trap_69	DHT	trap	2.21	99.8	0.732	2.6	57.75	66.84	70.92
R18_nop_179	R1881	apo	1.84	99.3	0.671	2.2	56.61	66.55	71.12
R18_nop_60	R1881	apo	1.85	97.1	0.599	2.1	56.43	66.46	71.00
R18_nop_57	R1881	apo	1.93	99.9	0.641	0.9	57.43	67.12	72.39
R18_nop_02	R1881	apo	1.94	99.9	0.618	2.3	59.17	67.37	73.67
R18_nop_124	R1881	apo	1.95	99.7	0.618	2.2	58.52	67.06	73.77
R18_nop_188	R1881	apo	1.96	91.4	0.644	2.2	56.74	66.72	71.30
R18_nop_58	R1881	apo	1.97	99.8	0.566	2.6	56.78	66.89	70.11
R18_nop_11	R1881	apo	2.01	98.9	0.650	1.9	57.28	66.54	71.53
R18_nop_121	R1881	apo	2.04	99.8	0.556	2.4	58.88	67.24	73.57
R18_nop_122	R1881	apo	2.08	99.7	0.629	2.3	59.49	67.58	73.60
R18_nop_180	R1881	apo	2.09	94.1	0.547	2.3	56.80	66.63	71.25
R18_ntd_153	R1881	ntd	1.70	98.0	0.675	2.1	56.61	66.90	71.35
R18_ntd_01	R1881	ntd	1.75	95.5	0.746	2.1	56.65	67.02	71.75
R18_ntd_154	R1881	ntd	1.92	97.6	0.565	2.7	56.61	67.18	71.60
R18_ntd_116	R1881	ntd	2.05	96.7	0.586	2.4	56.82	66.72	71.56
R18_ntd_117	R1881	ntd	2.14	99.7	0.724	2.6	56.80	67.15	71.66
R18_ntd_118	R1881	ntd	2.15	99.6	0.561	2.9	56.88	66.92	71.63
R18_pak_155	R1881	pak	2.04	97.9	0.555	2.5	55.69	66.94	70.45
R18_pak_08	R1881	pak	2.17	99.5	0.585	2.1	56.14	67.12	71.97
R18_pak_17	R1881	pak	2.22	95.5	0.603	2.6	56.56	66.9	70.52
R18_pak_09	R1881	pak	2.25	99.1	0.692	2.0	56.15	67.12	72.05
R18_trap_158	R1881	trap	1.96	99.9	0.505	2.5	59.55	67.56	73.31
R18_trap_185	R1881	trap	2.12	99.6	0.556	3.0	56.91	67.05	71.20
R18_trap_157	R1881	trap	2.16	99.5	0.483	2.8	56.82	66.83	71.38
TES_nop_139	TES	apo	1.57	99.5	0.715	1.9	57.74	66.27	71.47

filename	Ligands		High Resolution Shell				P_{212121}		
	HBP	AF2	Resolution (Å)	Completeness	R_{merge}	<I> / σ(I)	a (Å)	b (Å)	c (Å)
TES_nop_140	TES	apo	1.70	99.0	0.679	2.4	57.65	66.39	71.45
TES_nop_149	TES	apo	1.71	84.5	1.061	1.9	57.24	66.47	71.40
TES_nop_136	TES	apo	1.72	99.5	0.651	2.2	57.20	66.20	71.05
TES_nop_159	TES	apo	1.72	95.3	0.856	1.7	57.42	66.43	71.39
TES_nop_137	TES	apo	1.74	97.2	0.784	2.3	55.02	65.86	70.45
TES_nop_128	TES	apo	1.76	99.8	0.682	2.2	57.61	66.49	71.55
TES_nop_163	TES	apo	1.77	94.2	0.838	2.3	58.37	67.43	73.23
TES_nop_72	TES	apo	1.77	100.0	0.649	1.9	58.24	67.44	73.30
TES_nop_162	TES	apo	1.80	97.3	0.859	2.1	57.92	67.20	73.25
TES_nop_87	TES	apo	1.80	99.7	0.661	2.2	58.01	67.43	73.21
TES_nop_88	TES	apo	1.80	98.9	0.657	2.3	57.61	66.35	71.50
TES_nop_161	TES	apo	1.81	99.4	0.635	2.1	57.66	66.35	71.48
TES_nop_74	TES	apo	1.83	96.2	0.567	2.1	58.16	67.34	73.19
TES_nop_85	TES	apo	1.83	99.9	0.621	2.1	58.00	67.30	73.08
TES_nop_06	TES	apo	1.87	99.8	0.637	2.1	58.61	67.13	73.36
TES_nop_15	TES	apo	1.87	94.7	0.645	2.2	58.2	67.00	73.51
TES_nop_62	TES	apo	1.87	99.9	0.681	2.2	57.21	66.38	71.27
TES_nop_148	TES	apo	1.89	99.2	0.707	2.0	55.30	66.43	70.65
TES_nop_49	TES	apo	1.89	99.7	0.724	2.1	57.49	66.77	73.47
TES_nop_73	TES	apo	1.89	99.7	0.629	2.2	57.86	66.99	72.88
TES_nop_76	TES	apo	1.90	96.9	0.696	2.3	57.43	66.69	71.57
TES_nop_82	TES	apo	1.90	99.6	0.648	2.1	57.88	67.02	72.90
TES_nop_164	TES	apo	1.92	99.4	0.642	2.3	57.66	66.74	71.53
TES_nop_138	TES	apo	1.95	98.4	0.660	2.0	55.43	66.31	70.91
TES_nop_07	TES	apo	1.96	98.9	0.766	2.3	56.98	66.67	70.87
TES_nop_86	TES	apo	1.97	99.7	0.601	2.4	58.04	67.40	73.03
TES_nop_79	TES	apo	2.02	99.3	0.828	1.9	57.89	67.15	72.83
TES_nop_123	TES	apo	2.04	100	0.524	1.9	58.07	66.67	73.43
TES_MtoL_10	TES	MtoL	2.39	98.7	0.738	2.4	58.68	67.10	73.53
TES_ntd_165	TES	ntd	1.82	98.8	0.828	2.1	57.11	67.21	73.11
TES_ntd_160	TES	ntd	1.86	91.7	0.677	2.3	56.99	67.06	71.39
TES_ntd_166	TES	ntd	1.89	97.6	0.580	2.3	56.75	66.88	72.07
TES_ntd_28	TES	ntd	1.90	99.6	0.617	2.3	57.46	66.53	71.83
TES_pak_130	TES	pak	1.51	75.5	0.344	2.2	58.23	67.48	68.19
TES_pak_141	TES	pak	1.65	89.6	0.679	2.5	57.98	67.40	68.37

filename	Ligands		High Resolution Shell				P_{212121}		
	HBP	AF2	Resolution (Å)	Completeness	R_{merge}	<I> / σ(I)	a (Å)	b (Å)	c (Å)
TES_pak_129	TES	pak	1.68	96.0	0.635	2.0	58.06	67.31	68.29
TES_pak_142*	TES	pak	1.60	99.7	0.690	2.2	57.75	66.31	71.55
TES_pgc_61	TES	pgc	2.71	40.2	0.375	1.7	55.25	67.29	71.66
TES_trap_143	TES	trap	1.89	100	0.737	2.1	57.07	66.43	70.95
TES_trap_144	TES	trap	1.94	98.7	0.568	2.2	56.98	66.64	70.87
TES_trap_145	TES	trap	2.13	99.8	0.561	2.5	57.30	66.80	71.26
TES_trap_167	TES	trap	2.56	95.8	3.020	0.1	58.99	66.93	71.64

Appendix 2:

Software for the parameterization of steady-state binding data

Steady State Parameterization:

This section presents Matlab code for the parameterization of isotherm data. These are the three models that I have found most useful.

(I). The Langmuir binding model was developed in 1916 by Irving Langmuir to describe the adsorption of molecules on a solid surface. There are three fit parameters, all of which are physically meaningful. The B_{max} is the measured signal at maximum occupancy. The K_D is the concentration at which 50% occupancy occurs. The offset is dispensable for data that extrapolates to zero as concentration goes to zero. Concentration is represented by c, and the equation is presented in the SPR convention of response as a function of concentration, R(c).

$$R(c) = B_{\max} * \frac{c}{c + K_D} + offset$$

(II). The Hill equation was formulated in 1910 by Archibald Hill to describe the sigmoidal / cooperative binding of molecular oxygen to hemoglobin. In addition to the three physically meaningful parameters of the Langmuir model, the Hill equation includes an additional fit parameter, n_H, that appears as an exponent. The so-called Hill coefficient yields a fit that is able to cross through the 50% occupancy with slope not equal to one. Note that the concentration at half occupancy in this equation is no longer the dissociation constant. This is an equation, not a model.

$$R(c) = B_{\max} * \frac{c^{n_H}}{c^{n_H} + K_{1/2}^{n_H}} + offset$$

(III). The adsorption energy distribution (AED) approach to fitting isotherm data, as implemented here, enters the fitting process assuming n-different Langmuir binding processes.

Expectancy maximization reduces this distribution, and sensitively detects concentration dependent events. The greatest advantage of this model is in the explicit treatment of non-specific binding processes, which are quite common in biomolecular interactions.

Running the code:

Both the Langmuir and Hill fitting algorithms are intended for bootstrap with replacement. For this reason, input isotherms should include many points and replicates. I developed a distaste for chi-squared based errors in parameter estimation. For this reason, I turned to the bootstrap with replacement strategy which produces many different parameter combinations. Simple statistical measures, such as mean and standard deviation, of the resultant parameter distributions inform how well determined the fit parameters are assuming the model. To run these matlab functions, there are two input arguments: 1. two column isotherm vector, [conc resp(c)]; and 2. the number of bootstrap cycles to perform, n. In practice, the number of bootstrap cycles to perform will vary depending on data quality. As a first pass, I recommend starting with 500 cycles.

For the AED solver, there are three input arguments: 1. two column isotherm vector, [conc resp(c)]; 2. number of grid points to include in the energy distribution, n_grid_points; and 3. the number of expectancy maximization iterations to perform, its. Peak shape analyses of the resultant distribution provide an error in the fit affinity as the half-width at half maximum.

Using these models, I present the results for fitting sextuplicate isotherm data for interaction of PAK with the DHT liganded androgen receptor. The data is first scaled across experiments for different maximum responses by a sum of squared difference minimization of the individual isotherms.

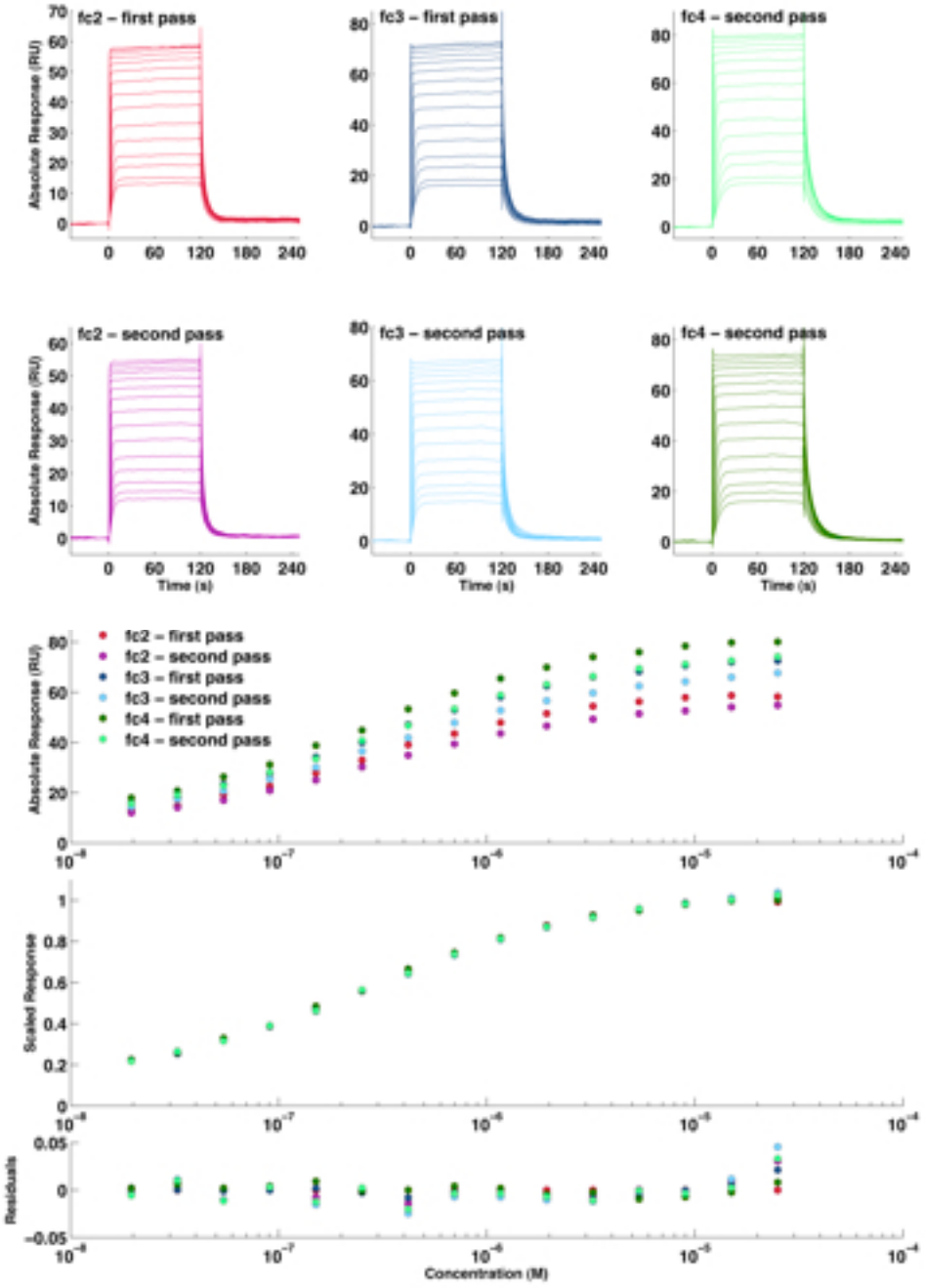


Figure 1. Raw traces and isotherm scaling.

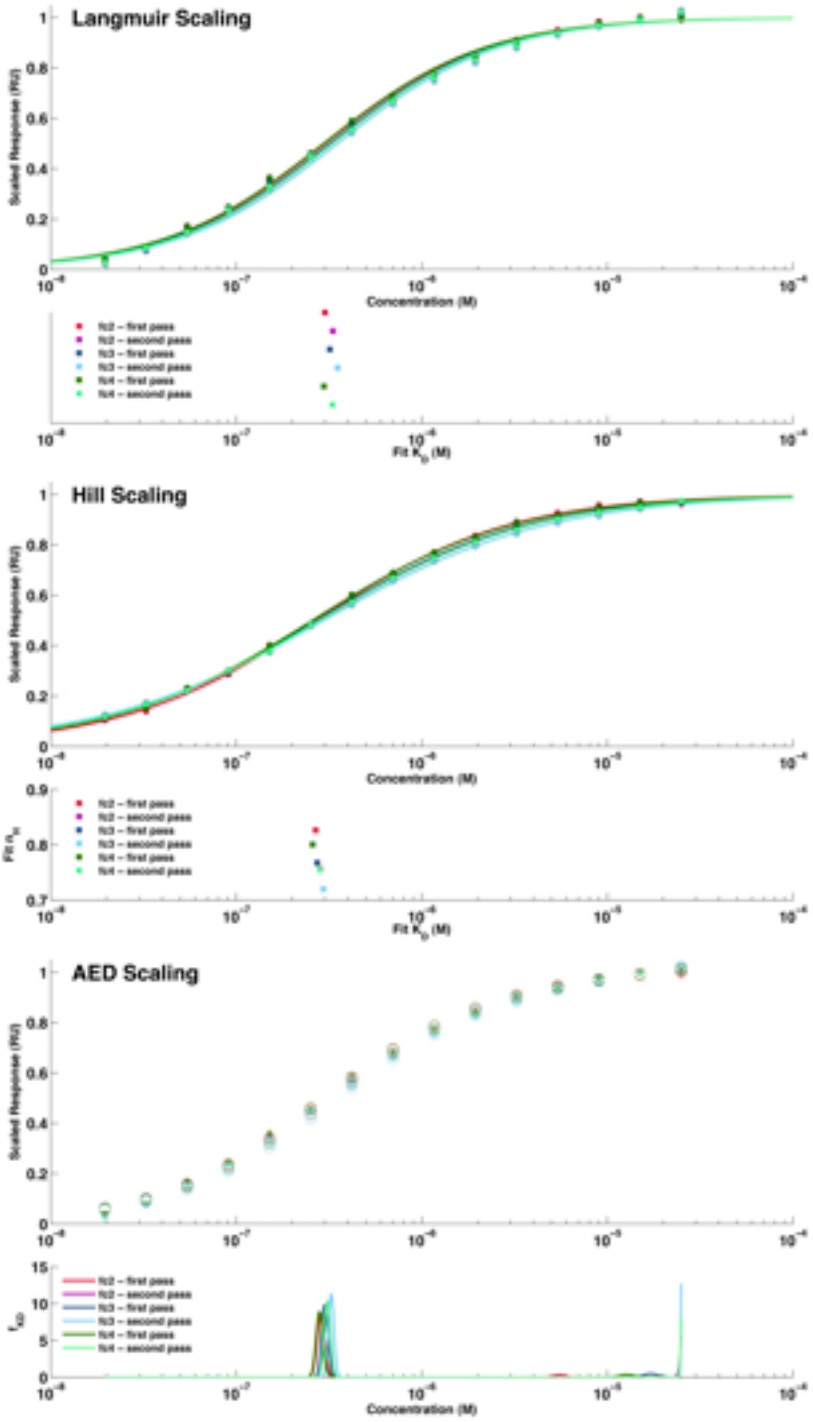


Figure 2. Application of the different models to the individual (n=6) isotherms.

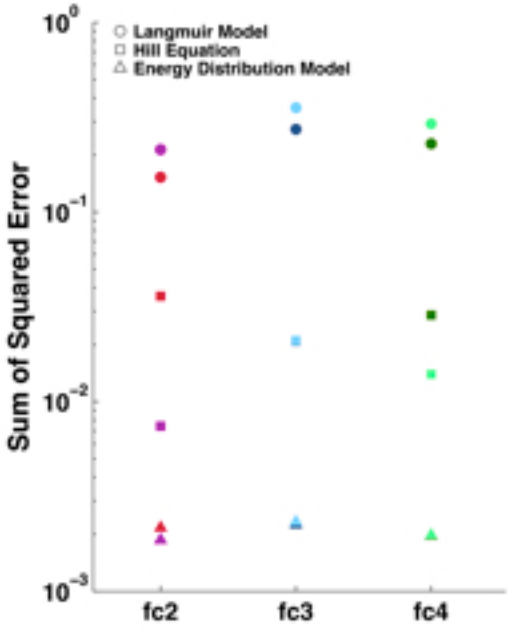


Figure 3. Error by model and flow cell.

	fc	Langmuir	Hill		AED
			$K_{1/2}$	n_H	
pass 1	2	300nM	268nM	0.83	284nM
pass 2		331nM	283nM	0.76	309nM
pass 1	3	320nM	273nM	0.77	302nM
pass 2		352nM	295nM	0.72	324nM
pass 1	4	297nM	258nM	0.80	281nM
pass 2		330nM	281nM	0.76	309nM
mean ± 2σ		322 ± 42nM	276 ± 26nM	0.77 ± 0.07	301 ± 33nM
global fit (sse)		321nM (0.034)	276nM	0.77	301 nM (0.0212)
			(0.0099)		
Bootstrap (n = 1000)		322 ± 20nM	276 ± 14nM	0.77 ± 0.03	n/a

Table 1. Parameterization and errors from individual and global fitting of the isotherms using the Langmuir model, Hill equation, and the Adsorption Energy Distribution.

```
function out = LANGstrap(isos,n)
% isos is a two column vector input
%       column 1 is concentration
%       column 2 is recorded steady state response
% n is the number of bootstrap cycles to perform

% preallocate fits matrix
fits = zeros(n,4);
pts = size(isos,1);
% bootstrap for n cycles
for k = 1:n
    which = randi([1,pts],pts,1); % random selection of points to include
    boot = f_kk_fitlang_offset(isos(which,:));
    fits(k,1) = boot.Kd;
    fits(k,2) = boot.Bmax;
    fits(k,3) = boot.offset;
    fits(k,4) = boot.sse;
    disp(strcat(sprintf('%0.2e  ',fits(k,1)),sprintf('  %0.2g  ',fits(k,2:4))))
end
% sort fits by associated errors
fits = sortrows(fits,4);
% histogram fit parameters
plotter(fits);
% return fits matrix
out = fits;
end

function p_out = f_kk_fitlang_offset(D)
% initial parameters ( [K_1/2 B_max offset] )
x0 = [mean(D(:,1)); max(D(:,2)); 0];
% log transform parameters so minimization doesn't go negative
x0(1:2) = log(x0(1:2));
% set options fminsearch
options = optimset('MaxFunEval',1e5,'TolX',1e-12,'TolFun',1e-12);
% minimization algorithm
[x_out fval] = fminsearch(@(x_) err_func(x_,D), x0,options);
% convert minimized parameters to structure
p_out.Kd = exp(x_out(1));
p_out.Bmax = exp(x_out(2));
p_out.offset = x_out(3);
p_out.sse = fval;

function err = err_func(x_,D)
    ec = D(:,1); % experimental concentrations
    er = D(:,2); % experimental responses
    % return log transformed parameters to original
    x_(1:2) = exp(x_(1:2));
    % calculate fit for current parameter set
    fit = x_(2) * (ec ./ (ec + x_(1))) + x_(3);
    % calculate sum of squared error
    err = (sum((er - fit).^2));
end
end
function plotter(fits)
clf
subplot(4,1,1)
hist(fits(:,1)); set(gca,'color','none'); title('Fit K_D');
subplot(4,1,2)
hist(fits(:,2)); set(gca,'color','none'); title('Fit B_m_a_x');
subplot(4,1,3)
hist(fits(:,3)); set(gca,'color','none'); title('Fit Offset');
subplot(4,1,4)
plot(fits(:,4),'r'); set(gca,'color','none'); title('ERRORS')
end
```

```
function out = HILLstrap(isos,n)
% isos is a two column vector input
%       column 1 is concentration
%       column 2 is recorded steady state response
% n is the number of bootstrap cycles to perform

% preallocate fits matrix
fits = zeros(n,5);
pts = size(isos,1);
% bootstrap for n cycles
for k = 1:n
    which = randi([1,pts],pts,1); % random selection of points to include
    boot = f_kk_fithill_offset(isos(which,:));
    fits(k,1) = boot.Kd;
    fits(k,2) = boot.nH;
    fits(k,3) = boot.offset;
    fits(k,4) = boot.Bmax;
    fits(k,5) = boot.sse;
    disp(fits(k,:))
end
% sort fits by associated errors
fits = sortrows(fits,5);
% histogram fit parameters
plotter(fits);
% return fits matrix
out = fits;
end
%%
function p_out = f_kk_fithill_offset(D)
% initial parameters ( [K_1/2 n_H B_max offset] )
x0 = [mean(D(:,1)); 1; max(D(:,2)); 0];
% log transform parameters so minimization doesn't go negative
x0(1:3) = log(x0(1:3));
% set options fminsearch
options = optimset('MaxFunEval',1e5,'TolX',1e-12,'TolFun',1e-12);
% minimization algorithm
[x_out fval] = fminsearch(@(x_) err_func(x_,D), x0,options);
% convert minimized parameters to structure
p_out.Kd = exp(x_out(1));
p_out.nH = exp(x_out(2));
p_out.Bmax = exp(x_out(3));
p_out.offset = x_out(4);
p_out.sse = fval;

function err = err_func(x_,D)
    ec = D(:,1); % experimental concentrations
    er = D(:,2); % experimental responses
    % return log transformed parameters to original
    x_(1:3) = exp(x_(1:3));
    % calculate fit for current parameter set
    fit = x_(3) * ec.^x_(2) ./ (ec.^x_(2) + x_(1).^x_(2)) + x_(4);
    % calculate sum of squared error
    err = (sum((er - fit).^2));
end
end

function plotter(fits)
clf
subplot(4,1,1)
hist(fits(:,1)); set(gca,'color','none'); title('Fit K_1_/_2')
subplot(4,1,2)
hist(fits(:,2)); set(gca,'color','none'); title('Fit n_H')
subplot(4,1,3)
hist(fits(:,4)); set(gca,'color','none'); title('Fit B_m_a_x')
subplot(4,1,4)
plot(fits(:,5),'r'); set(gca,'color','none'); title('ERRORS')
end
```

```
function out = AEDist_solver(isos,n_grid_points,its)
% isos is a two column vector input
%       column 1 is concentration
%       column 2 is recorded steady state response
% n_grid_points determines the sampling of K_D space between the low and
%       high concentrations
% its is number of expectancy maximization cycles
cdata = isos(:,1);
rdata = isos(:,2);
fin = ones(n_grid_points,1); % Initial Distribution Weights
%% make energy distribution vector (even spacing)
log_cdata = log(cdata);
lnkd_min = min(log_cdata);
lnkd_max = max(log_cdata);
log_space = lnkd_max - (lnkd_max - lnkd_min)*(0:(n_grid_points-1))./(n_grid_points-1);
grid_vec = [exp(log_space)]';
delta_lnkd = (lnkd_max-lnkd_min)/(n_grid_points-1);
%% make theta matrix
theta = zeros(n_grid_points, size(cdata,1));
for k = 1:n_grid_points
    for j = 1:size(cdata,1)
        theta(k,j) = cdata(j)./(cdata(j)+grid_vec(k));
    end
end
%% expectency maximization
f_k = fin;
k = 0;
while k < its
    k = k+1; % count iterations
    % calculate R(C) w/ current distribution f_k
    R = theta'*f_k*delta_lnkd;
    % get ratio vector
    R_ratio = rdata./R;
    % calculate f_k_plus_1
    f_k_plus_1 = (f_k.*(theta*R_ratio))./sum(theta,2);
    % plot progress every 10000 iterations
    if rem(k,10000) == 0
        clf
        subplot(2,1,1) % data as *, current fit as o
            semilogx(cdata, rdata,'b*', cdata, R,'ro','MarkerSize',8,'LineWidth',0.8); set(gca,'Color','none')
            legend('expt','fit',4); legend('Boxoff')
            title(sprintf('iterations: %d',k)) % title with iterations
            ylabel('Response (RU)'); xlabel('Concentration (M)') % axis labels
        subplot(2,1,2)
            semilogx(grid_vec,f_k_plus_1,'Color',[0 0.2 1]); set(gca,'Color','none')
            xlabel(sprintf('sum(f): %0.4g',sum(f_k)*delta_lnkd))
        drawnow
    end
    % update f_k
    f_k = f_k_plus_1;
end
%% return distribution / fit / data / predicted bmax / error
out.DIST = [grid_vec f_k];
out.FIT = [cdata R];
out.DATA = [cdata rdata];
out.bmax = sum(f_k)*delta_lnkd;
out.sse = sum((rdata-R).^2);
end
```

Appendix 3:

Software for the analysis of ensemble models of protein structure

Workflow for analysis of ensemble models:

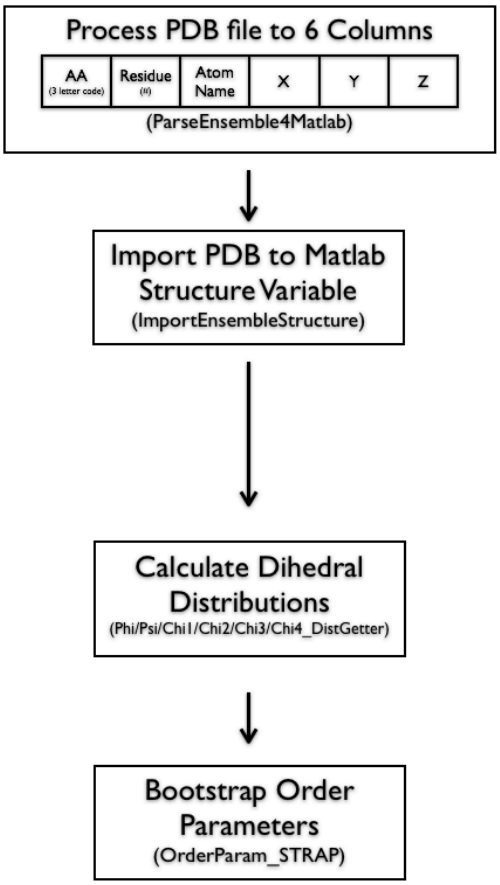


Figure 1. Workflow for Matlab processing of ensemble refined structures.

As implemented in Phenix, ensemble refinement produces a PDB file that contains multiple models of the structure. This data is rich, but there are no plug and play analytics in place. For this reason, I wrote a number of different Matlab scripts and functions in order to calculate dihedral angles, order parameters, and correlation coefficients between different models.

The lynchpin of my structural analytics is the use of structure variables within Matlab. The ability to loop over fieldnames and call upon the same elements from separate variables is wonderfully implemented in Matlab. Besides helping maintain a clean workspace, the dynamic assignment of fieldnames has proved very helpful in the analysis of large datasets.

For importing the ensemble models, the first step is converting the phenix generated pdb into a simple text file of six columns. For my purposes, I stripped all heteroatoms and hydrogens using awk one liners before rearranging the columns as shown in Figure 1. The following text can serve as a general starting point for a shell script to accomplish this:

```
echo "current file is $1"
echo "new file is $2"
awk '$3 !~ "H"' $1 > tmp
grep -v 'TER\|ANISOU\|OXT\|REMARK\|HOH\|CRYST\|SCALE\|MODEL' tmp > tmp2
awk '/./' tmp2 > tmp3
awk '$5 !~ "B"' tmp3 > tmp4
awk '{print $4,$6,$3,$7,$8,$9}' tmp4 > $2
rm tmp*
exit 0
```

The remaining steps are all accomplished through Matlab functions as scripted on the following page. This workflow calls on different functions, which are also included in the following pages.

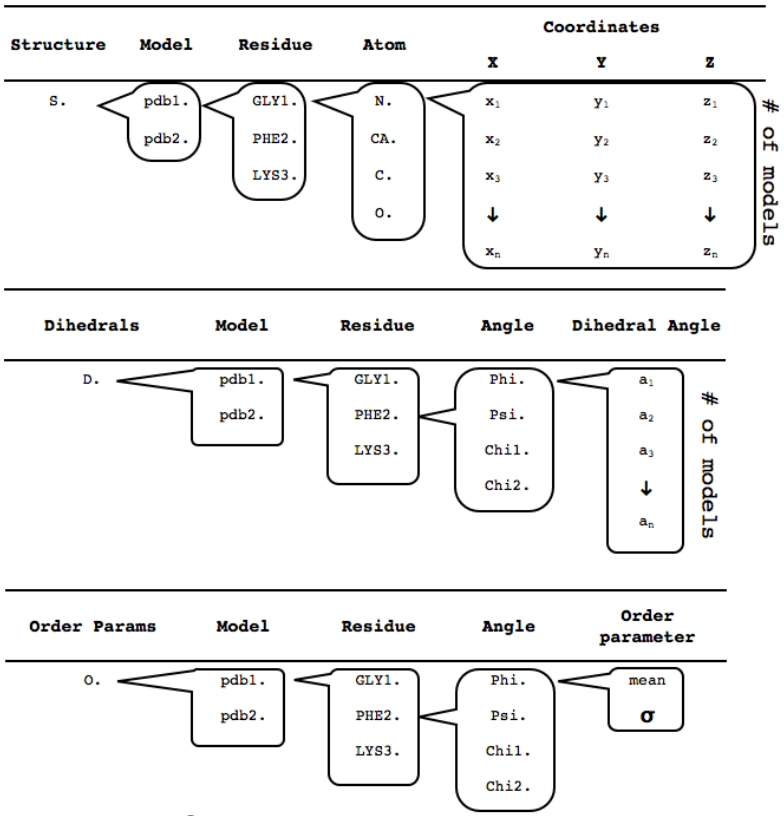


Figure 2. Structure variables for processing of ensemble models.

Script to import ensemble model and create dihedral and order parameter structure variables:

```
tmp = importdata('{fullpath}');
%% Convert Parsed PDB Text to Matlab Structure Variable
for k = 1:size(tmp.data,1)
    res = strcat(tmp.textdata(k,1),tmp.textdata(k,2));
    atom = tmp.textdata(k,3);
    try
        j = size(S.pdb1.(char(res)).(char(atom)),1);
        S.pdb1.(char(res)).(char(atom))(j+1,1:3) = tmp.data(k,:);
    catch
        S.pdb1.(char(res)).(char(atom))(1,1:3) = tmp.data(k,:);
    end
end
%%  Create Dihedral Angle Structure Variable
D.pdb1.Phi = kk_PhiDist_getter(S.pdb1);
D.pdb1.Psi = kk_PsiDist_getter(S.pdb1);
D.pdb1.Chi1 = kk_Chi1Dist_getter(S.pdb1);
D.pdb1.Chi2 = kk_Chi2Dist_getter(S.pdb1);
D.pdb1.Chi3 = kk_Chi3Dist_getter(S.pdb1);
D.pdb1.Chi4 = kk_Chi3Dist_getter(S.pdb1);
%% Boot Strap Dihedral Order Parameters
angs = fieldnames(D.pdb1);
resis = fieldnames(D.pdb1.Phi);
for a = 1:size(angs,1)
    for r = 1:size(resis,1)
        try
            O.pdb1.(char(angs(a))).(char(resis(r))) = kk_OrderParam_STRAP
(D.pdb1.(char(angs(a))).(char(resis(r))),200);
        end
    end
end
```

Functions Called:

Calculate Phi distribution:

```
function out = kk_PhiDist_getter(in)
resis = fieldnames(in);
r = 1;
while r < size(resis,1) - 1
    naa = in.(char(resis(r)));
    caa = in.(char(resis(r+1)));
    p1s = naa.C; p2s = caa.N; p3s = caa.CA; p4s = caa.C;
    out.(char(resis(r))) = kk_Dihedral(p1s,p2s,p3s,p4s);
    r = r+1;
end
end

function out = kk_Dihedral(p1s,p2s,p3s,p4s)

b1s = (p2s-p1s) ./ repmat(sum((p2s-p1s).^2,2).^0.5,1,3);
b2s = (p3s-p2s) ./ repmat(sum((p3s-p2s).^2,2).^0.5,1,3);
b3s = (p4s-p3s) ./ repmat(sum((p4s-p3s).^2,2).^0.5,1,3);
n1s = cross(b1s,b2s) ./ repmat(sum(cross(b1s,b2s).^2,2).^0.5,1,3);
n2s = cross(b2s,b3s) ./ repmat(sum(cross(b2s,b3s).^2,2).^0.5,1,3);
m1s = cross(n1s,b2s);
xs = dot(n1s,n2s,2);
ys = dot(m1s,n2s,2);
out = -(atan2(ys,xs)*180/pi);
end
```

Calculate Psi distribution:

```
function out = kk_PsiDist_getter(in)
resis = fieldnames(in);
r = 1;
while r < size(resis,1) - 1
    naa = in.(char(resis(r)));
    caa = in.(char(resis(r+1)));
    p1s = naa.N; p2s = naa.CA; p3s = naa.C; p4s = caa.N;
    out.(char(resis(r))) = kk_Dihedral(p1s,p2s,p3s,p4s);
    r = r+1;
end
end

function out = kk_Dihedral(p1s,p2s,p3s,p4s)

b1s = (p2s-p1s) ./ repmat(sum((p2s-p1s).^2,2).^0.5,1,3);
b2s = (p3s-p2s) ./ repmat(sum((p3s-p2s).^2,2).^0.5,1,3);
b3s = (p4s-p3s) ./ repmat(sum((p4s-p3s).^2,2).^0.5,1,3);
n1s = cross(b1s,b2s) ./ repmat(sum(cross(b1s,b2s).^2,2).^0.5,1,3);
n2s = cross(b2s,b3s) ./ repmat(sum(cross(b2s,b3s).^2,2).^0.5,1,3);
m1s = cross(n1s,b2s);
xs = dot(n1s,n2s,2);
ys = dot(m1s,n2s,2);
out = -(atan2(ys,xs)*180/pi);
end
```

Calculate Chi1 distribution:

```
function out = kk_Chi1Dist_getter(in)
resis = fieldnames(in);
r = 1;
while r < size(resis,1)
    curr = in.(char(resis(r)));
    p1s = curr.N; p2s = curr.CA;
    try
        p3s = curr.CB;
    catch
        p3s = 0;
    end
    try
        p4s = curr.CG;
    catch
        try
            p4s = curr.SG;
        catch
            try
                p4s = curr.CG1;
            catch
                try
                    p4s = curr.OG;
                catch
                    try
                        p4s = curr.OG1;
                    catch
                        p4s = 0;
                    end
                end
            end
        end
    end
    if p4s == 0 | p3s ==0
        r = r+1;
    else
        out.(char(resis(r))) = kk_Chi1(p1s,p2s,p3s,p4s);
        r = r+1;
    end
end
end

function out = kk_Chi1(p1s,p2s,p3s,p4s)

b1s = (p2s-p1s) ./ repmat(sum((p2s-p1s).^2,2).^0.5,1,3);
b2s = (p3s-p2s) ./ repmat(sum((p3s-p2s).^2,2).^0.5,1,3);
b3s = (p4s-p3s) ./ repmat(sum((p4s-p3s).^2,2).^0.5,1,3);
n1s = cross(b1s,b2s) ./ repmat(sum(cross(b1s,b2s).^2,2).^0.5,1,3);
n2s = cross(b2s,b3s) ./ repmat(sum(cross(b2s,b3s).^2,2).^0.5,1,3);
m1s = cross(n1s,b2s);
xs = dot(n1s,n2s,2);
ys = dot(m1s,n2s,2);
out = -(atan2(ys,xs)*180/pi);
end
```

Calculate Chi2 distribution:

```
function out = kk_Chi2Dist_getter(in)
%%
resis = fieldnames(in);
r = 1;
while r < size(resis,1)
    curr = in.(char(resis(r)));
    p1s = curr.CA;
    try
        p2s = curr.CB;
    catch
        p2s = 0;
    end
    try
        p3s = curr.CG;
    catch
        try
            p3s = curr.CG1;
        catch
            p3s = 0;
        end
    end

    try
        p4s = curr.CD; %CD1; ND1; OD1;
    catch
        try
            p4s = curr.CD1;
        catch
            try
                p4s = curr.ND1;
            catch
                try
                    p4s = curr.OD1;
                catch
                    try
                        p4s = curr.SD;
                    catch
                        p4s = 0;
                    end
                end
            end
        end
    end
    if p4s == 0 | p3s == 0 | p2s == 0
        r = r+1;
    else
        out.(char(resis(r))) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    end
end
end
 function out = kk_Dihedral(p1s,p2s,p3s,p4s)
b1s = (p2s-p1s) ./ repmat(sum((p2s-p1s).^2,2).^0.5,1,3);
b2s = (p3s-p2s) ./ repmat(sum((p3s-p2s).^2,2).^0.5,1,3);
b3s = (p4s-p3s) ./ repmat(sum((p4s-p3s).^2,2).^0.5,1,3);
n1s = cross(b1s,b2s) ./ repmat(sum(cross(b1s,b2s).^2,2).^0.5,1,3);
n2s = cross(b2s,b3s) ./ repmat(sum(cross(b2s,b3s).^2,2).^0.5,1,3);
m1s = cross(n1s,b2s);
xs = dot(n1s,n2s,2);
ys = dot(m1s,n2s,2);
out = -(atan2(ys,xs)*180/pi);
end
```

Calculate Chi3 distribution:

```
function out = kk_Chi3Dist_getter(in)
%%
resis = fieldnames(in);
r = 1;
while r < size(resis,1)
    curr = resis(r);  %aas with chi3: ARG LYS MET GLU GLN
    prs = regexp(curr,{'ARG','GLN','GLU','LYS','MET'});
    if prs{1} == 1
        p1s = in.(char(curr)).CB; p2s = in.(char(curr)).CG; p3s = in.(char(curr)).CD; p4s = in.
(char(curr)).NE;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    elseif prs{2} == 1
        p1s = in.(char(curr)).CB; p2s = in.(char(curr)).CG; p3s = in.(char(curr)).CD; p4s = in.
(char(curr)).OE1;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    elseif prs{3} == 1
        p1s = in.(char(curr)).CB; p2s = in.(char(curr)).CG; p3s = in.(char(curr)).CD; p4s = in.
(char(curr)).OE1;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    elseif prs{4} == 1
        p1s = in.(char(curr)).CB; p2s = in.(char(curr)).CG; p3s = in.(char(curr)).CD; p4s = in.
(char(curr)).CE;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    elseif prs{5} == 1
        p1s = in.(char(curr)).CB; p2s = in.(char(curr)).CG; p3s = in.(char(curr)).SD; p4s = in.
(char(curr)).CE;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    else
        r = r+1;
    end
end
end

function out = kk_Dihedral(p1s,p2s,p3s,p4s)

b1s = (p2s-p1s) ./ repmat(sum((p2s-p1s).^2,2).^0.5,1,3);
b2s = (p3s-p2s) ./ repmat(sum((p3s-p2s).^2,2).^0.5,1,3);
b3s = (p4s-p3s) ./ repmat(sum((p4s-p3s).^2,2).^0.5,1,3);
n1s = cross(b1s,b2s) ./ repmat(sum(cross(b1s,b2s).^2,2).^0.5,1,3);
n2s = cross(b2s,b3s) ./ repmat(sum(cross(b2s,b3s).^2,2).^0.5,1,3);
m1s = cross(n1s,b2s);
xs = dot(n1s,n2s,2);
ys = dot(m1s,n2s,2);
out = -(atan2(ys,xs)*180/pi);
end
```

(By the time I started writing code for Chi3 calculation, I abandoned the 'try' and 'catch' strategy and moved to a regular expression match to correctly call atom names for the dihedrals.)

Calculate Chi4 distribution:

```
function out = kk_Chi4Dist_getter(in)
%%

resis = fieldnames(in);
r = 1;
while r < size(resis,1)
    curr = resis(r);  %aas with chi4: ARG LYS
    prs = regexp(curr,{'ARG','LYS'});
    if prs{1} == 1
        p1s = in.(char(curr)).CG; p2s = in.(char(curr)).CD; p3s = in.(char
(curr)).NE; p4s = in.(char(curr)).CZ;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    elseif prs{2} == 1
        p1s = in.(char(curr)).CG; p2s = in.(char(curr)).CD; p3s = in.(char
(curr)).CE; p4s = in.(char(curr)).NZ;
        out.(char(curr)) = kk_Dihedral(p1s,p2s,p3s,p4s);
        r = r+1;
    else
        r = r+1;
    end
end
end

function out = kk_Dihedral(p1s,p2s,p3s,p4s)

b1s = (p2s-p1s) ./ repmat(sum((p2s-p1s).^2,2).^0.5,1,3);
b2s = (p3s-p2s) ./ repmat(sum((p3s-p2s).^2,2).^0.5,1,3);
b3s = (p4s-p3s) ./ repmat(sum((p4s-p3s).^2,2).^0.5,1,3);
n1s = cross(b1s,b2s) ./ repmat(sum(cross(b1s,b2s).^2,2).^0.5,1,3);
n2s = cross(b2s,b3s) ./ repmat(sum(cross(b2s,b3s).^2,2).^0.5,1,3);
m1s = cross(n1s,b2s);
xs = dot(n1s,n2s,2);
ys = dot(m1s,n2s,2);
out = -(atan2(ys,xs)*180/pi);
end
```

Boot strap order parameters:

```
function out = kk_OrderParam_STRAP(dist,n)
 N = size(dist,1);
orderparam = zeros(n,1);
for k = 1:n
    which = randi(N,N,1);
    unvs = [cosd(dist(which)) sind(dist(which))];
    suv = sum(unvs,1);
    orderparam(k) = 1/N * norm(suv);
end

out = [mean(orderparam);std(orderparam)];
end
```

Appendix 4:

Chromodomain-mediated oligomerization of HP1 suggests a nucleosome bridging mechanism for heterochromatin assembly

In collaboration with:

Daniele Canzio, Evelyn Chang, Smita Shankar, Matthew D. Simon, Diten D. Madhani, Geeta J. Narlikar, and Bassem Al-Sady.

Reprinted with permission from:

Canzio D, Chang E, Shankar S, Kuchenbecker KM, Simon M, Madhani H, Narlikar G, Al-Sady B. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome bridging mechanism for heterochromatin assembly. Mol Cell. 2011 Jan 7;41(1):67-81.

Foreword:

Swi6 oligomerization results from the integration of three signals.

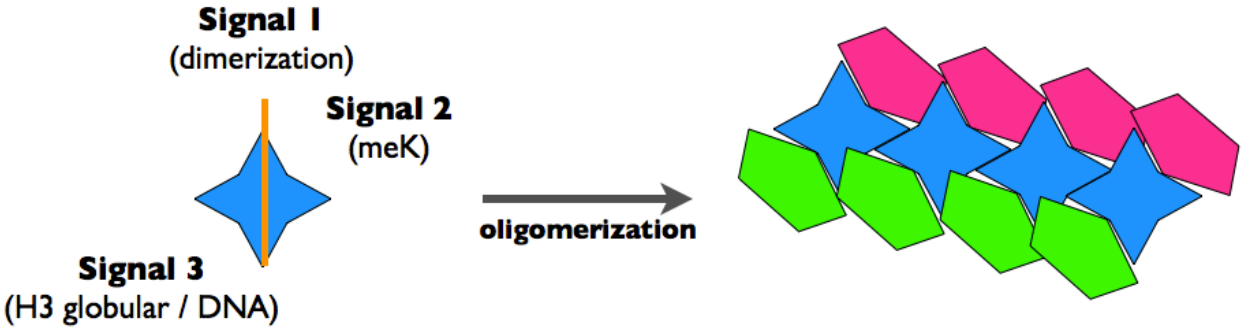


Figure f1. HP1 protein undergoes four binding processes. Dimerization enables methyl lysine recognition, which promotes weak binding to the globular region of histone H3 and DNA-protein interactions. This confluence of signals enables oligomerization and bridging between nucleosomes.

The fission yeast (*S. pombe*) homolog of heterochromatin protein 1, Swi6, is involved in transcriptional repression by maintenance of regions of tightly packed DNA. The basic unit of these heterochromatic regions is the nucleosome, which is comprised of eight histone proteins that are encircled by ~146 base pairs of DNA. These regions are dynamic and changes to cellular transcription programs can be mediated by spreading or sliding of heterochromatic regions. While the higher order structure and dynamics are under active investigation, it was this work by Daniele Canzio (Narlikar Lab) that helped lay the mechanistic framework for the role of Swi6 in local nucleosome structure.

My knowledge of higher order DNA structures was limited to the trivial fact that each human cell contains ~six feet of DNA in its nucleus. In order to package the DNA inside the ~six micron nucleus, the DNA must adopt a condensed structure. This tight packing provides a straightforward mechanism of transcriptional regulation as long stretches of DNA will be inaccessible to the transcriptional machinery. As a mode of transcriptional regulation, these structures require both thermodynamic stability and kinetic flexibility. Post translational

modification of the nucleosomal core (histone proteins) provides an enzymatic switch that, if coupled to the binding of scaffold proteins, could promote local reorganization and subsequent gene transcription.

I came to this project following the first round of reviews. The reviewers had requested a more sensitive direct binding measurement, and I still remember Daniele approaching me to ask if we could use SPR to measure the specifity of a protein interaction with methylated vs. unmethylated nucleosomes. At this point in my graduate career, I had begun to realize the potential of ultra-low density surfaces for SPR measurement. The huge size of nucleosomes (200kDa, ~50Å x 100Å) would mandate ultra-low density surfaces and presented a good test case for this novel approach to the SPR measurement.

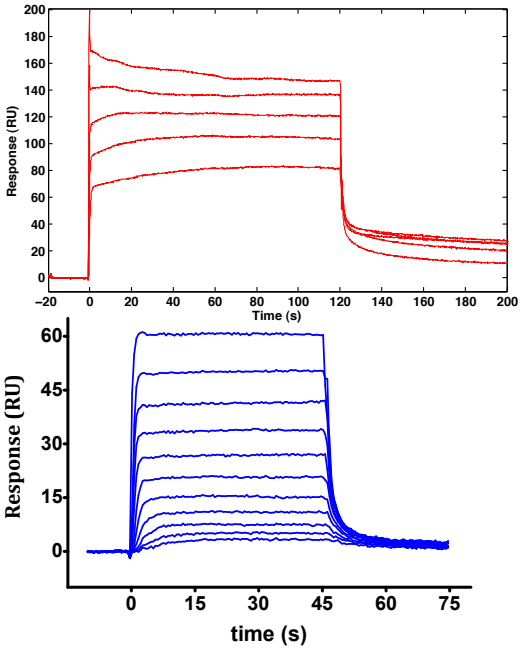


Figure f2. Improvements to the quality of the data was due to protein purification and surface density optimization.

In Figure f2, the initial data collection is shown in the red traces. There are a number of problems with this data. Perhaps most notable is the huge amount of protein sticking to the surface following injection. In the bottom panel, data from the optimized system is plotted in blue. There are three major differences between these experiments. The red traces were collected using a commercial streptavidin chip, giving poor control over the immobilization level. The blue traces were collected against a home-made flat streptavidin surface that allowed for precise control of capture level and diminished non-specific interactions with the dextran matrix. The final experimental conditions attempted to capture one biotinylated nucleosome

per 1000Åx1000Å area of the SPR chip. This capture level corresponded to approximately 30RU. Buffer was also optimized from low salt / high detergent (100mM/0.1%np20) to physiological salt / low detergent (150mM / 0.005%np20). However, I believe the biggest contributor to data improvement was a better preparation of the protein. The red traces are from a ~90% pure His-tagged sample, while the blue traces were obtained from a 99% pure sample following cleavage of the His-tag and three rounds of gel filtration: Initial sizing; Anion exchange; Final sizing. One of my favorite aspects of SPR is that bad data should scream at you while good data looks like biology.

Establishing the assay took a considerable amount of experimental work. And upon having a well-controlled stable system for data collection we were met with the challenge of an unfittable isotherm. The isotherms displayed complex curvature and failed to saturate, see Figure f3. Excited with the reproducibility of the data, but frustrated with the lack of an appropriate binding model, I turned to Scatchard plots.

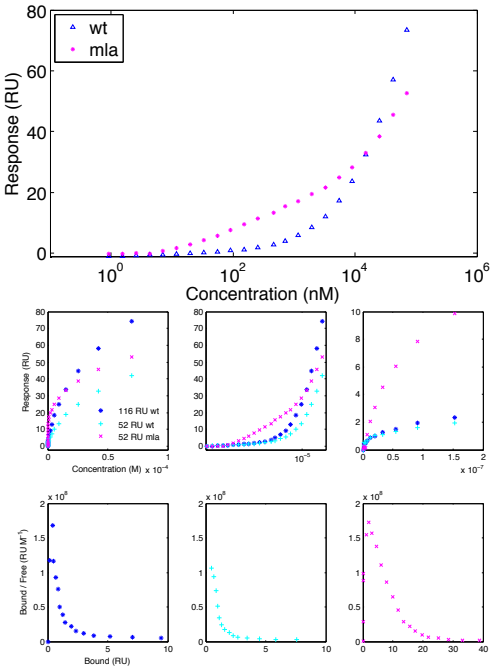


Figure f3. Non-ideality in the binding of Swi6 to nucleosomes.

As one of my favorite quotes from Robert over my time in his lab, upon showing him the presence of both positive and negative cooperativity in the isotherms, he exclaimed: "Scatchard plots are revelatory." As the end goal of this project was to describe the difference between the binding of Swi6 to methylated vs. non-methylated nucleosomes, I focused on the distinction in two major features of the Scatchard plots. For the methylated nucleosomes, the early hump is evidence of positive

cooperativity at very low concentrations (< 20nM), which indicates a high affinity site that may promote dimerization of the Swi6 molecule in the context of nucleosome binding. Also, for both methylated and non-methylated nucleosomes there is a tight event and a weak event, but for the methylated nucleosomes, this tight event has a significantly different x-intercept, indicating a different stoichiometry. So, besides the evidence of slight cooperativity for methylated nucleosomes, the methylated nucleosomes seem to more readily promote higher order assembly.

This project marked my first legitimate foray into real modeling of binding processes. When I began, I happily used Kaleidagraph. But in order to model this more complex process, I discovered a profound appreciation for Matlab.

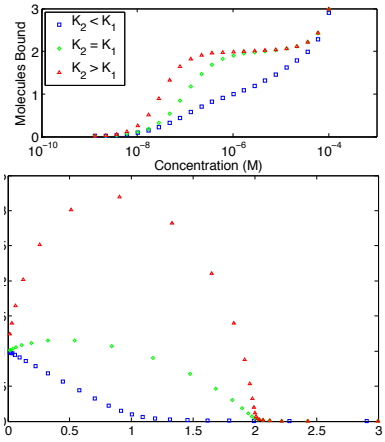


Figure f4. A two-site BET adsorption model, gives differential cooperativity while accounting for the unsaturability.

In feed forward modeling to account for the two principal features of the Scatchard analysis, I came across two models. The BET adsorption model was developed in 1938 by Brunauer, Emmett, and Teller. This model allows for a pseudo polymerization whereby binding of the first ligand creates a binding site for the second, etc... This model helped provide a conceptual framework for explanation of the shared unsaturability between the methylated and non-methylated nucleosomes, with a physical basis for the early rise and plateau of the methylated binding. The other model I explored for this system explicitly treated the multimerization of Swi6, with different oligomers possessing different affinities for the nucleosome. Ultimately, both of these models were inapplicable because we did not have adequate knowledge of the overall oligomeric partition function. So, our concentration axis was arguably meaningless.

Chromodomain-mediated oligomerization of HP1 suggests a nucleosome bridging mechanism for heterochromatin assembly

Daniele Canzio[1,2], Evelyn Y. Chang[1,3], Smita Shankar[1], Kristopher M. Kuchenbecker[1,4], Matthew D. Simon[5], Hiten D. Madhani[1], Geeta J. Narlikar[1,*] and Bassem Al-Sady[1,*]

[1] Department of Biochemistry and Biophysics, University of California San Francisco, 94158, USA.
[2] Chemistry and Chemical Biology Graduate Program University of California San Francisco, 94158,USA. *[3] Tetrad Graduate Program University of California San Francisco, 94158, USA.*
[4] Biophysics Graduate Group University of California San Francisco, 94158,USA.
[5] Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114 and Department of Genetics, Harvard Medical School, Boston, MA 02115.
*To whom correspondence should be addressed geeta.narlikar@ucsf.edu, bassem.al-sady@ucsf.edu

Summary

HP1 proteins are central to the assembly and spread of heterochromatin containing histone H3K9 methylation. The chromodomain (CD) of HP1 proteins specifically recognizes the methyl mark on H3 peptides, but the same extent of specificity is not observed within chromatin. The chromoshadow domain of HP1 proteins promotes homodimerization, but this alone cannot explain heterochromatin spread. Using the *S. pombe* HP1 protein, Swi6, we show that recognition of H3K9 methylated chromatin *in vitro* relies on a newly identified interface between two CDs. This interaction causes Swi6 to tetramerize on a nucleosome, generating two vacant CD sticky ends. On nucleosomal arrays, methyl-mark recognition is highly sensitive to inter-nucleosomal distance, suggesting that the CD sticky ends bridge nearby methylated nucleosomes. Strengthening the CD-CD interaction enhances silencing and heterochromatin spread *in vivo*. Our findings suggest that recognition of methylated nucleosomes and HP1 spread on chromatin are structurally coupled, and imply that methylation and nucleosome arrangement synergistically regulate HP1 function.

Introduction

Histone H3 lysine 9 methylated (H3K9me3) heterochromatin, conserved from yeast to humans, is a highly versatile nuclear structure. It is required for centromere formation, heritable gene silencing, repression of recombination, sister chromatid cohesion, and maintenance of telomere stability (Grewal and Jia, 2007). A hallmark of this type of heterochromatin is the formation of macromolecular assemblies that can spread along chromatin from specific nucleation sites (Hall et al., 2002). The structural features that allow H3K9me3 based heterochromatin to spread and fulfill its various nuclear functions, however, are not well understood.

At the core of heterochromatic macromolecular assemblies lies the HP1-H3K9me3 chromatin complex, which is thought to mediate the many functions of heterochromatin through the recruitment of diverse sets of regulators (Grewal and Jia, 2007; Smothers and Henikoff, 2000). In gene silencing, HP1 proteins are thought to reduce RNA polymerase occupancy by both recruiting accessory silencing factors (Fischer et al., 2009) and by forming less accessible chromatin structures (Danzer and Wallrath, 2004). HP1 proteins have been proposed to enable post-transcriptional gene silencing by recruiting RNA processing machinery (Iida et al., 2008). Understanding how HP1 proteins recognize and bind H3K9me3 chromatin is thus central to understanding both the molecular mechanisms of heterochromatin assembly and how this type of heterochromatin fulfills its wide range of functions.

Previous work has described individual aspects of the HP1/H3K9me3 nucleosome complex. HP1 proteins contain three recognized protein domains: 1) a chromodomain (CD), 2) an evolutionarily related chromoshadow domain (CSD), and 3) a poorly defined hinge (H) region between the CD and CSD. The CD is part of a family of proteins that contain a

specialized hydrophobic cage, formed by aromatic residues, that bind methyl marks on histones with high specificity but low affinity (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). The CSD is involved in dimerization of HP1 proteins (Cowieson et al., 2000) and is important for the silencing function of HP1 proteins (Sadaie et al., 2008). The H region is thought to be required for non-specific binding of HP1 proteins to DNA, as observed *in vitro* (Meehan et al., 2003; Zhao et al., 2000). Despite these key findings, several questions remain about how the functions of these individual domains are integrated to allow stable recognition of the physiological template, H3K9 methylated chromatin. For example, it is not clear whether the weak binding of the CD for methylated tail peptides observed *in vitro* is sufficient to guide heterochromatin assembly to the correct sites *in vivo*. In particular, the strong non-specific binding of HP1 proteins to inter-nucleosomal DNA (Meehan et al., 2003; Yamada et al., 1999) raises the question of how specificity for the methyl mark is attained in the context of chromatin**.** Finally, while HP1 proteins can dimerize via the CSD, such homodimerization alone appears insufficient to explain the ability of these proteins to spread along chromatin.

To address these questions, we used the *S. pombe* HP1 protein, Swi6, as a model system. *S. pombe* contains only a single H3K9 methyltransferase, Clr4, along with two HP1 proteins, Chp2 and Swi6, of which Swi6 is more abundant (Grewal and Jia, 2007; Sadaie et al., 2008). We reconstituted the core Swi6-H3K9me3 chromatin complex using recombinant Swi6 and chromatin templates that are homogeneously methylated at H3K9 using methyl lysine analog (MLA) technology (Simon et al., 2007). We analyzed the biochemical properties of this complex and tested our key conclusions *in vivo*. Our results suggest a mechanism of heterochromatin formation in which HP1 proteins utilize a process of step-wise higher order oligomerization.

This process is mediated by interactions between CDs to interpret information encoded in both the methylation state and the underlying nucleosomal arrangement of chromatin.

Results

Swi6 recognizes the H3K9 methyl mark within mononucleosomes and forms oligomers on mononucleosomes and in solution

Previous studies have reported on the ability of Swi6 to preferentially bind the H3K9me3 mark in the context of H3 tail peptides (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002; Yamada et al., 2005). However, the magnitude of discrimination observed within H3 tail peptides has not been recapitulated in the context of chromatin, largely due to the challenge of generating homogeneously methylated chromatin. We produced homogenously methylated nucleosomes using methyl lysine analogs (MLAs), then investigated the ability of recombinant Swi6 to specifically recognize methylated nucleosomes using two different equilibrium approaches. For both approaches, unmodified (H3K9) and methylated (H3Kc9me3) nucleosomes were assembled on 147 base pairs of the nucleosome positioning sequence 601 (Figure 1a).

In the first approach, surface plasmon resonance (SPR) was used to assay binding of Swi6 to H3K9 and H3Kc9me3 nucleosomes (Figure 1b). Analysis of the binding kinetics (traces in Figure 1b, inset) revealed no large differences in the association rates, but comparison of the dissociation traces reveals that Swi6 dissociates more rapidly from H3K9 nucleosomes compared to H3Kc9me3 nucleosomes, consistent with specific binding of Swi6 to methylated nucleosomes (Figure 1b). Because kinetic analysis of SPR data can be problematic and at times unreliable, we further optimized the assay for equilibrium measurements. The equilibrium binding isotherms clearly reveal two features (Figure 1c; see also Figure S1b&c). At low

concentrations (10 nM – 1 μM), there is a methylation specific interaction that approaches but does not reach saturation. At high concentrations (>1 μM), there is apparently a weak, non-saturable interaction, and the concentration dependence of this interaction is similar for the H3K9 and the H3Kc9me3 nucleosome surfaces. We were, however, unable to fit a physically meaningful model to the data because (i) the data do not reveal saturation and therefore cannot be used to determine a final stoichiometry and (ii) HP1 proteins are known to oligomerize in solution, so the concentration will change as function of the oligomeric state of Swi6 (See Figure S1e&f for more detailed discussion).

Despite the inability to fit a quantitative model to the data, the Swi6 concentration dependence reveals interesting features of the interaction of Swi6 with nucleosomes. The results imply the presence of at least two types of Swi6 binding events: one that occurs at concentrations below 1 μM and involves recognition of the methyl mark, and a second that occurs primarily at higher concentrations, is less sensitive to the presence of the methyl mark and is suggestive of step-wise Swi6 oligomerization.

To further investigate the Swi6 behavior observed by SPR, we measured Swi6 binding to core nucleosomes using a fluorescence polarization based approach. Using nucleosomal DNA labeled at one end with fluorescein, we monitored the gain in fluorescence polarization as a function of Swi6 concentration (Figure 1d, schematic, also see Extended Experimental Procedures). Analogous to the SPR data, we observe a binding profile that contains a methylation specific concentration regime and a non-saturable concentration regime.

The above results raised the question of what physical processes underlie the different types of binding events implied by the unusual concentration dependence. We hypothesized that the binding events in the methyl mark specific concentration regime reflect direct binding of

Swi6 to the nucleosome and the H3K9 residue while the binding events in the non-saturable concentration regime reflect mainly Swi6-Swi6 interactions that are scaffolded by the initial Swi6-nucleosome complex. The non-saturable behavior would then arise because addition of each Swi6 molecule would generate a new binding site for another Swi6 molecule, reflecting an intrinsic property of Swi6 to self-associate. To test this hypothesis, we investigated the oligomeric states adopted by Swi6 in solution under the two concentration regimes.

To determine the oligomeric state of Swi6 in the methylation specific concentration regime, we used two complementary approaches: (i) a cross-linking based approach and (ii) isothermal titration calorimetry (ITC). Over concentrations ranging from 25-5000 nM, cross-linker treated wild-type Swi6 migrates on SDS-PAGE gels at a mass consistent with a dimer, while the previously described dimer-disrupting CSD mutant, L315D, migrates at a mass consistent with a monomer (Cowieson et al., 2000) (Figure 2a). We then used ITC to obtain a more quantitative estimate of the K_d of the known dimerization domain of Swi6, the CSD (Figure 2b). Consistent with the cross-linking data, titrations of the WT Swi6 CSD into buffer produced no detectable heat release even at 17 nM indicating that K_d for CSD self-association is below 17 nM (Figure 2b, left panel). In contrast, titrations for the CSD domain containing the L315D mutation produce significant heat release and suggest a K_d for self-association of this mutant CSD in the high micromolar range (Figure 2b, right panel). Together, these two approaches suggest that at low nanomolar concentrations, Swi6 mainly exists as a dimer in the absence of nucleosomes.

We next determined the oligomeric states that can be adopted by Swi6 in the non-saturable concentration regime. We had noticed that under cross-linking conditions, Swi6 can form oligomers larger than a dimer (Figure 2a, indicated by asterisk), consistent with previous studies

on HP1 (Yamada et al., 1999; Zhao et al., 2000). To investigate the formation of defined higher-order oligomers and obtain true masses independent of oligomer shape we used a multi-angle light scattering (MALS) approach (Extended Experimental Procedures). The WT Swi6 protein forms mainly dimers at 20 μM (Figure 2c). Interestingly, approximately 5% of the protein is tetrameric, suggesting that Swi6 is capable of forming oligomers beyond a dimer. In contrast, the L315D mutation drastically reduces the ability of Swi6 to dimerize: more than 90% of the L315D is monomeric at 20 μM, in agreement with the ITC data (Figure 2b). The inter-molecular cross-linking approach described above enabled further stabilization of the higher order oligomeric states for analysis by MALS. Using this approach we found that WT Swi6 can form discrete complexes corresponding to dimeric, tetrameric, and octameric states (Figure 2d), whereas the L315D mutant is strongly impaired in forming such oligomeric states (Figure S2a). These data indicate that Swi6 can form well-defined higher order complexes in solution. Further, the Swi6 concentration regime in which states beyond dimer become populated correlates with the non-saturable concentration regimes of Figures 1c and d, suggesting that the non-saturable concentration regime mainly reflects Swi6-Swi6 interactions.

The above characterization of the oligomeric states of Swi6 indicates that Swi6 exists as a preformed dimer in the concentration regime in which we observed discrimination between H3Kc9me3 and H3K9 mononucleosomes. Further, the intrinsic property of Swi6 to form higher order oligomers suggests a potential for such oligomerization in binding across multiple nucleosomes within a nucleosomal array. To examine this possibility, we isolated the steps involved in direct recognition of the H3 tail within a mononucleosome, then used the information derived from these studies to better understand how Swi6 functions in the context of multiple nucleosomes.

Swi6 displays lower specificity for the H3K9me3 mark in mononucleosomes compared to H3 tail peptides

We reasoned that, by following the disappearance of the unbound nucleosomes in a gel mobility shift assay, we could better separate direct binding of Swi6 to the nucleosome from subsequent binding events that might entail mainly Swi6-Swi6 contacts. We measured the Swi6 concentration dependence for disappearance of unbound nucleosomes and obtained a value for $K_{1/2}$, which represents the concentration of Swi6 at which half of the nucleosomes remain unshifted. Most of the unbound MLA nucleosomes completely disappear by 1 μM Swi6 (Figure 3a). At higher concentrations we observe further, apparently continuous upshifting of the complexes, consistent with the nucleosome-scaffolded oligomerization behavior inferred from Figures 1c and d.

Using the above approach of quantifying $K_{1/2}$ values, we found that Swi6 prefers H3Kc9me3 over H3K9 nucleosomes by 5-fold (Figure 3a, right panel; specificity is expressed as a ratio of $K_{1/2}$ for H3K9 to that for H3Kc9me3 nucleosomes). Swi6 binds H3Kc9me0 nucleosomes with the same affinity as H3K9 nucleosomes (Figure S3a). We obtained the same 5-fold specificity for H3Kc9me3 over H3K9 nucleosomes using an equilibrium binding assay, in which the two types of nucleosomes compete with a fluorescently labeled DNA molecule for binding to Swi6 (Figure S3e).

Both the above assays indicate that the specificity for the methyl mark on core nucleosomes is substantially lower than that observed for the methyl mark on H3 tail peptides (Figures 3b&c and S3b). The results suggest a model in which Swi6 can bind to a core nucleosome in alternative orientations that lack interactions between the H3K9 residue and the CD, in addition to orientations that recognize the H3K9 residue. The binding orientations that

lack interactions between the CD and H3K9 could arise from the previously described abilities of the hinge and the CSD domains to interact with other regions of the nucleosome, such as the DNA and a globular region of H3, respectively (Dawson et al., 2009; Lavigne et al., 2009; Meehan et al., 2003). The above model predicts that increasing alternative binding interactions between Swi6 and the nucleosome will decrease the observed specificity for the methyl mark, as a smaller proportion of Swi6 molecules would bind in H3K9 recognizing orientations. At the same time we expect that the overall affinity will increase, as increasing the number of alternative binding orientations will increase the binding options of Swi6. Given that the affinity of HP1 proteins for free DNA increases with DNA length (Zhao et al., 2000) and our observations for Swi6 (Figure S3d), increasing the flanking DNA could be one way to increase the number of alternative binding modes. Consistent with these predictions, we find that increasing the flanking DNA length on one or both sides of a nucleosome results in a reduction in specificity but a gain in overall affinity (Figure 3d and Figure S3c).

Application of a simple quantitative model suggests that, for Swi6-H3Kc9me3 core nucleosome complexes, 94% of the Swi6 molecules are bound in H3K9-specific orientations and 6% are bound in alternative orientations (Extended Experimental Procedures). In contrast, for Swi6-H3K9 core nucleosome complexes, only 0.1% of the Swi6 molecules are bound in H3K9-specific orientations and >99% are bound in alternative orientations. Thus, in the context of H3K9 nucleosomes, the large fraction of Swi6 molecules bound in alternative orientations is expected to mask the binding contributions from molecules bound in H3K9-specific orientations. Together, the above observations raise the possibility that the specificity of HP1 proteins for the H3K9me3 mark could be controlled in part by regulating alternative binding

orientations. The experiments that follow provide a structural and energetic framework to understand how such regulation might occur.

The core unit of Swi6 binding to a mononucleosome is a tetramer

Quantification of the gel mobility shifts results suggests that binding of Swi6 to either H3Kc9me3 or H3K9 core nucleosomes occurs cooperatively with Hill coefficients of ~1.7 and 2.0, respectively (Figure 3b) suggesting that at least two molecules of Swi6 bind to one nucleosome. Further, the analysis in figures 2a and b indicates that, at the concentrations used in the native gel-shift assay, Swi6 is a dimer in solution. The cooperative binding could then reflect an additional interaction between two or more Swi6 dimers on the nucleosome. Indeed, the MALS data from Figure 2 indicate that Swi6 can form tetramers and octamers in the absence of nucleosomes at high concentrations. Alternatively, the two Swi6 dimers may not directly interact, but binding by two or more dimers may be required to stably upshift the nucleosomes on a native gel.

To directly determine the stoichiometry of the Swi6-core nucleosome complex in solution, under the methylation specific concentration regimes of Figures 1c, 1d and 3a, we used sedimentation velocity analytical ultracentrifugation (SV-AUC). SV-AUC allows the differentiation of multiple species present in the sample based on their mass dependent migration. Recent improvements in the analysis tools for SV-AUC data allow the determination of masses of multi-protein complexes while directly accounting for differences in shapes (Brown and Schuck, 2006). We performed three independent experiments each for samples containing H3Kc9me3 core nucleosomes alone (Figure 4a), H3Kc9me3 core nucleosomes bound by L315D Swi6 (Figure 4b), or H3Kc9me3 core nucleosomes bound by WT Swi6 (Figure 4c). We used Swi6 and nucleosome concentrations based on titration experiments (See Experimental

Procedures). Each experiment was analyzed using two independent models for data fitting: (i) a continuous two-dimensional function $c(s,f/f_0)$ for sedimentation coefficient s and hydrodynamic translational frictional ratio f/f_0, (Figure S4a), and (ii) a continuous function $c(s)$ for sedimentation coefficient s with a bi-modal f/f_0 distribution (Figure S4b) (f/f_0 is a measure of the shape of the complex, see Experimental Procedures).

Both analysis methods indicate that the majority of the complexes have a stoichiometry of four WT Swi6 proteins to one core nucleosome (Figure 4c). These findings suggest that two WT Swi6 dimers bind to the nucleosome to form a tetramer. Further, the molar mass obtained for the core nucleosome-L315D Swi6 complex reveals a stoichiometry of two Swi6 proteins to one core nucleosome (Figure 4b).

These observations suggest a model in which the two unoccupied Swi6 chromodomains in the Swi6 tetramer can serve as sticky ends (Figure 4c, black arrows) that can bind methyl marks on nearby nucleosomes. Binding of proximal nucleosomes via this specific type of sticky ends architecture would be predicted to energetically favor H3K9-specific binding orientations of the Swi6 tetramer over alternative binding orientations, resulting in greater specificity for the methyl mark. To test this hypothesis, we determined whether Swi6 binds with greater specificity to methylated di- and polynucleosome constructs compared to mononucleosomes.

Swi6 binds with similar specificity to mono and dinucleosomes

Dinucleosomes were first assembled on a DNA construct containing 15 base pairs of linker DNA (L15) between two 601 positioning sequences (Figure 5a, diagram). The relatively short linker length is designed to mimic inter-nucleosomal distances prevalent in *S. pombe* (Godde and Widom, 1992; Lantermann et al., 2010). Native gel mobility shift assays show that Swi6 binds to methylated 2N(L15) with approximately 2.5-fold higher affinity than to the

unmethylated control (Figure 5a). This specificity is comparable to that observed for mononucleosomes containing 20 bp of flanking DNA and is likely due to nonspecific binding of Swi6 to the linker DNA (Figure 3d and Fig. S3c). These results suggest that L15 dinucleosomes do not increase specific binding by Swi6. To test if these closely spaced nucleosomes sterically interfere with Swi6 binding to the H3 tails of both nucleosomes, we also measured Swi6 binding to a 2N(L47) dinucleosome, linked by 47 base pairs of DNA. Swi6 binds to methylated 2N(L47) with 2.5-fold higher affinity than unmethylated 2N(L47), ruling out a simple steric interference model (Figure 5b).

The above results suggest that, in the context of Swi6 binding, a dinucleosome substrate behaves like two unlinked mononucleosomes with flanking DNA and does not show any amplification of specificity. *In vivo*, however, Swi6 binds along many nucleosomes (Noma et al., 2001), leaving the possibility that the sticky ends mechanism may have evolved to have a larger effect in the context of a long stretch of nucleosomes.

Nucleosome arrays provide a highly specific substrate for Swi6

We next measured Swi6 binding to a 12 nucleosome array containing the same 15 base pair linker length as used in the dinucleosome construct (Figure 5c, diagram and Figure S5a). Native gel mobility shifts show that this 12N(L15) array substrate substantially increases Swi6 specificity for the methyl mark, to ~25-fold (Figure 5c). This represents a ~10-fold amplification in specificity compared to that measured for the corresponding dinucleosome construct. The large gain in specificity on nucleosomal arrays is consistent with our model (Figure 4c), in which bridging interactions between nucleosomes, mediated by vacant CD sticky ends, favor binding of Swi6 in H3K9-specific orientations over alternative orientations. The observation that H3K9me specificity is amplified only in the context of 12N arrays but not dinucleosome substrates,

suggests that Swi6 bridging requires nucleosome conformations that cannot be accessed by dinucleosomes.

If bridging nearby nucleosomes is required for correctly orienting Swi6 complexes, then increasing the distance between nucleosomes on a 12N array is expected to reduce such bridging and result in lower specificity. We therefore measured the specificity of Swi6 for the methyl mark in the context of arrays with more widely spaced nucleosomes, containing 47 base pair linker DNA (Figure S5b). As predicted, Swi6 binds to the methylated 12N(L47) substrate with lower specificity (5.4-fold) than to methylated 12N(L15) arrays (Figure 5d). These results indicate the importance of appropriate nucleosome placement for achieving high specificity.

Our finding that specificity for the methyl mark is amplified in a manner that is sensitive to internucleosomal distance is consistent with model in which the tetrameric Swi6 architecture depicted in Figure 4c enables bridging across nucleosomes. In this context, the intrinsic ability of Swi6 to form a tetrameric state (Figure 2c,d) suggests that, in addition to the CSD-CSD interface, there are other Swi6-Swi6 interfaces that promote tetramerization.

Swi6 tetramerization is mediated by the chromodomain

Because the CSD domain alone shows no oligomer formation beyond a dimer, even at concentrations where the intact Swi6 protein forms tetrameric species (Figure 2c and S2b), we used a domain deletion approach using Swi6 constructs lacking the CSD domain to identify the domain responsible for Swi6 tetramer formation. We used gel filtration to measure the extent of protein self-association for these proteins. At high concentrations, the CD alone (aa 81-137) is able to dimerize in solution, to the same extent as a Swi6 protein lacking only the CSD (NCDH) (Figure S6a). This suggested that the CD is the major component of the additional protein-protein interface. The weak CD self-association could be further stabilized by cross-

linking (Figure 6a). The CD-CD interaction also helps explain the non-saturable addition of Swi6 to nucleosomes observed by SPR and fluorescence anisotropy.

Our data agree with previous reports that the CD of human HP1 forms higher-order oligomers when cross-linked (Yamada et al., 1999). However, the interface through which such CD-CD interactions occur has not yet been identified. Given the high level of structural similarity between the CD and the evolutionary related CSD (Figure 6b), we hypothesized that the region of the CD corresponding to the sole alpha helix in the CSD that is primarily responsible for CSD dimerization might play a similar role in CD self-association. This hypothesis was further supported by analysis of the previously determined crystal structure of the dHP1 CD (Jacobs and Khorasanizadeh, 2002). The crystallographic unit of this structure contains two CD monomers that appear to engage in contacts via the alpha helix (Figure S6b). Over 30 different point mutants were made in an attempt to reduce CD dimerization, but all of these also resulted in a loss of H3K9me3 peptide binding function (data not shown). We were, however, able to obtain two gain-of-function mutants that increase CD dimerization without significantly, or not all, disrupting peptide binding: the single mutant Y131W and the double mutant V82E-Y131W (Figures 6b). The single mutant V82E replaces a Swi6 residue with a residue normally found at this location in Chp1, another chromodomain containing protein in *S. pombe* (Schalch et al., 2009).

When introduced in the full-length protein, the Y131W single and the V82E-Y131W double mutant respectively displayed ~ 1.6-fold and ~3.5-fold increased tetramer formation over WT as determined by MALS, suggesting that this region of the CD is involved in Swi6 tetramerization (Figure 6c&d and S6c). The V82E single mutant by itself does not significantly increase tetramer formation (Figure S6c). We find that the V82E single substitution increases

binding to H3K9me3 tail peptides by ~ 3-fold, consistent with previous work (Schalch et al., 2009). The double mutant V82E-Y131W however displays similar affinity for the H3K9me3 tail peptide as WT (Figure 6d).

Specificity for the methyl mark is dependent on both the CD-CD and the CSD-CSD interactions

The observation that CD-CD self association helps form Swi6 tetramers was particularly intriguing because our model, in which both H3K9 methyl marks are bound by CDs of different Swi6 dimers, places those two CDs in close proximity to self-associate (Figure 4c). We therefore hypothesized that binding in the specific orientation would strongly favor Swi6 tetramerization via CD-CD self-association and conversely, Swi6 tetramerization via self-association of two CDs would strongly favor binding in the specific orientation. If so, any disruption of the tetramer architecture depicted in Figure 4c would reduce specific recognition of the H3K9me3 mark on the nucleosome, while any strengthening of the specific architecture would increase specificity for the H3K9me3 mark.

To test these predictions, we measured specificity towards H3Kc9me3 core nucleosomes for WT Swi6 and for the L315D and V82E-Y131W mutants (Figures 6e & S6c). The L315D mutation, which significantly decreases higher-order oligomerization by disrupting CSD self-association (Figures 2c and S2a), displays 2.5-fold reduced specificity for methylated core nucleosomes relative to WT Swi6 (Figure 6e). Conversely, the V82E-Y131W double mutant, which increases tetramer formation 3.5-fold in solution by increasing CD self-association (Figure 6c&d), displays 2-fold increased specificity for methylated core nucleosomes (Figure 6e). Interestingly, both the L315D and the V82E-Y131W mutant proteins bind the H3K9me3 tail peptide with specificities comparable to the WT protein (Figure 6d). The observation that the

mutations alter methyl mark discrimination only in the nucleosomal context suggests that the effects are a result of altered oligomerization states. These results indicate that specific recognition of the nucleosomal H3K9me3 mark by Swi6 is dependent on both CSD-mediated dimerization and CD-mediated tetramerization on the nucleosome surface.

A specific CD-CD interface implies that the sticky ends that bridge nearby nucleosomes would entail CD-CD interactions in addition to interactions between the unoccupied CD and a nearby methyl mark. Therefore, in the context of nucleosomal arrays, the CD-CD interaction would further promote the H3K9 specific orientations via bridging interactions with nearby nucleosomes (see also Supplementary discussion). Such a model then makes two key predictions: (i) strengthening the CD-CD interaction would increase the specificity on nucleosomal arrays to a greater extent than on mononucleosomes, and (ii) any amplification of specificity would be very sensitive to the inter-nucleosomal distance. To test these predictions, we compared the specificity of the V82E-Y131W mutant to that of WT Swi6 on the 12N(L15) and 12N(L47) nucleosomal arrays.

As predicted by the model, we found that the V82E-Y131W mutant shows a large increase in specificity (~7-fold) on the 12N(L15) arrays compared to WT Swi6 (Figure 6e). Interestingly, this raises the specificity for the H3K9me3 mark to the ~130 fold observed on H3 tail peptides (Figure 3c). Further, most of the observed gain in specificity arises from a large decrease in binding to the H3K9 array and a small increase in binding to the H3Kc9me3 array (Figure S6d). These results suggest that the combination of strengthening the CD-CD interface and binding across multiple nucleosomes eliminates most of the alternative binding modes adopted by Swi6. No significant amplification of specificity is observed in the context of the 12N

(L47), confirming that the CD-CD nucleosome bridging interaction is sensitive to inter-nucleosomal distance (Figure 6e).

To further investigate the role of Swi6 oligomerization in the context of nucleosomal arrays, we tested the effects of disrupting the CSD-CSD interface, which is also expected to disrupt the ability of Swi6 dimers to bridge across nucleosomes (Figure 6e). The L315D mutant shows greatly reduced specificity on the 12N(L15) template. Intriguingly, the L315D mutant discriminates between methylated and unmethylated 12N(L15) arrays to a similar degree (2-3 fold) as in the context of 12N(L47) arrays and mononucleosomes (Figure 6e). Therefore, the L315D Swi6 mutant is insensitive to the distance between nucleosomes. The L315D mutant thus uncovers the baseline ability of HP1 proteins to recognize a nucleosomal H3K9 methyl mark in the absence of significant oligomerization and nucleosome bridging-dependent effects.

Increased tetramerization of Swi6 results in increased silencing at an artificial heterochromatic locus and higher recruitment to centromeres

To test whether these biochemically derived mechanistic conclusions are relevant to the ability of Swi6 to form functional heterochromatin *in vivo*, we investigated whether strengthening the CD-CD interface via the V82E-Y131W double mutant causes enhanced silencing and Swi6 occupancy *in vivo*.

To test for such an effect, we utilized a reporter system that measures silencing of the *ura4+* gene at its endogenous location on Chromosome 3 (S.S., K. Finn, H.D.M., unpublished). In this reporter construct, a centromeric fragment, under control of a promoter, is inserted 1.8 kb downstream of the *ura4+* gene (Figure 7a). We chose a 1.7 kb fragment (Fragment A; Fr A) from a library of fragments derived from the centromeric *dh* repeats. Fr A shows very weak silencing of the *ura4+* gene, leading to minimal growth of cells on 5-FOA, which provides a sensitive assay

for mutants that enhance silencing (Figure 7b). We introduced the *swi6*^{V82E,Y131W} allele by chromosomal integration into strains containing Fr A. As a control, we constructed isogenic *swi6+* strains. To control for strain-to-strain variability, we isolated and characterized 6 independent genetic isolates for both *swi6*^{V82E,Y131W} and *swi6+* alleles in the Fr A background. As shown in Figure 7b for two independent strains, *swi6*^{V82E,Y131W} increased Fr A-dependent silencing of *ura4+* (compare rows 6 and 8 to rows 5 and 7). A side by side comparison of all 6 independent isolates of the *swi6+* and *swi6*^{V82E,Y131W} alleles further confirms stronger growth on 5-FOA for all the *swi6*^{V82E,Y131W} strains (Fig. S7). The mutant Swi6 protein is not expressed at a higher level than the WT Swi6 protein, ruling out a trivial explanation for the gain of silencing effects (Figure 7c).

We next probed the molecular features of the silenced region using ChIP. We first examined Swi6 localization across the Fr A cassette locus, and found a reproducible 2-3 fold increase in Swi6 enrichment in the *swi6*^{V82E,Y131W} alleles versus the *swi6+* alleles (Figure 7d), consistent with the increased specific binding observed on nucleosomal arrays *in vitro.* However, the overall enrichment was low, probably reflecting the low degree of silencing at this artificial locus. Next, we examined H3K9me2 levels at and around the Fr A locus. Since Fr A-dependent Swi6 localization spreads beyond Fr A into adjacent euchromatic regions at the unbounded 5' end (Figure 7d), H3K9 methylation may also exhibit some Swi6-dependent spread (Hall et al., 2002). Indeed, we found that H3K9me2 levels are robustly increased in the *swi6*^{V82E,Y131W} alleles, and remain elevated at regions well outside (~20 kb) the Fr A initiating element (Figure 7e). The fact that H3K9me2 enrichment can be observed outside the zone of detectable Swi6 enrichment is likely due to the differential sensitivity of the two ChIP experiments. Increased localization of Swi6 in the context of the V82E-Y131W mutation, concomitant with robustly

increased H3K9me2 levels and elevated *ura4+* silencing, suggests that increasing the oligomerization capacity of the Swi6 protein enhances the ability of Swi6 to establish and spread heterochromatin at the artificial locus.

Next we asked whether the V82E-Y131W mutation has an effect on Swi6 activity at endogenous heterochromatin loci. We examined recruitment of Swi6 by ChIP at the *dg* repeat of centromere 1. Since H3K9 methylation at the centromere is Swi6-independent (Nakayama et al., 2001) examining centromeric heterochromatin should allow us to uncouple Swi6 recruitment from deposition of H3K9 methylation. In such a situation, changes in Swi6 recruitment should directly report on the ability of the protein to recognize H3K9 methylated chromatin *in vivo*. Indeed, when we examined H3K9me2 methylation at the *dg* repeat in the no Fr A control, *swi6+* and *swi6*^{V82E,Y131W} Fr A-containing alleles, we found no change in the enrichment level of H3K9me2 at the *dg* repeat (Figure 7f, top panel). In contrast, when we tested for Swi6 recruitment, we found a small but reproducible increase of Swi6 residence at the *dg* repeat only in the context of the Swi6 V82E-Y131W mutant (Figure 7f, bottom panel). This result suggests that when Swi6 oligomerization is increased, Swi6 recruitment is increased at endogenous heterochromatin loci where H3K9 methylation is Swi6-independent. These data help strengthen our model that CD-mediated oligomerization is critical for Swi6-dependent heterochromatin formation.

Discussion

To understand how HP1 proteins assemble on physiological chromatin templates, we reconstituted and characterized the core HP1-H3K9me3 chromatin complex. Our data suggest that recognition of H3K9me3 by HP1 proteins is coupled to its oligomerization on the nucleosome through a chromodomain-chromodomain interface that promotes silencing *in vivo*.

The mechanistic implications of these observations are discussed below.

A new role for the HP1 chromodomain in heterochromatin formation

Swi6 mutants that have increased tetramerization mediated via the CD-CD interaction exhibit increased specificity for H3K9me3 nucleosomes, suggesting that interactions between two CDs on a nucleosome restricts the number of non-H3K9me3 specific binding modes. Mutants that increase tetramerization, and thus H3K9me3 specificity *in vitro*, also exhibit increased heterochromatin spread and silencing at an artificially induced heterochromatic locus *in vivo*. The CD of HP1 proteins was previously known to recognize peptides containing methylated H3K9 (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). Our work suggests that the CD has an additional critical role in the context of chromatin: orienting HP1 proteins via CD-CD interactions to ensure that HP1 proteins can distinguish the methyl mark from other overlapping binding surfaces presented by a nucleosome.

Sticky end chromodomains in the Swi6 tetramer-nucleosome complex present polymerizable surfaces for higher order oligomerization across chromatin

Our data imply that H3K9me3 recognition and chromatin coating by Swi6 are mechanistically coupled and intrinsic to the fundamental architecture of the tetrameric HP1/Swi6 complex on the nucleosome (Figure 6f) as follows: i) Dimerization via the strong CSD-CSD interaction and tetramerization via the weaker CD-CD interaction couples recognition of the two methyl marks in a nucleosome to the generation of two unoccupied CDs. These unoccupied CDs can serve as sticky ends that bridge and recruit neighboring methylated nucleosomes, which might be either adjacent or located on different chromatin fibers. ii) Interactions between the CD of a chromatin bound HP1 dimer and that of an incoming HP1

dimer can promote deposition of the incoming HP1 dimer in an H3K9me3-recognizing orientation.

This ability to bridge nucleosomes via polymerizable CDs may represent the primary underlying mechanism that allows HP1 proteins to spread (Hall et al., 2002) along the chromatin fiber and establish the extent of the heterochromatic domain. Further, since this mechanism is dependent on a high density of H3K9me3 methylation on chromatin (Nakayama et al., 2001; Noma et al., 2001), it may enable Swi6 to sense regions of high local Clr4 methylase activity, preventing ectopic heterochromatin formation. Coupling between oligomerization and recognition of H3K9methyl marks has also been proposed in the context of the vertebrate chromodomain containing protein CDYL1b (Franz et al., 2009).

Chromatin architecture and implications for heterochromatin spread

The bridging architecture depicted in Figure 6f places specific steric and distance constraints on any Swi6-mediated heterochromatin assembly and spread, restricting the number of chromatin architectures accessible to heterochromatin assembly by Swi6. In fact, we find that Swi6 gains specificity on nucleosome arrays over unlinked nucleosomes only in the context of short DNA linkers (Figures 5c&d and 6e). We therefore hypothesize that HP1 proteins assess both the nucleosome arrangement in addition to the H3K9me3 mark, thereby integrating two signals for heterochromatin assembly.

If only a subset of chromatin architectures is permissive to template the assembly of HP1 proteins on H3K9me3 chromatin, such architectures might be regulated *in vivo* to allow specification of HP1 protein binding sites. In fact, in metazoans, the nucleosome architecture of heterochromatic loci shows significant differences compared to euchromatin sites. For example, in *Drosophila melanogaster*, constitutive heterochromatin is characterized by more evenly spaced

nucleosomes (Danzer and Wallrath, 2004; Wallrath and Elgin, 1995) compared to euchromatin. It has been suggested that the ACF chromatin remodeling complex is involved in generating such chromatin architectures in *Drosophila* (Fyodorov et al., 2004). In *S. pombe*, there is some evidence that local nucleosome arrangement in heterochromatin impacts Swi6 association. Several protein complexes collaborate in *S. pombe* to maintain heterochromatin regions. A key such effector is a bi-functional enzyme complex called SHREC, containing both the histone deacetylase (HDAC) Clr3 and the SNF2 chromatin-remodeling factor homolog Mit1 (Sugiyama et al., 2007). The Clr3 subunit of SHREC is required for Swi6 localization, in a manner that appears uncoupled from its effects on H3K9 methylation (Nakayama et al., 2001; Yamada et al., 2005); S.I.S. Grewal, personal communication). This effect may result either from 1) the absence of Clr3's HDAC activity, resulting in an increase in acetylated and phosphorylated histones that may affect Swi6's ability to associate with those nucleosomes (Yamada et al., 2005), or 2) effects of the SHREC complex on nucleosome arrangement (Sugiyama et al., 2007). We speculate that SHREC and/or other chromatin regulators may promote a nucleosome arrangement that enables Swi6 to bridge H3K9me3-marked nucleosomes and therefore to spread. Further work will be needed to identify what exact chromatin architectures are compatible with Swi6 bridging and how such structures may be generated and maintained *in vivo*.

Experimental Procedures

(Detailed Methods are described in "Extended Experimental Procedures" in Supplementary Information)

Protein cloning and purification

Full length Swi6 was cloned into pET30a (Novagen), mutants made using site directed mutagenesis and proteins purified from *E. coli*. Tagged Swi6 containing N-terminal 6xHis and C-terminal FLAG tags was used for the MALS, AUC, cross-linking, native gel-shift, nucleosome competition and peptide binding assays. Untagged Swi6 was used for the SPR-based and polarization-based nucleosome binding measurements. Removing the tags slightly increases overall affinity for nucleosomes and arrays (~2.5 fold) but does not affect specificity (data not shown).

Mononucleosome, dinucleosome, and array reconstitution

Gradient salt dialysis was used to assemble mononucleosomes on DNA templates containing the 147 bp long 601 positioning sequence, dinucleosomes on DNA templates containing two 601 sequences linked by 15 or 47 bp of DNA and arrays on DNA templates containing 12 copies of the 601 nucleosome positioning sequence separated by either 15 or 47 bp linkers. H3Kc9me3, H3Kc9me0 and H3K9 histones were prepared as described(Luger et al., 1999; Simon et al., 2007). Arrays with >95% assembly were used for gel shift assays.

Surface plasmon resonance

A BiaCore T100 instrument was used for SPR analysis of Swi6 interaction with the mononucleosome substrates. H3K9 and H3Kc9me3 substrates assembled on a 5' biotinylated 601 sequence were immobilized to 25 to 60 RU on active flow cells . Immobilization levels of mononucleosomes ranged from 25 to 60 RU. Surface stability and assay quality were judged by

the reproducibility of a 10 µL control sample (100 nM Swi6) injection that followed each sample concentration (Figure S1d).

Fluorescence polarization binding measurements

All H3 tail peptides were produced as described (Simon et al., 2007). Nucleosomes for polarization-based binding measurements were assembled on a 6-carboxyfluorescein-labeled 601 positioning sequence.

Native gel mobility shift assays

Different concentrations of Swi6 protein were incubated with 5-10 nM mononucleosome or 1.25 nM dinucleosome. Samples were run on native polyacrylamide gels, stained with SyBR Gold (Invitrogen), visualized on a Molecular Dynamics Typhoon scanner and quantified using Image Quant software. The $K_{1/2}$ for each binding curve and Hill coefficient were calculated with Kaleidograph software using a simple equilibrium model. Swi6 gel shifts with 12N arrays were performed using agarose gels with 1 nM array (12 nM nucleosomes) and analyzed as described for mononucleosomes.

Protein cross-linking

Cross-linking assays were performed using EDC/NHS cross-linking (Pierce) (see Extended Experimental Procedures). The samples were boiled and analyzed on 4-12% NuPAGE gradient gels (Invitrogen) under denaturing conditions, then visualized on a Typhoon scanner by Sypro Red staining or by anti-FLAG western blotting.

Isothermal titration calorimetry

The heat released by dissociation of CSD dimers into monomers was measured with a Microcal, Inc., Omega microcalorimeter. Dilution ITC experiments involved sequential injections

from a concentrated protein stock (5 μM for WT Swi6 and 625 μM for L315D Swi6) in 5μl increments into the 1.4 ml calorimeter cell initially containing only buffer.

Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS/UV/RI)

Protein samples were injected into an analytical size exclusion silica gel KW804 chromatography column (Shodex). The chromatography system was coupled to an 18-angle light scattering detector (DAWN HELEOS II, Wyatt Technology) and a differential refractometer (Optilab rEX, Wyatt Technology).

Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity experiments were conducted using an analytical ultracentrifuge (Beckman Coulter) equipped with an absorption optical scanner. The binding reaction was set up such that a) both nucleosome and Swi6 concentrations were above the $K_{1/2}$ value measured by native gel and b) the Swi6 concentration was sufficient to titrate all the nucleosomes as assayed by native gel shift.

Data were analyzed using the Sedfit software (Schuck, 2004). Three independent analyses: c(s), c(s,fr), and c(s,bi-modal fr) were used to study the sedimentation properties and the molar mass of each sample. Solution density (ρ), solution viscosity (η) were calculated in SEDNTERP (Schuck, 2004).

In vivo silencing assays

A fragment of the *dh* centromeric repeat was placed 1.8 kb downstream of the *ura4+* gene on chromosome III using homologous recombination. Transcription of this fragment is driven by the p*adh1+* promoter and is sufficient to induce silencing of *ura4+* in a manner that requires *clr4+* and *dcr1+* (S.S., K. Finn and H.D.M., unpublished). Silencing of *ura4+* gene was

assayed by growth on 2 mg/mL 5-fluoroorotic acid (5-FOA). The endogenous *swi6+* was replaced with either *swi6*^{V82E,Y131W} marked with a 5' *G418*^{R} selectable marker, or the wild-type allele and the same marker.

Chromatin immunoprecipitation (ChIP)

ChIP experiments with *Fr A*$^{-}$, *swi6+* and *swi6*^{V82E,Y131W} strains were performed using anti-H3K9me2 (Abcam) or anti-Swi6 (Nakayama et al., 2000) antisera. Details see Extended Experimental procedures.

References

Cowieson, N.P., Partridge, J.F., Allshire, R.C., and McLaughlin, P.J. (2000). Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Curr Biol *10*, 517-525.

Danzer, J.R., and Wallrath, L.L. (2004). Mechanisms of HP1-mediated gene silencing in Drosophila. Development *131*, 3571-3580.

Dawson, M.A., Bannister, A.J., Gottgens, B., Foster, S.D., Bartke, T., Green, A.R., and Kouzarides, T. (2009). JAK2 phosphorylates histone H3Y41 and excludes HP1 alpha from chromatin. Nature *461*, 819-U879.

Fischer, T., Cui, B., Dhakshnamoorthy, J., Zhou, M., Rubin, C., Zofall, M., Veenstra, T.D., and Grewal, S.I.S. (2009). Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. Proc Natl Acad Sci USA *106*, 8998-9003.

Franz, H., Mosch, K., Soeroes, S., Urlaub, H., and Fischle, W. (2009). Multimerization and H3K9me3 binding are required for CDYL1b heterochromatin association. J Biol Chem *284*, 35049-35059.

Fyodorov, D.V., Blower, M.D., Karpen, G.H., and Kadonaga, J.T. (2004). Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. Genes Dev *18*, 170-183.

Godde, J.S., and Widom, J. (1992). CHROMATIN STRUCTURE OF SCHIZOSACCHAROMYCES-POMBE - A NUCLEOSOME REPEAT LENGTH THAT IS SHORTER THAN THE CHROMATOSOMAL DNA LENGTH. Journal of Molecular Biology *226*, 1009-1025.

Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet *8*, 35-46.

Hall, I.M., Shankaranarayana, G.D., Noma, K.I., Ayoub, N., Cohen, A., and Grewal, S.I.S. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Iida, T., Nakayama, J.-i., and Moazed, D. (2008). siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. Mol Cell *31*, 178-189.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Lantermann, A.B., Straub, T., Stralfors, A., Yuan, G.C., Ekwall, K., and Korber, P. (2010). Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nature Structural & Molecular Biology *17*, 251-U215.

Lavigne, M., Eskeland, R., Azebi, S., Saint-Andre, V., Jang, S.M., Batsche, E., Fan, H.Y., Kingston, R.E., Imhof, A., and Muchardt, C. (2009). Interaction of HP1 and Brg1/Brm with the globular domain of histone H3 is required for HP1-mediated repression. PLoS Genet *5*, e1000769.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. In Chromatin (San Diego, Academic Press Inc), pp. 3-19.

Meehan, R.R., Kao, C.F., and Pennings, S. (2003). HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. Embo Journal *22*, 3164-3174.

Nakayama, J., Klar, A.J., and Grewal, S.I. (2000). A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. Cell *101*, 307-317.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103-107.

Noma, K., Allis, C.D., and Grewal, S.I.S. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science *293*, 1150-1155.

Sadaie, M., Kawaguchi, R., Ohtani, Y., Arisaka, F., Tanaka, K., Shirahige, K., and Nakayama, J. (2008). Balance between Distinct HP1 Family Proteins Controls Heterochromatin Assembly in Fission Yeast. Molecular and Cellular Biology *28*, 6973-6988.

Schalch, T., Job, G., Noffsinger, V.J., Shanker, S., Kuscu, C., Joshua-Tor, L., and Partridge, J.F. (2009). High-affinity binding of Chp1 chromodomain to K9 methylated histone H3 is required to establish centromeric heterochromatin. Mol Cell *34*, 36-46.

Schuck, P. (2004). A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. Biophysical Chemistry *108*, 187-200.

Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. Cell *128*, 1003-1012.

Smothers, J.F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr Biol *10*, 27-30.

Sugiyama, T., Cam, H.P., Sugiyama, R., Noma, K.-i., Zofall, M., Kobayashi, R., and Grewal, S.I.S. (2007). SHREC, an effector complex for heterochromatic transcriptional silencing. Cell *128*, 491-504.

Wallrath, L.L., and Elgin, S.C.R. (1995). Position Effect Variegation in Drosophila Is Associated with an Altered Chromatin Structure. Genes & Development *9*, 1263-1277.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D., and Grewal, S.I.S. (2005). The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol Cell *20*, 173-185.

Yamada, T., Fukuda, R., Himeno, M., and Sugimoto, K. (1999). Functional domain structure of human heterochromatin protein HP1 (Hsalpha): involvement of internal DNA-binding and C-terminal self-association domains in the formation of discrete dots in interphase nuclei. J Biochem *125,* 832-837.

Zhao, T., Heyduk, T., Allis, C.D., and Eissenberg, J.C. (2000). Heterochromatin protein 1 binds to nucleosomes and DNA in vitro. J Biol Chem *275*, 28332-28338.

Figure Legends:

Figure 1: Swi6 recognizes the H3K9 methyl mark within mononucleosomes and forms oligomers on mononucleosomes

(A) Schematics of the unmodified K9 and methyl lysine analog (MLA) Kc9me0 and Kc9me3 substrates (top panel) and of the unmodified (H3K9) and MLA methylated (H3Kc9me3) mononucleosomes assembled on the 147 bp 601 sequence (bottom).

(B) Bottom Panels: Representative dose responses for H3K9 (black) and H3Kc9me3 (red) mononucleosomes. Schematic: H3K9 and H3Kc9me3 mononucleosomes captured on a streptavidin derivatized SPR chip. Top panels: Close up of the kinetics of association and dissociation.

(C) Scaled isotherms for three independent dose responses of Swi6 against H3K9 (open diamonds) and H3Kc9me3 (filled circles) mononucleosome surfaces plotted on a semi-log scale. Plotted points represent the response at equilibrium, determined by averaging the signal over the final ten seconds of the sample injection. Inset: isotherms plotted on a linear scale.

(D) Schematics: Mononucleosomes with fluorescein (green star) attached by a flexible linker at one end of the 147 bp DNA template. Average of three independent fluorescent polarization experiments for H3K9 (open diamonds) and H3Kc9me3 (filled circles) mononucleosomes are shown. Error bars represent s.e.m. All Swi6 concentrations represent monomer concentrations.

Figure 2: Swi6 forms distinct oligomeric states in the absence of chromatin

(A) Wild-type Swi6 (schematic on top) is largely a pre-formed dimer at low nM concentrations. Swi6 WT (left) and L315D (right) were treated at indicated concentrations with EDC and NHS cross-linkers. Treated proteins were separated by SDS-PAGE and detected by anti-FLAG western. Swi6 concentrations: uncross-linked 50 nM, cross-linked 25, 50, 100, 200, 500, 1000 and 5000 nM.

(B) The CSD-CSD dimerization K_d < 17 nM. Top: Representative ITC thermogram profiles for the dissociation of WT CSD dimer (left) and L315D CSD dimer (right) at 15°C. Bottom: Graphs represent the respective binding isotherms plotted as heat changes per injection (q_i) vs. total monomer concentration.

(C) Multiangle light scattering (MALS) measurements for 20 µM WT Swi6 (blue) and 20 µM L315D mutant (red). Relative refractive index signals (solid lines, left y-axis) and derived molar masses (dotted lines, right y-axis) shown as a function of the elution volume. M: monomer, D: dimer, T: tetramer. (D) Top panel: Higher order oligomeric species of Swi6 stabilized by cross-linking. MALS measurements conducted as in (A). M: monomer, D: dimer, T: tetramer, O: octamer. Bottom panel: Aliquots of fractions collected from chromatography in (Top) were separated by denaturing SDS-PAGE and visualized by Sypro Red staining. The distribution of distinct oligomeric states thus visualized directly correlates with the oligomeric masses observed by MALS, while the presence of un-cross-linked Swi6 demonstrates Swi6 is not over cross-linked. All Swi6 concentrations represent monomer concentrations.

Figure 3: Swi6 displays lower specificity for the H3K9me3 mark in mononucleosomes compared to that in H3 tail peptides

(A) Representative gel shift using H3K9 or H3Kc9me3 mononucleosomes. Swi6 concentrations vary from 0 to 12 μM (0.6 fold dilutions). Unbound nucleosomes (N).

(B) Quantification of three gel shift experiments using H3K9 (open diamonds) and H3Kc9me3 (filled circles) to determine $K_{1/2}$ and specificity ($K_{1/2}$ H3K9 / $K_{1/2}$ H3Kc9me3). Hill coefficient = 1.7 (H3Kc9me3) and = 2 (H3K9).

(C) Swi6 specificities for H3Kc9me3 mononucleosome and H3K9me3 peptide. $K_{1/2}$ values (μM) for peptides were measured by fluorescence anisotropy and $K_{1/2}$ (μM) for nucleosomes are from (B) with n=3.

(D) Increasing linker DNA length (L, in "bp") decreases Swi6's ability to discriminate the methyl mark on mononucleosomes. Left graph: Swi6 discrimination for H3Kc9me3 over unmodified mononucleosomes. Right graph: Swi6 affinity for H3Kc9me3 mononucleosomes, normalized to core (L=0) nucleosomes. All error bars represent s.e.m. All Swi6 concentrations represent monomer concentrations.

Figure 4: The core unit of Swi6 binding to a mononucleosome is a tetramer

(A) Sedimentation Velocity Analytical Ultracentrifugation (SV-AUC) on H3Kc9me3 core nucleosomes. Left: The *c(M, f/f0)* distribution generated from SV–AUC experiments shown as a two-dimensional distribution. x-axis: molecular weight (Mwt); y-axis: hydrodynamic translational frictional ratio (*f/f0*). Below, the *c(M, f/f0)* surface is shown as a contour plot of the distribution projected onto the *M-f/f0* plane, where the magnitude of *c(M, f/f0)* is indicated by contour lines at constant *c(M, f/f0)* for equidistant intervals of *c*. Right: Table showing measured average masses (versus theoretically predicted masses) from three independent experiments using either a continuous two-dimensional function *c(s,f/f0)* for sedimentation coefficient *s* and hydrodynamic translational frictional ratio f/f_0, or a continuous function *c(s)* with a *bi-modal* f/f_0 distribution *c(s, bimodal f/f0)*. Errors represent s.e.m.

(B) SV-AUC on H3Kc9me3 core nucleosome with L315D Swi6. Representation and table as in (A). Red asterisk: free L315D Swi6.

(C) SV-AUC on H3Kc9me3 core nucleosome with WT Swi6. Representation and table as in (A). Blue asterisk: free WT Swi6. Black arrows represent sticky ends.

The measured masses are used to derive structural models for the stoichiometry of the complexes as shown.

Figure 5: Amplification of Swi6 specificity towards H3Kc9me3 occurs on nucleosome arrays and is sensitive to nucleosomal placement

Dinucleosome (2N) or 12-nucleosome arrays (12N) constructs contain either 15 bp (L15) or 47 bp (L47) internucleosomal linkers.

(A) Swi6 displays 2.5 fold specificity towards 2N(L15) H3Kc9me3 dinucleosomes. Representative gel shift shown. $K_{1/2}$ for H3Kc9me3 and H3K9 2N(L15) substrates are 62 and 156 nM, respectively. Specificity = $K_{1/2}$ H3Kc9me3 / $K_{1/2}$ H3K9.

(B) Swi6 displays similar specificity towards H3Kc9me3 2N(L47) as for 2N(L15) dinucleosomes. Gel shift and analysis as in (A). $K_{1/2}$ for H3Kc9me3 and H3K9 2N(L47) substrates were 12 and 32 nM, respectively.

(C) Swi6 displays ~10x amplified specificity towards H3Kc9me3 12N(L15) arrays vs. H3Kc9me3 2N(L15) dinucleosomes. Swi6-bound and unbound arrays were separated by agarose gel electrophoresis. Representative gel shift shown. $K_{1/2}$ and specificity were determined as above.

(D) Amplification of Swi6 specificity on 12N(L15) arrays is reduced on 12N(L47) arrays. Gel shift and analysis as in (C).

$K_{1/2}$ for array substrates: see Figure S6d. All error bars represent s.e.m. Swi6 concentrations represent monomer concentrations.

Figure 6: The chromodomain contains the Swi6 tetramerization interface and couples tetramerization on the nucleosome surface to H3K9me3 recognition

(A) The chromodomain of Swi6 can homodimerize. MALS measurements for uncross-linked (black) and cross-linked (green) Swi6 chromodomain (CD) showing UV absorbance signal in mA (solid lines, left y-axis) and derived molar masses (dotted lines, right y-axis) as a function of the elution volume. The CD was injected at ~50 µM. M: monomer, D:dimer. Cross-linked CD shows increased D. Inset: SDS-PAGE analysis for the uncross-linked (-) and cross-linked (+) samples used in the MALS measurements.

(B) Top: Superimposition of the structure of monomeric dHP1 CD (black, pdb 1KNE) with dimeric Swi6 CSD (light brown, pdb 1E0B) shows structural similarity between the two evolutionarily related domains. Bottom: Alignment of the CD of the three HP1-like proteins in *S. pombe* with dHP1 CD and Swi6 CSD. Yellow boxes: conserved residues V82 and Y131. Purple box: hydrophobic residue (L or I) central to CSD dimerization. Red star: CD hydrophobic cage residues required for H3K9me3 recognition. Gray: secondary structure schematic for dHP1 CD and Swi6 CSD.

(C) MALS measurements for WT Swi6 (blue) and V82E Y131W Swi6 (yellow), shows UV absorbance signal (solid lines, left y-axis) and derived molar masses (dotted lines, right y-axis) as a function of the elution volume. WT and V82E-Y131W Swi6 were injected at ~20 µM. The V82E-Y131W protein shows a higher proportion of species in tetrameric (T) and octameric (O) oligomeric states.

(D) Relationship between peptide specificity and oligomeric states (tetramer and beyond) for WT, CSD mutant (L315D) and the CD double mutant V82E-Y131W. H3K9me3 specificity for

each protein is calculated as $K_{1/2}$H3K9/$K_{1/2}$H3K9me3. All data are reported as fold differences relative to the WT protein. Errors represent s.e.m.

(E) H3K9me3 specificity is regulated by the oligomeric state of Swi6.

y-axis: Fold specificity for methylated mononucleosome (1N) and indicated 12N array substrates.

(F) A model to depict how the CSD-CSD and CD-CD interactions enable orientation of Swi6 to correctly recognize the methyl mark in a nucleosome and generate sticky ends that bridge nearby nucleosomes and further enhance specific orientations.Swi6 concentrations represent monomer concentrations.

Figure 7: Increased tetramerization of Swi6 translates into increased silencing and heterochromatin spreading at an artificial heterochromatic locus

(A) Schematic of the reporter cassette integrated downstream of endogenous *ura4+* gene. Cassette contains a promoter (P) driving the expression of a centromeric *dh* fragment (Fr A), an intergenic region from two convergent regions (t), a boundary element (B) that contains synthetic TFIIIC binding sites (known to limit the spread of heterochromatin in *S. pombe*) and a Nat drug resistance marker (Nat^R)**.**

(B) The V82E-Y131W mutant shows increased silencing of the fragment A cassette (Fr A). Serial dilutions of indicated *S. pombe* strains. Strains containing Fr A show silencing of *ura4+* and are able to grow on media containing 5-FOA. *swi6+* or *swi6*$^{VY\rightarrow EW}$ alleles were introduced into strains containing the whole cassette with or without Fr A (Fr A^-); 2 independent clones are shown for each swi6 allele. Fr A^- strains contain the entire cassette as shown in (A) but lack the centromeric fragment.

(C) The V82E-Y131W mutant expresses slightly lower levels of Swi6 than WT. Extracts from respective strains were separated on SDS-PAGE gels and probed for γ-tubulin (green) or Swi6 (red). Quantification of the Swi6 band normalized for the γ-tubulin control is shown relative to the value obtained for *swi6+* clone 1.

(D) The Swi6 V82E-Y131W mutation induces increased Swi6 recruitment to the Fr A locus. Chromatin immunoprecipitation (ChIP) with anti-Swi6 antisera was performed in *swi6+, swi6* $^{VY\rightarrow EW}$ or Fr A^- backgrounds. Fr A specific Swi6 enrichment is represented as the ratio of the actin-normalized signal at indicated amplicons in *swi6+* or *swi6*$^{VY\rightarrow EW}$ strains divided by the actin-

normalized signal in the Fr A$^-$ strain. Error bars represent s.e.m. of unicate ChIP experiments from the 6 genetic isolates of *swi6*+ or *swi6*$^{VY \rightarrow EW}$ alleles. X-axis: distance in base pairs relative to the Fr A cassette promoter (P). Genomic features near the Fr A cassette insertion site are aligned below the graph.

(E) The Swi6 V82E-Y131W mutation increases H3K9 methylation at and beyond the Fr A cassette. ChIP experiments were performed with anti-H3K9me2 antisera. The Fr A specific H3K9me2 enrichment is calculated as in (D). Error bars as in (D).

(F) The Swi6 V82E-Y131W mutation leads to increased Swi6 recruitment at endogenous heterochromatin. Top: H3K9me2 ChIP. H3K9me2 fold enrichment over actin at the centromeric *dg* repeat for the Fr A- strain and *swi6*+ or *swi6*$^{VY \rightarrow EW}$ alleles. Bottom: Swi6 ChIP. Swi6 fold enrichment over actin at the centromeric *dg* repeat for the same strains as (Top). Error bars for *swi6*+ and *swi6*$^{VY \rightarrow EW}$ as in (D). Error bar for Fr A$^-$ (bottom) represents s.e.m. for three independent IPs from the Fr A$^-$ strain.

Figure 1

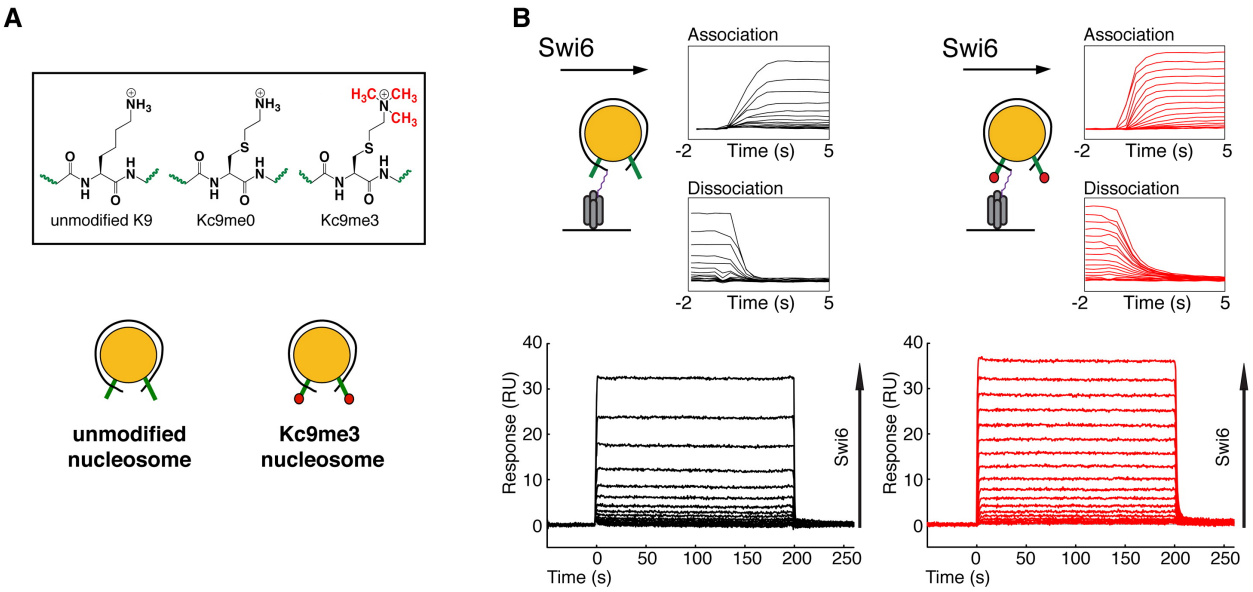

A
unmodified K9
Kc9me0
Kc9me3
unmodified nucleosome
Kc9me3 nucleosome
B
Swi6
Association
Dissociation
-2
Time (s)
5
Response (RU)
Swi6
Time (s)
C
H3Kc9me3
H3K9
Scaled Response (RU)
Swi6 (μM)
D
H3Kc9me3
H3K9
+ Swi6
+ Swi6
Polarization (mP)
Swi6 (μM)

Figure 2

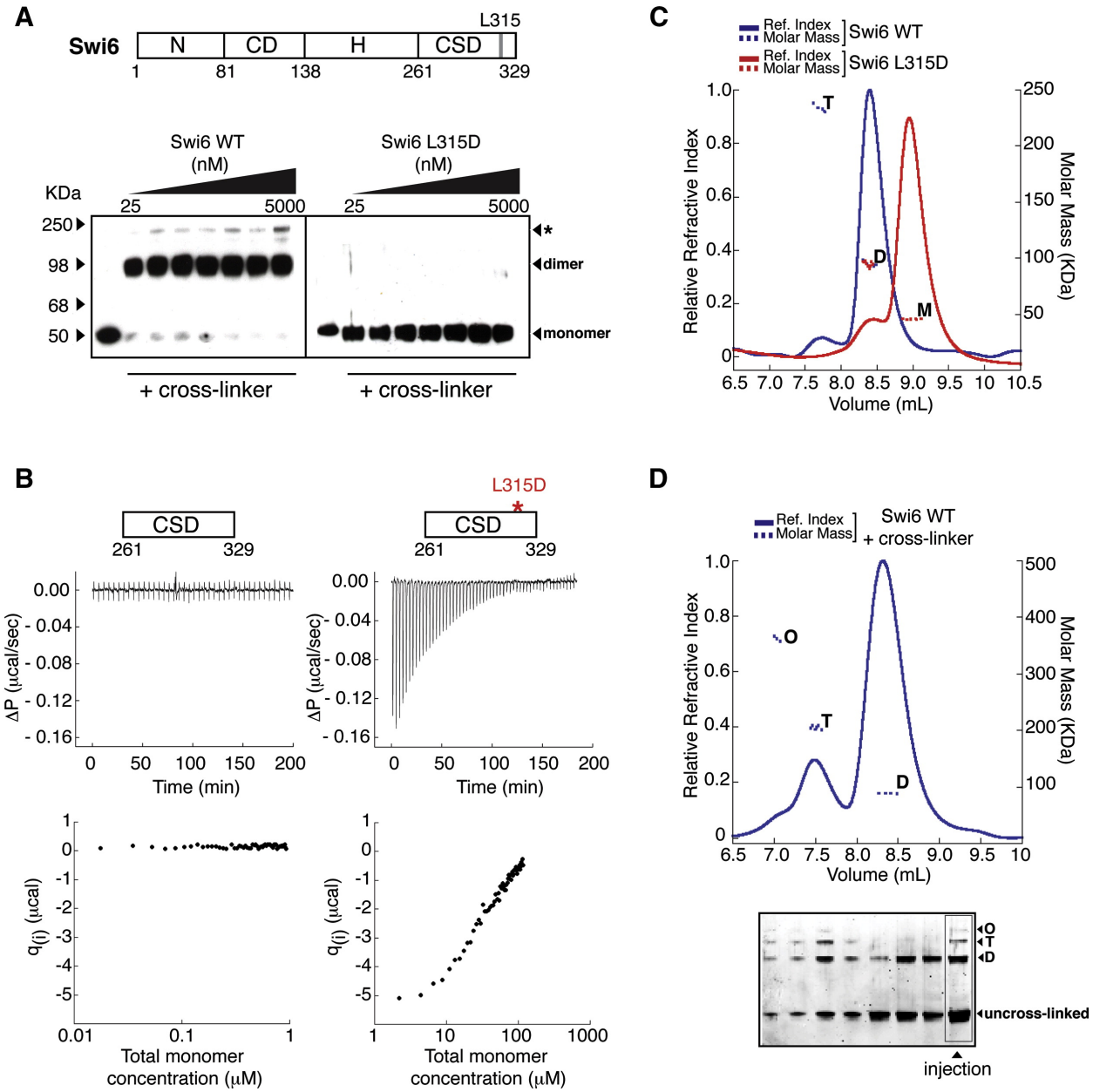
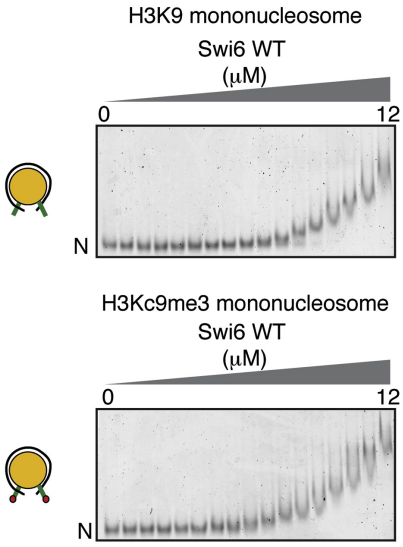

A
Swi6
N
CD
H
CSD
L315
1
81
138
261
329
Swi6 WT
(nM)
Swi6 L315D
(nM)
KDa
25
5000
25
5000
250
98
68
50
*
dimer
monomer
+ cross-linker
+ cross-linker
B
CSD
261
329
L315D
*
CSD
261
329
ΔP (μcal/sec)
Time (min)
q(i) (μcal)
Total monomer
concentration (μM)
C
Ref. Index
Molar Mass
Swi6 WT
Ref. Index
Molar Mass
Swi6 L315D
T
D
M
Relative Refractive Index
Molar Mass (KDa)
Volume (mL)
D
Ref. Index
Molar Mass
Swi6 WT
+ cross-linker
O
T
D
Relative Refractive Index
Molar Mass (KDa)
Volume (mL)
O
T
D
uncross-linked
injection

Figure 3

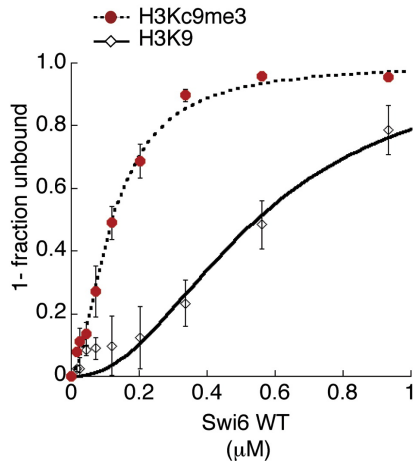
A



C

peptide		core nucleosome	
affinity ($K_{1/2}$, μM)		*affinity ($K_{1/2}$, μM)*	
H3K9	H3K9me3	H3K9	H3Kc9me3
170 ± 5.3	1.29 ± 0.03	0.56 ± 0.08	0.12 ± 0.01
specificity		*specificity*	
132.1 ± 6.8		**4.6 ± 0.6**	

B



D

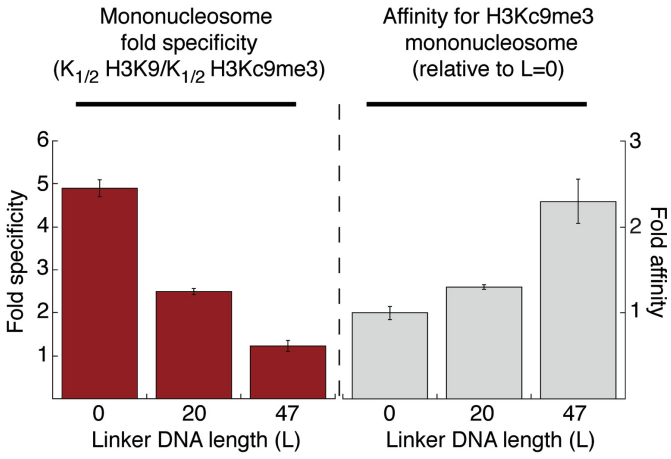
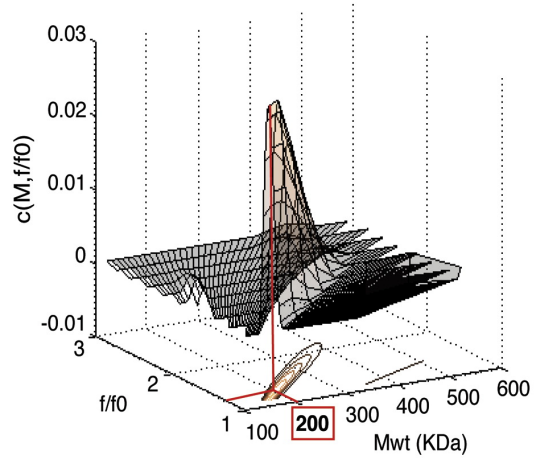


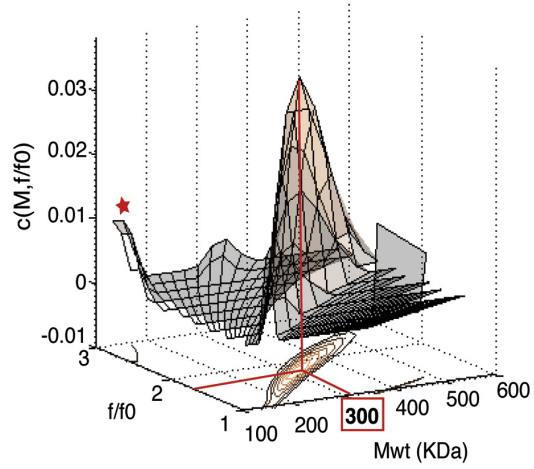
Figure 4

A H3Kc9me3 mononuclesome

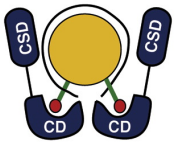


Predicted mwt (KDa)	Measured mwt (KDa)	
	c(s,f/f0)	c(s,bimodal f/f0)
197	204 ± 10	199 ± 9.2

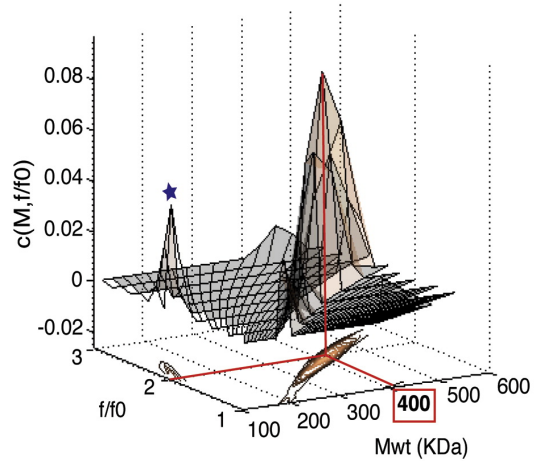
B H3Kc9me3 mononuclesome + Swi6 L315D



Predicted mwt (KDa)	Measured mwt (KDa)	
	c(s,f/f0)	c(s,bimodal f/f0)
285	312 ± 8.7	306 ± 7.5



C H3Kc9me3 mononuclesome + Swi6 WT



Predicted mwt (KDa)	Measured mwt (KDa)	
	c(s,f/f0)	c(s,bimodal f/f0)
370	367 ± 14	354 ± 3.4

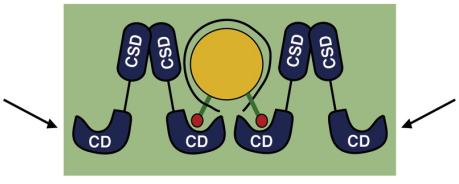


Figure 5

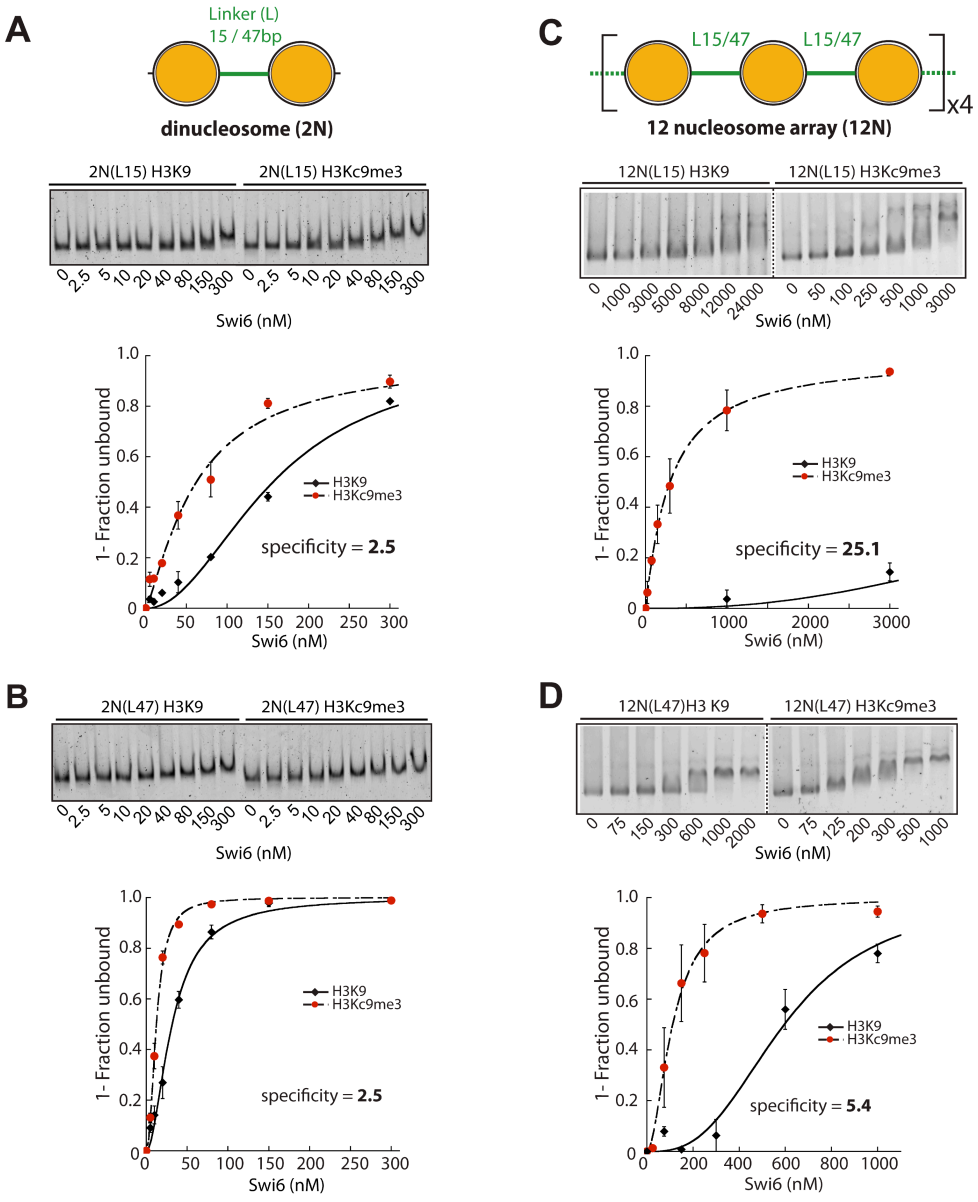

A
Linker (L)
15 / 47bp
dinucleosome (2N)
2N(L15) H3K9
2N(L15) H3Kc9me3
0 2.5 5 10 20 40 80 150 300 0 2.5 5 10 20 40 80 150 300
Swi6 (nM)
1- Fraction unbound
H3K9
H3Kc9me3
specificity = 2.5
Swi6 (nM)
C
L15/47
L15/47
x4
12 nucleosome array (12N)
12N(L15) H3K9
12N(L15) H3Kc9me3
0 1000 3000 5000 8000 12000 24000 0 50 100 250 500 1000 3000
Swi6 (nM)
1- Fraction unbound
H3K9
H3Kc9me3
specificity = 25.1
Swi6 (nM)
B
2N(L47) H3K9
2N(L47) H3Kc9me3
0 2.5 5 10 20 40 80 150 300 0 2.5 5 10 20 40 80 150 300
Swi6 (nM)
1- Fraction unbound
H3K9
H3Kc9me3
specificity = 2.5
Swi6 (nM)
D
12N(L47)H3 K9
12N(L47) H3Kc9me3
0 75 150 300 600 1000 2000 0 75 125 200 300 500 1000
Swi6 (nM)
1- Fraction unbound
H3K9
H3Kc9me3
specificity = 5.4
Swi6 (nM)

Figure 6

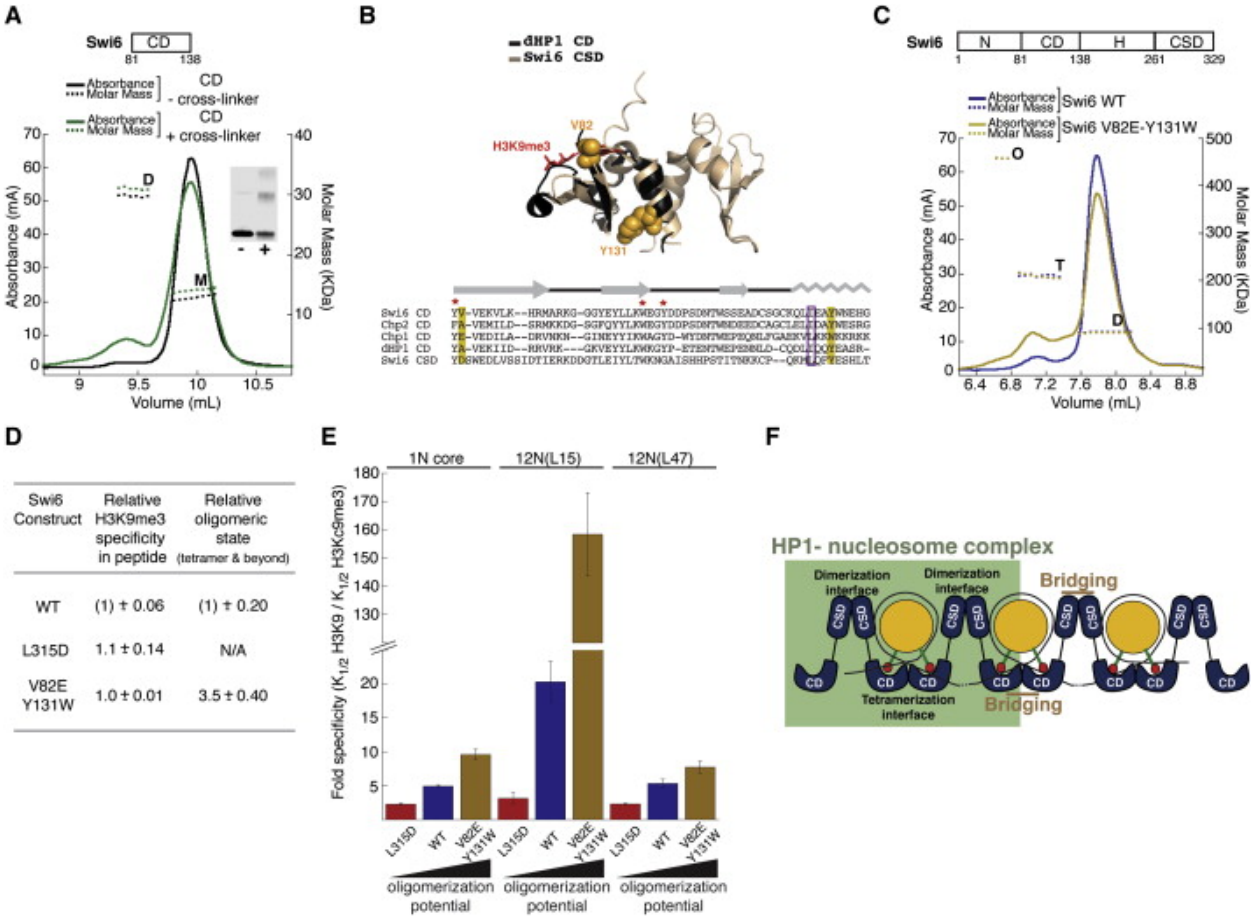
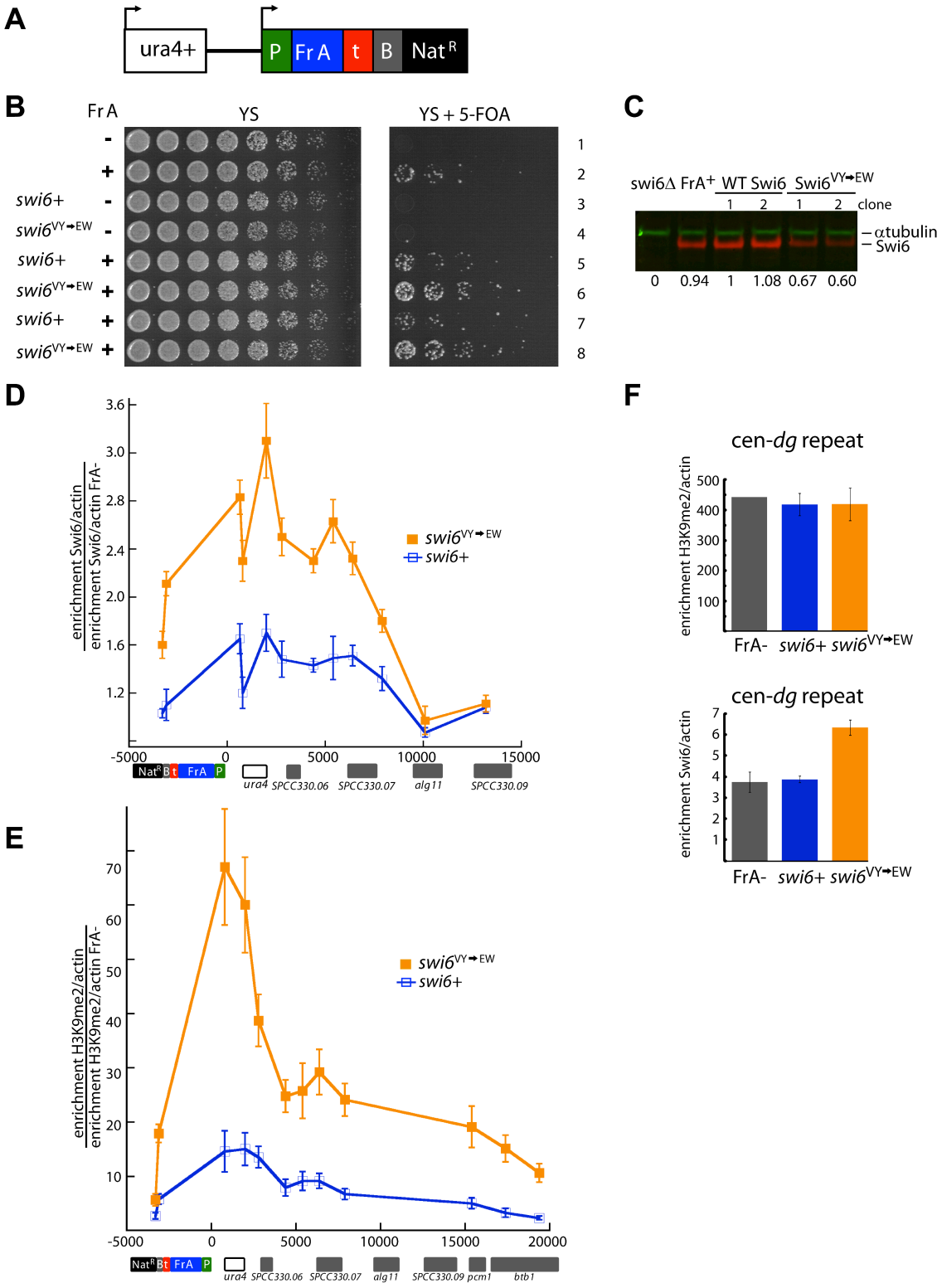

A
Swi6
CD
81
138
Absorbance
Molar Mass
CD
- cross-linker
CD
+ cross-linker
D
M
Absorbance (mA)
Molar Mass (kDa)
Volume (mL)
B
dHP1 CD
Swi6 CSD
V82
H3K9me3
Y131
C
Swi6
N
CD
H
CSD
1
81
138
261
329
Swi6 WT
Swi6 V82E-Y131W
O
T
D
Absorbance (mA)
Molar Mass (KDa)
Volume (mL)
D
Swi6 Construct
Relative H3K9me3 specificity in peptide
Relative oligomeric state (tetramer & beyond)
WT
(1) ± 0.06
(1) ± 0.20
L315D
1.1 ± 0.14
N/A
V82E Y131W
1.0 ± 0.01
3.5 ± 0.40
E
1N core
12N(L15)
12N(L47)
Fold specificity (K1/2 H3K9 / K1/2 H3Kc9me3)
L315D
WT
V82E Y131W
oligomerization potential
F
HP1- nucleosome complex
Dimerization interface
Dimerization interface
Bridging
CSD
CD
Tetramerization interface
Bridging

Figure 7


A
ura4+
P
FrA
t
B
Nat R
B
FrA
YS
YS + 5-FOA
swi6+
swi6VY→EW
C
swi6Δ FrA+
WT Swi6
Swi6VY→EW
clone
αtubulin
Swi6
0 0.94 1 1.08 0.67 0.60
D
enrichment Swi6/actin
enrichment Swi6/actin FrA-
swi6VY→EW
swi6+
ura4
SPCC330.06
SPCC330.07
alg11
SPCC330.09
E
enrichment H3K9me2/actin
enrichment H3K9me2/actin FrA-
pcm1
btb1
F
cen-dg repeat
enrichment H3K9me2/actin
FrA-
swi6+
swi6VY→EW
enrichment Swi6/actin

Acknowledgements

We thank Shiv I.S. Grewal for a generous gift of anti-Swi6 antisera, Christine Rumpf for help with ITC measurements, Dan Southworth and Janet Yang for assistance with MALS set-up, Kalyan Sinha and Peter Schuck for critical advice and comments on the AUC experiments and data interpretation and Robert Fletterick for critical advice and comments on the SPR experiments and data interpretation. We thank Barbara Panning, Jonathan Weissman, Karim-Jean Armache and members of the Narlikar Lab for helpful discussion and comments on the manuscript. This work was supported by grants from the National Institutes of Health (5R01GM071801 to H.D.M. and 1R01GM073767 to G.J.N.), the UCSF Program for Breakthrough Biomedical Research (PBBR) Award to G.J.N, and the Beckman Foundation (to G.J.N.). G.J.N. and H.D.M. are Scholars of the Leukemia and Lymphoma Society. D.C. is a Genentech, Inc. Predoctoral Fellow. E.C. is a ARCS Foundation, Inc. Fellow. S.S. is a fellow of the Leukemia and Lymphoma Society. K.M.K is supported by a training grant to the Biophysics Graduate Group (NIH T32GM008284).M.D.S. is a fellow of the Helen Hay Whitney foundation. B.A. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

Supplemental Data

Supplementary Figures **S1-S7**

Supplementary Information

Chromodomain-mediated oligomerization of HP1 suggests a nucleosome bridging mechanism for heterochromatin assembly

Daniele Canzio, Evelyn Y. Chang, Smita Shankar, Kristopher M. Kuchenbecker, Matthew D. Simon, Hiten D. Madhani, Geeta J. Narlikar and Bassem Al-Sady

Inventory of Supplementary Information

Figure S1 (related to Figure 1)

Raw data and analysis of the surface plasmon resonance assay

Figure S2 (related to Figure 2)

Loss of higher-order oligomerization in L315D Swi6 mutant

Figure S3 (related to Figure 3)

Comparison of H3K9 and H3Kc9me0 nucleosome:Swi6 affinities by gel mobility shift assay. Quantification of H3K9 and H3K9me3 tail peptide:Swi6 interactions by fluorescence polarization. Quantification of gel mobility shift assays of Swi6 with mononucleosomes containing varying linker DNAs on both sides of 147bp 601 sequence and Swi6 with DNA constructs of varying length. Estimation of Swi6 binding specificity for H3Kc9me3 over H3K9 core nucleosomes using a DNA - nucleosome competition assay

Figure S4 (related to Figure 4)

Details of two analysis methods for mononucleosome:Swi6 SV-AUC data

Figure S5 (related to Figure 5)

Quality control for arrays used in Figure 5. Gel mobility shift assay based co-operativity estimates for Swi6:H3Kc9me3 12N array interactions. Sensitivity analysis for gel mobility shift assays in Figure 5c.

Figure S6 (related to Figure 6)

Swi6 deletion analysis. Analysis of dHP1 chromodomain crystallographic unit. H3K9me3 peptide affinities and specificities of Swi6 mutants. Affinity of Swi6 for H3Kc9me3 and H3K9 12N arrays.

Figure S7 (related to Figure 7)

Fr A cassette silencing phenotype of 6 *swi6+* and 6 *swi6*$^{VY \rightarrow EW}$ isolates.

Extended Experimental Procedures

Supplementary discussion

Estimation of the fraction of Swi6 molecules bound in H3K9 recognizing orientations.

Detailed description of model for the role of CD-CD interactions and the effects of altering flanking DNA.

Supplementary information references

Supplemental Figure Legends

Figure S1

Raw data and analysis of the surface plasmon resonance assay

(A) Coomassie-stained SDS-PAGE protein gel loaded with ~4μg of the cleaved, recombinant Swi6 protein used in the SPR assay.

(B) Data from three independent dose responses of Swi6 to nucleosomes captured on the streptavidin derivatized SPR chip, carried out at different nucleosome densities (Low, Medium and High) for both H3Kc9me3 (red) and H3K9 (black) nucleosomes.

(C) Unscaled binding isotherms plotted on a semi-log scale for H3Kc9me3 (circles) and H3K9 (diamonds) of the dose response titrations from (B). The inset shows the isotherms plotted on a linear scale. The scaled isotherms for three independent dose reponses of Swi6 for both H3Kc9me3 and H3K9 are shown in Figure 1c.

(D) Raw traces of the interaction between a fixed concentration of Swi6 (control sample) and H3Kc9me3 nucleosomes. The control sample was injected following every concentration of Swi6 in the titration to assay for nucleosome surface stability throughout the duration of the experiment.

(E) Fits of the scaled isotherms based on a 1:1 binding model for both H3Kc9me3 (red) and H3K9 (black) nucleosomes.

(F) Fits of the scaled isotherms based on a two-site binding model for both H3Kc9me3 (red) and H3K9 (black) nucleosomes.

As described in the main text, despite the readily apparent equilibrium and kinetic differences between Swi6's interaction with H3K9 and H3Kc9me3 nucleosomes, we cannot directly fit a model to the data for two reasons. (i) Because the data does not reveal saturation, we do not

have knowledge of the final stoichiometry for this interaction. (ii) In Figure 1c, the data are plotted as a function of total Swi6 monomer concentration. However, HP1 proteins are known to oligomerize in solution. This presents a major complication as the concentration of molecules in solution will change as a function of oligomeric state. Because we are spanning over four orders of magnitude in our dose response curves, it is likely that we are titrating Swi6 at different oligomeric states. Without direct knowledge of the oligomeric partitioning and the final stoichiometry, we cannot obtain a meaningful model to extract rate and equilibrium constants from the kinetic and equilibrium measurements. The 1:1 binding model clearly does not fit well to the data, and while the two sites binding model appears to fit well to the data, we cannot compare the two K_d values for methylated and unmethylated nucleosomes. This is because the fits result in very different final stoichiometries of Swi6:nucleosome ($B_{max,1}$ and $B_{max,2}$) for each binding event for methylated and unmethylated nucleosomes.

K_{d1} represents the first binding step, and K_{d2} represents the second binding step.

Swi6 concentrations represent monomer concentrations.

Figure S2

Loss of higher-order oligomerization in L315D Swi6 mutant

(A) Higher order oligomeric species of L315D Swi6 can be stabilized by cross-linking.

Top panel: MALS measurements for cross-linked L315D Swi6 were conducted as in Figure 2d. L315D Swi6 was cross-linked at 100 µM using EDC/NHS chemistry and then injected at 20 µM. A mixture of masses corresponding to distinct monomer, dimer, and tetramer states was observed. Bottom panel: Aliquots from the size elution chromatography fractions analyzed in the top panel were separated on SDS-PAGE and visualized by Sypro Red staining. The distribution of distinct oligomeric states visualized by the denaturing gel directly correlates with the oligomeric

masses observed with MALS, while the presence of uncross-linked Swi6 migrating at 50 kDa demonstrates that the protein was not over cross-linked.

Swi6 concentrations represent monomer concentrations.

(B) The CSD does not display detectable higher order oligomerzation. Graph represents the binding isotherms plotted as heat changes per injection (q_i) versus total monomer concentration for WT CSD (blue) and L315D CSD (red). Heat release for the L315D mutant reflects dimer dissociation (see Figure 2b). Concentration regime was 0.4 to 24 μM CSD protein.

Figure S3

(A) Quantification of gel shift experiments using unmodified (open diamonds), H3Kc9me0 (filled squares), and H3Kc9me3 nucleosomes (filled circles). The average of three experiments is shown for H3Kc9me0 and H3K9 nucleosomes and the average of two experiments for H3Kc9me3 nucleosomes.

(B) Quantification of peptide binding experiments using H3K9 (open diamonds), and H3K9me3 (filled circle) peptide. $K_{1/2}$ (μM) for methylated and unmethylated peptide was measured by fluorescence polarization. Error bars represent s.e.m. The Hill coefficients (n) determined for these experiments were variable, with the median n = 1.38 (6 experiments) and n = 1.1 (4 experiments), for H3K9me3 and H3K9 peptides, respectively.

(C) Increasing linker DNA length (L) on both sides of the 147 bp 601 sequence decreases Swi6's ability to discriminate the methyl mark on mononucleosomes. Three linker DNA lengths were assayed: 10bp/10bp linker DNA, 20bp/20bp linker DNA and 47bp/47bp linker DNA. Left graph: Swi6 discrimination for H3Kc9me3 over unmodified mononucleosomes decreases with

the linker DNA length. Right graph: Swi6 affinity for H3Kc9me3 mononucleosomes versus linker DNA length, normalized to L=10/10 nucleosomes.

All error bars represent s.e.m.

(D) $K_{1/2}$ values for DNA constructs with 20, 47 and 147 bp. Maximum variation observed was 40%.

(E) Estimation of Swi6 binding specificity for H3Kc9me3 over H3K9 core nucleosomes using a DNA - nucleosome competition assay. Scheme: An 80 bp 5' and 3' fluorescein-lableled DNA probe (F80F) was prebound to Swi6. The extent of binding was measured by fluoresence polarization (FP). Swi6 was displaced from the F80F-Swi6 complex by the addition of H3Kc9me3 or H3K9 core nucleosomes at indicated concentrations, resulting in a decreased FP signal. K_i constants for H3Kc9me3 and H3K9 nucleosomes were determined by fitting the FP data from the competition curves (right) and determined to be 0.19 μM and 1.03 μM, respectively. These values result in a 5.3 fold binding preference of Swi6 for H3Kc9me3 over H3K9 nucleosomes.

The points represent averages of two repeats for each competition curve.

Swi6 concentrations represent monomer concentrations.

Figure S4

Two analyses of SV-AUC data by c(s,f/f0) and c(s, bimodal f/f0)

(A) Application of c(s,f/f0) analysis to H3Kc9me3 mononucleosome with WT Swi6.

Top: Two dimensional representation of the raw data profiles (black dots) as a function of time (min) and radius (cm). Data were collected by following sample absorbance at 260 nm. Traces were calculated at one-minute time intervals. For clarity, only every fourth scan is shown. Solid lines represent fits of the analysis (rmsd=0.0068). The two arrows indicate two clearly separated regions in the sedimentation coefficient distribution.

Middle: The c(s, f/f0) distribution as a two-dimensional distribution with x-axis representing the sedimentation coefficient (s-values) and y-axis representing the hydrodynamic translational frictional ratio (f/f0). Below the c(s, f/f0) surface is shown a contour plot of the distribution projected into the s-f/f0 plane, where the magnitude of c(s, f/f0) is indicated by contour lines for equidistant intervals of c.

Bottom: The c(s, M) distribution as a two-dimensional distribution with x-axis representing the sedimentation coefficient (s-values) and y-axis representing molecular weight (M, KDa). Below the c(s, M) surface is shown a contour plot of the distribution projected into the s-M plane, where the magnitude of c(s, M) is indicated by contour lines for equidistant intervals of c.

(B) Application of c(s, bimodal f/f0) analysis to H3Kc9me3 mononucleosome with Swi6 WT. Top: Two dimensional representation of the raw data profiles (black dots) as a function of time (min) and radius (cm). Data were collected and traces presented as in panel A. Solid lines represent the fits of the analysis (rmsd=0.09). Middle: The c(s, bimodal f/f0) distribution with x-axis representing the sedimentation coefficient (s-values) and y axis representing c(s, bimodal f/f0). Graph is based on a s-value window from 5 to 15. Bottom: The c(M, bimodal f/f0) distribution as a two-dimensional distribution with x-axis representing the molecular weight (KDa) and y axis representing c(M, bimodal f/f0).

Figure S5

Quality control of assembled 12N(L15) and 12N(L47) nucleosome arrays and co-operativity of Swi6 binding on H3Kc9me3 arrays

(A) 12N(L15) array. 12N(L15) DNA alone (lane 1), H3K9 (lane 2) and H3Kc9me3 (lane 3) 12N(L15) arrays assembled at 1.3:1 (H3K9) or 1.1:1 (H3Kc9me3) histone octamer:DNA molar ratios were digested with 100 U EcoR1 for 3 hrs and separated on 2% TBE-agarose gels. Release

of free 162 bp DNA from assembled arrays indicates under-assembly, while appearance of species corresponding to dinucleosomes or greater indicates over-assembly. Arrays with >90% signal deriving from mononucleosomes were chosen for gel shifts.

(B) 12N(L47) array. As above, except samples were digested with 10 U EcoR1. Histone octamer to DNA molar assembly rations were 1.2:1 for H3K9 arrays and 1.1:1 for H3Kc9me3 arrays.

(C) Binding of Swi6 to H3Kc9me3 12N(L47) and 12N(L15) arrays appears more co-operative than on mono- and dinucleosomes as measured by native gel electrophoretic mobility shift assay. The following binding model was used to derive Hill coefficients from the binding data: (fraction bound = $[Swi6]^n/([Swi6]^n + K_d)$ where n= Hill coefficient. n = 2.9 for H3Kc9me3 12N (L47) and n = 15 for H3Kc9me3 12N(L15).

Swi6 concentrations represent monomer concentrations.

These Hill coefficients are only a qualitative estimate of cooperativity as the observed cooperativity could also arise if binding of multiple Swi6 molecules is required to observe stably upshifted arrays. The data, however, do strongly indicate that binding of Swi6 across multiple nucleosomes helps correctly orient Swi6 for recognition of the H3Kc9me3 mark.

(D) Sensitivity analysis for 12N(L15):Swi6 gel shifts in Figure 5c. TOP: Three of the seven Swi6 concentration points from H3K9 (blue) or H3Kc9me3 (red) 12N(L15):Swi6 binding curves were individually excluded and the remaining six points refitted using the equation in the extended experimental procedures. The excluded Swi6 concentration points are indicated in the legend. BOTTOM: Residuals for plots in TOP.

(E) Table of $K_{1/2}$ values. Calculated $K_{1/2}$ values for the plots in (D) are shown. * The 95% confidence interval is 278.9-320.3 nM. ** The 95% confidence interval is 6911-8597 nM.

Figure S6

Characterization of mutations in the chromodomain dimerization interface

(A) Domain deletion approach.

Top: Molecular weights of each of the three constructs analyzed for Swi6 domain mapping.

Bottom Left: Elution profile followed at 280 nm of the NCDH, NCD, CD constructs off a S-75 Superdex column. Both the CD and NCD show a main monomer peak and a relatively small dimer peak.

Bottom right: Elution profile followed at 280 nm of the NCDH construct off a S-200 Superdex column. This shows a main monomer peak and a relatively small dimer peak. Arrows in insets represent dimer peaks.

(B) Crystallographic unit of dHP1 CD structure (pdb 1KNE) containing two dHP1 CD monomers (black and gray), which appear to engage in contacts via the CSD-homologous alpha helix.

(C) Affinity for H3K9me3 peptide, specificity for H3K9me3 over H3K9 peptide, specificity for H3Kc9me3 over H3K9 mononucleosomes and oligomerization state (tetramer and beyond) for WT, V82E and Y131W Swi6. All data are relative to WT Swi6.

All error bars represent s.e.m. Swi6 concentrations represent monomer concentrations.

(D) Affinity for H3Kc9me3 and H3K9 12N(L15) and 12N(L47) arrays for WT Swi6 and V82E-Y131W Swi6.

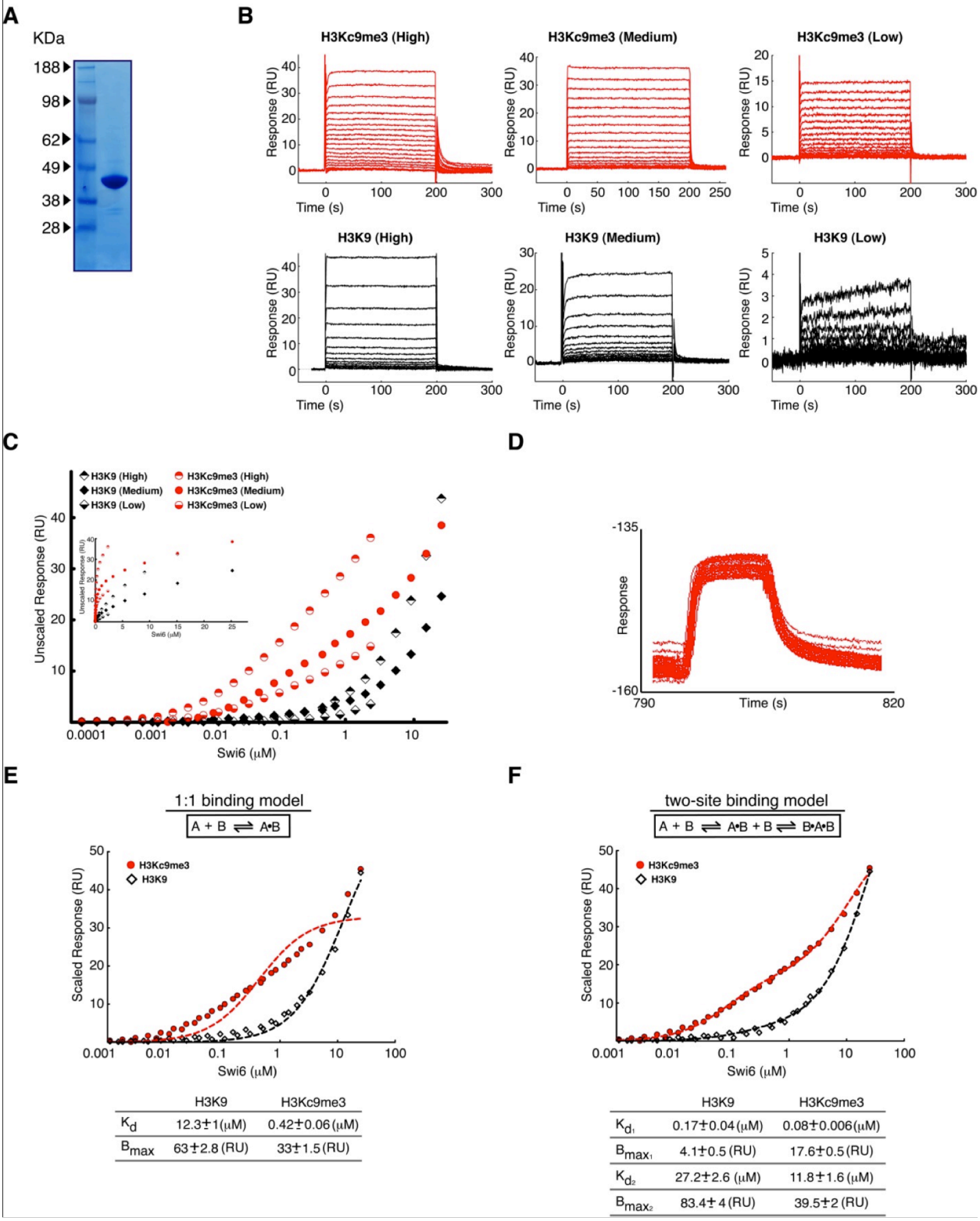
Note the $K_{1/2}$ measurements were determined by averaging the fits of three independent binding curves, while H3K9/H3Kc9me3 specificity is calculated as the average of three $K_{1/2}$ ratios from independent fits. All error bars represent s.e.m. Swi6 concentrations represent monomer concentrations.

Figure S7

Characterization of 6 independent genetic isolates for *swi6+* and *swi6*$^{VY\rightarrow EW}$ alleles.

Serial dilutions of saturated *S. pombe* cultures derived from 6 independent genetic isolates for *swi6+* and 6 independent isolates for *swi6*$^{VY\rightarrow EW}$ (genetic background as in Figure 7b). Dilutions were plated either on YS media or YS media supplemented with 5-FOA to assay silencing.

Figure S1



	H3K9	H3Kc9me3
K_d	12.3±1(μM)	0.42±0.06(μM)
B_{max}	63±2.8(RU)	33±1.5(RU)

	H3K9	H3Kc9me3
K_{d_1}	0.17±0.04(μM)	0.08±0.006(μM)
B_{max_1}	4.1±0.5(RU)	17.6±0.5(RU)
K_{d_2}	27.2±2.6(μM)	11.8±1.6(μM)
B_{max_2}	83.4±4(RU)	39.5±2(RU)

Figure S2

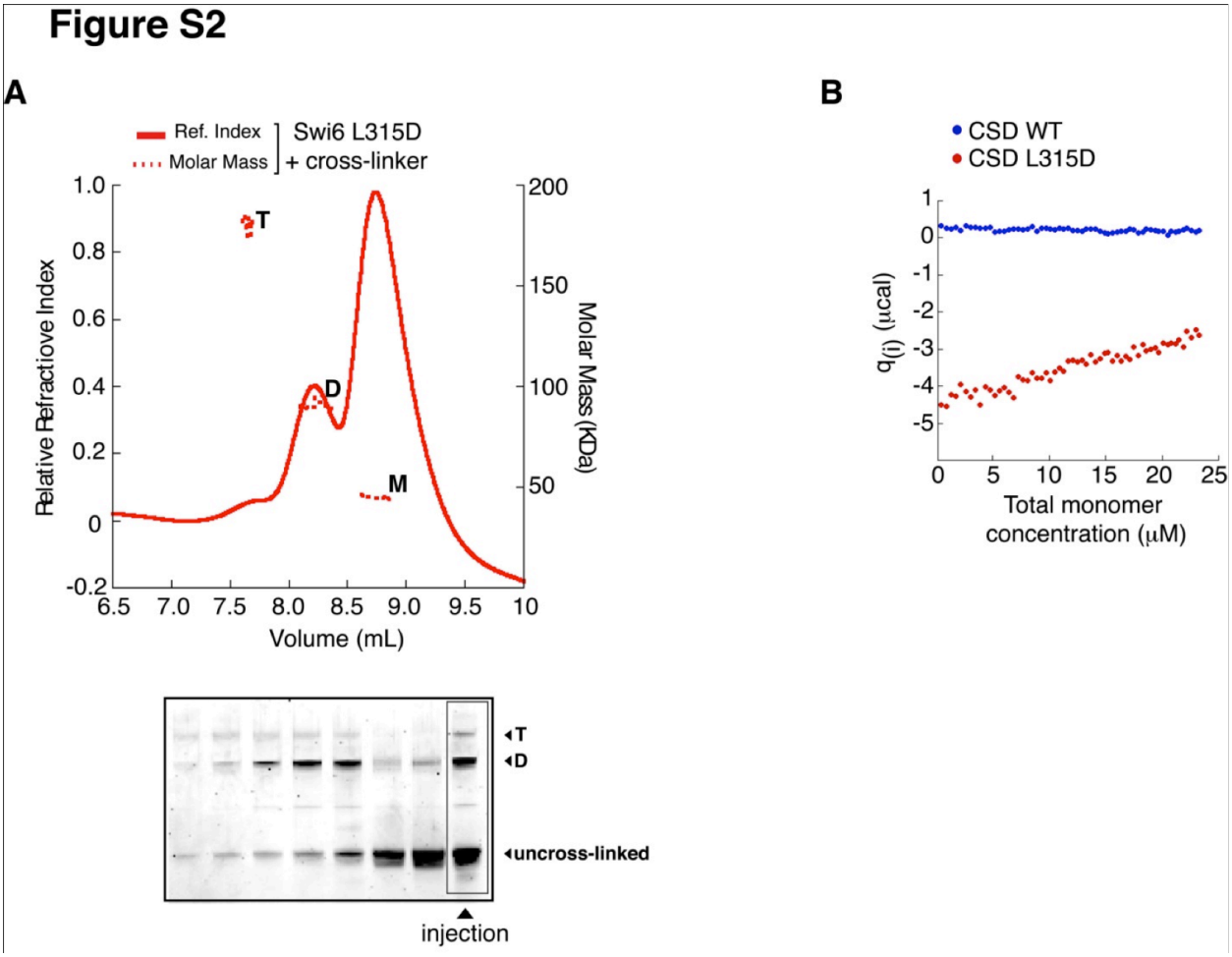
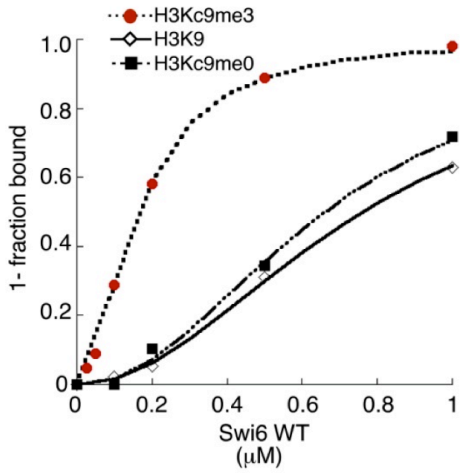
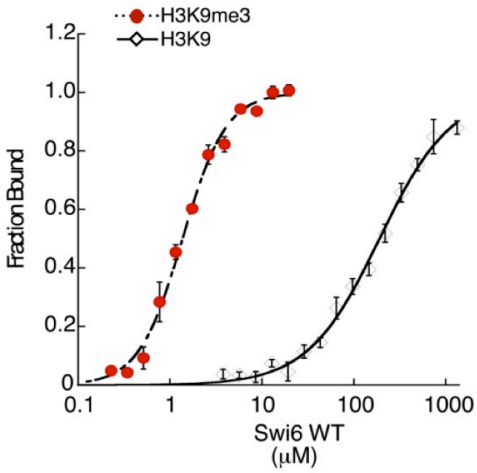

A
Ref. Index
Molar Mass
Swi6 L315D
+ cross-linker
Relative Refractiove Index
Molar Mass (KDa)
Volume (mL)
T
D
M
T
D
uncross-linked
injection
B
CSD WT
CSD L315D
q(i) (µcal)
Total monomer
concentration (µM)

Figure S3

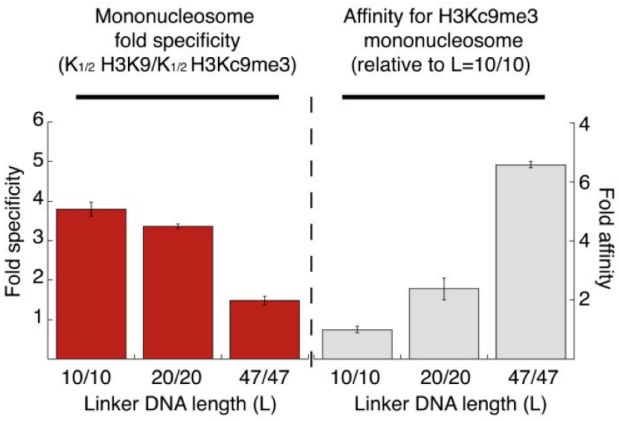
A



B



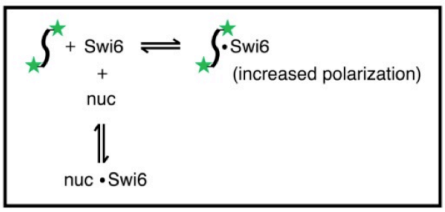
C



D

DNA Construct (bp)	Affinity ($K_{1/2}$, μM)
20	4.18
47	1.24
147	0.11

E



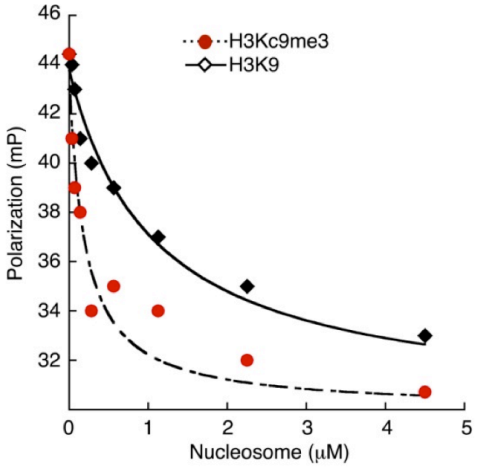


Figure S4

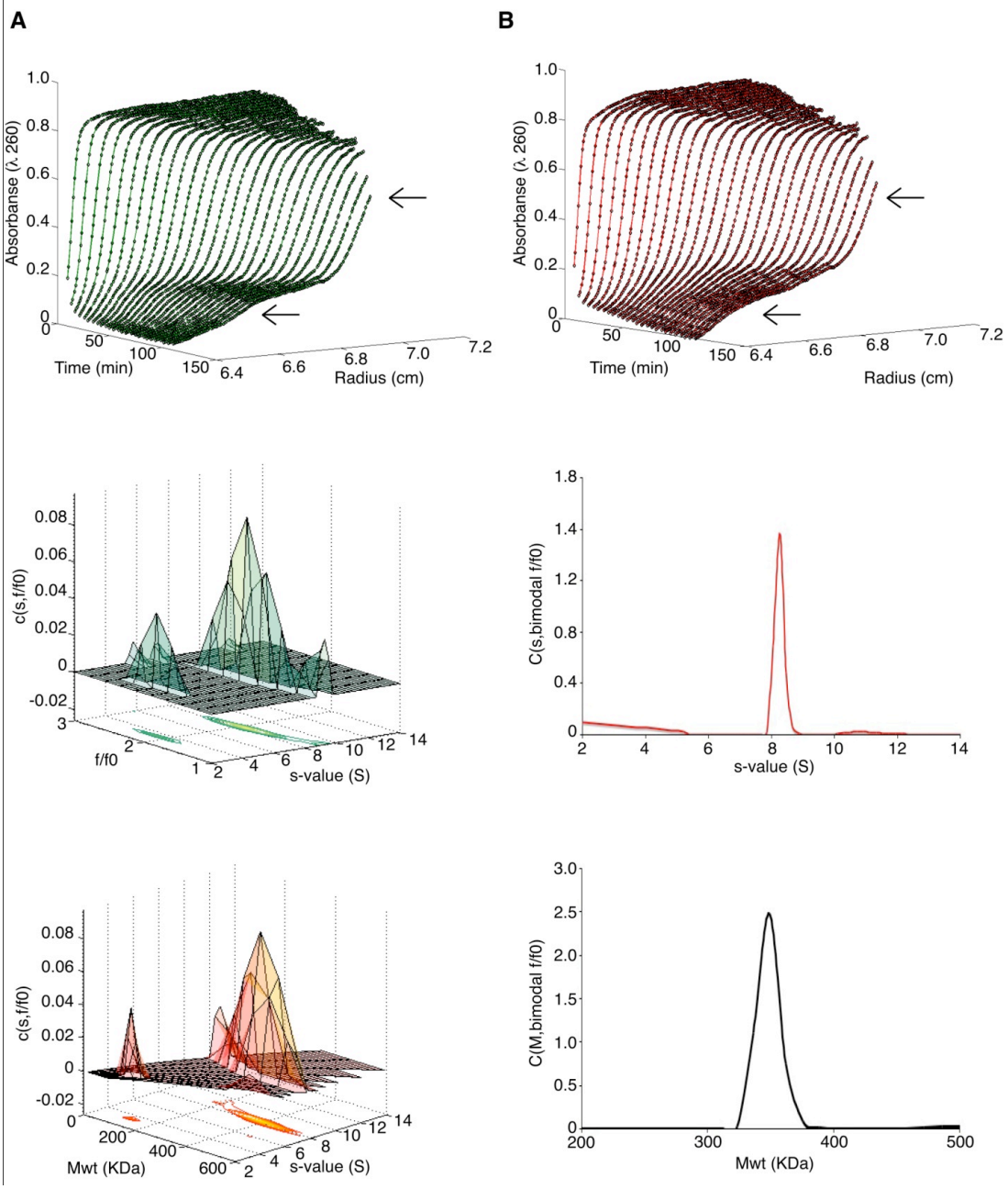
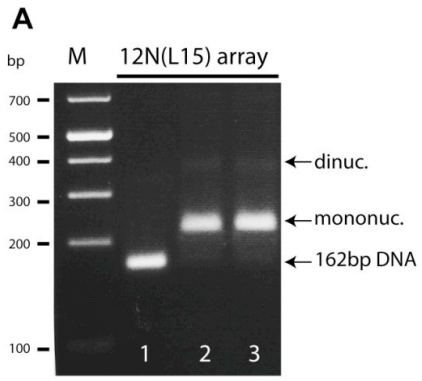
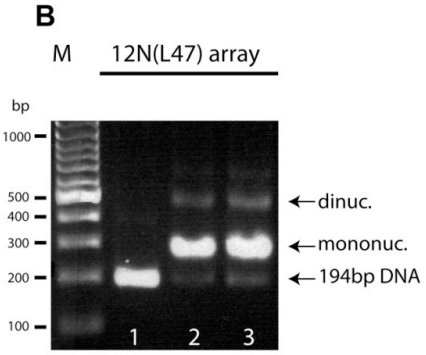
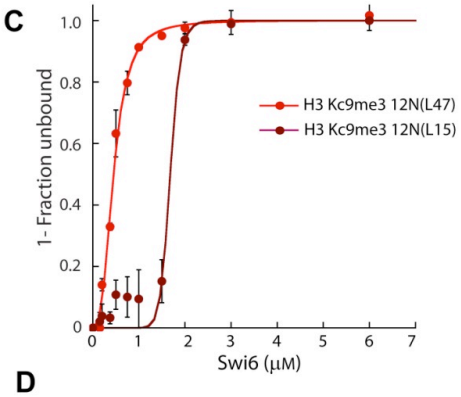

A
B
Absorbanse (λ 260)
Time (min)
Radius (cm)
c(s,f/f0)
f/f0
s-value (S)
C(s,bimodal f/f0)
Mwt (KDa)
C(M,bimodal f/f0)

Figure S5







E

Fit	$K_{1/2}$ H3K9	Fit	$K_{1/2}$ H3Kc9me3
full	7754nM*	full	299.6nM**
-3000nM	7604nM	-50nM	301.9nM
-12000nM	7775nM	-1000nM	298.6nM
-8000nM	7472nM	-250nM	276.9nM

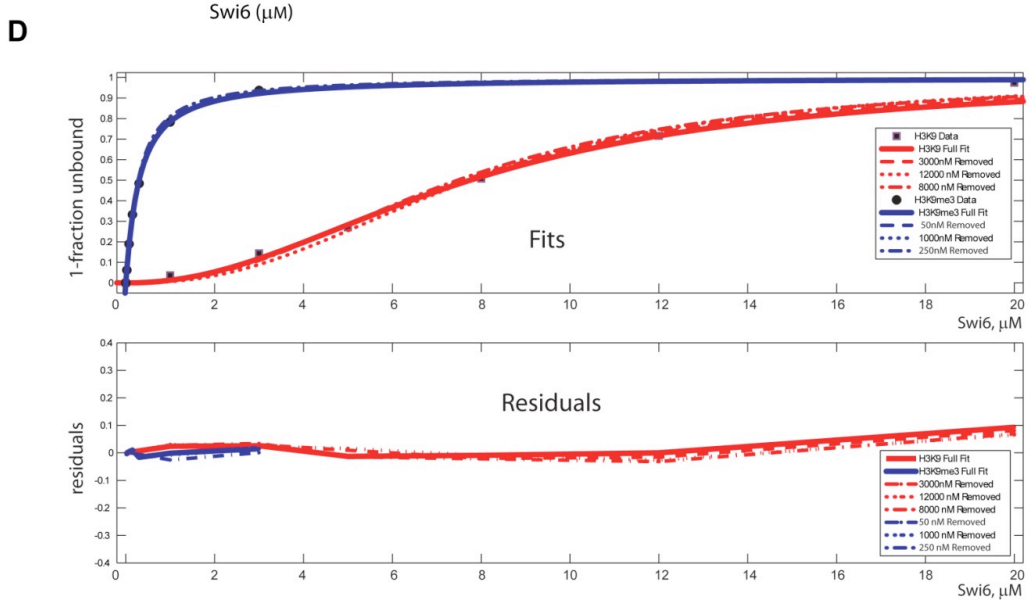
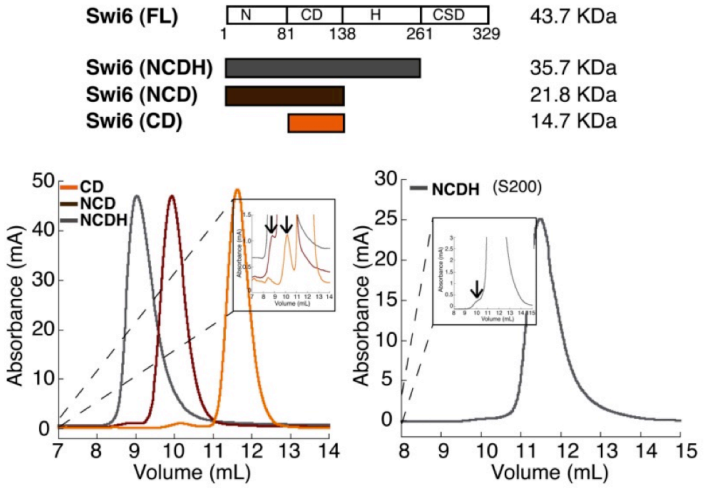
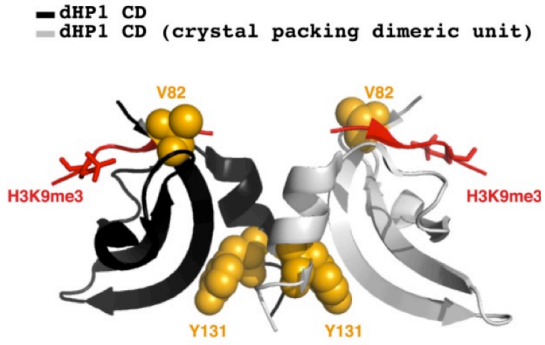


Figure S6

A



B



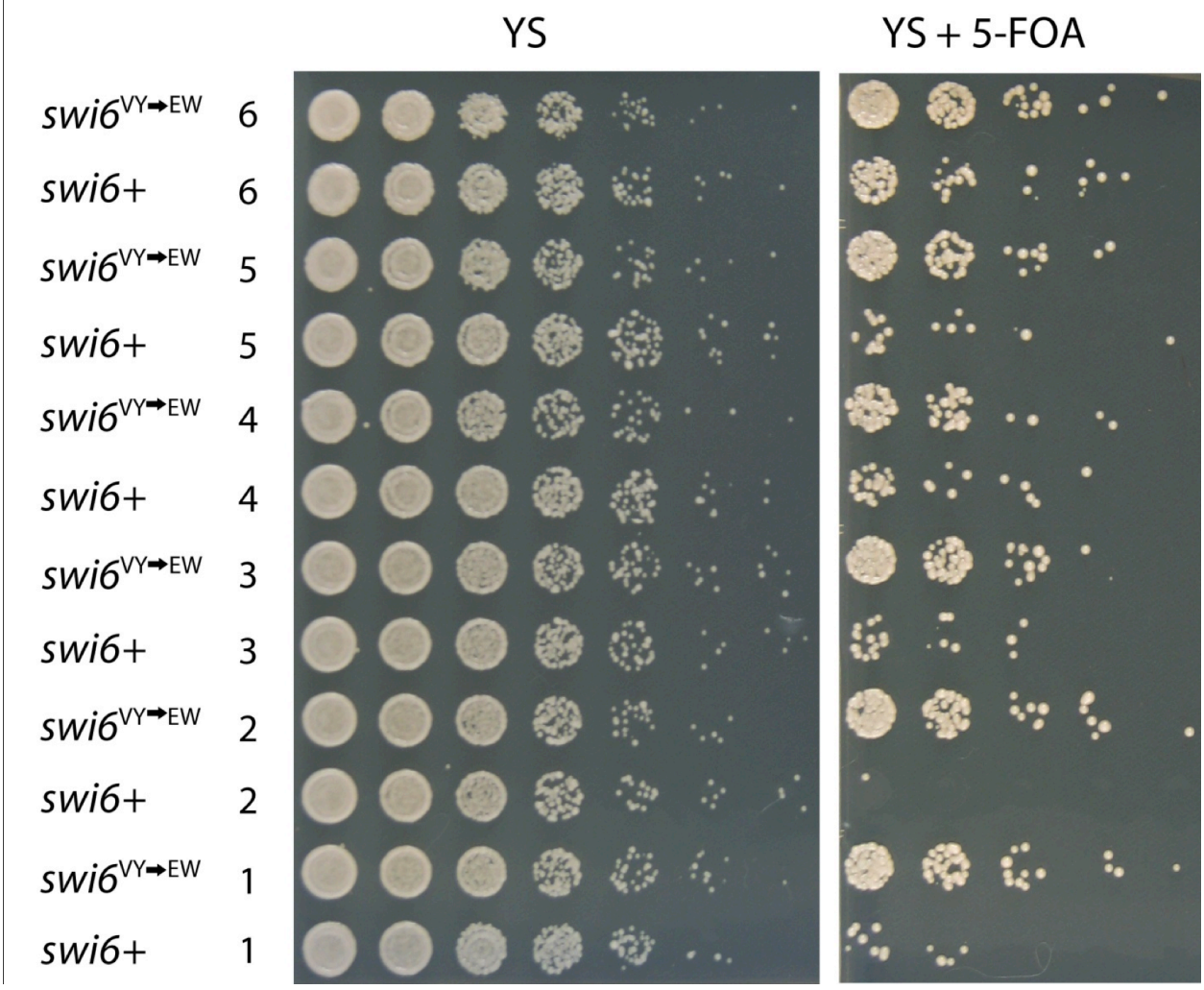
C

Swi6 Construct	H3K9me3 Peptide Affinity (K_d, μM)	Peptide Specificity (K_d H3K9/K_d H3K9me3)	Mononucleosome Specificity ($K_{1/2}$ H3K9/$K_{1/2}$ H3K9me3)	Oligomeric State (tetramer & beyond)
WT	(1) ± 0.05	(1) ± 0.06	(1) ± 0.04	(1) ± 0.20
V82E	0.32 ± 0.03	N/D	1.3 ± 0.11	1.2 ± 0.09
Y131W	2.17 ± 0.13	0.5 ± 0.03	0.9 ± 0.04	1.6 ± 0.21

D

Swi6 Construct	H3Kc9me3 12N(L15) Affinity ($K_{1/2}$, μM)	H3K9 12N(L15) Affinity ($K_{1/2}$, μM)	H3Kc9me3 12N(L47) Affinity ($K_{1/2}$, μM)	H3K9 12N(L47) Affinity ($K_{1/2}$, μM)
WT	0.33 ± 0.07	8.1 ± 1.37	0.12 ± 0.02	0.58 ± 0.08
V82E-Y131W	0.18 ± 0.04	28.1 ± 3.62	0.59 ± 0.11	4.41 ± 0.24

Figure S7


YS
YS + 5-FOA
swi6VY➡EW 6
swi6+ 6
swi6VY➡EW 5
swi6+ 5
swi6VY➡EW 4
swi6+ 4
swi6VY➡EW 3
swi6+ 3
swi6VY➡EW 2
swi6+ 2
swi6VY➡EW 1
swi6+ 1

Extended Experimental Procedures

Protein cloning and purification

Tagged and untagged Swi6 were cloned into pET30a at the BamH1 and Not1 sites. The constructs for both tagged and untagged Swi6 contain an N-terminal 6xHis tag encoded by the pET30a vector, but the untagged Swi6 contains a TEV cleavage site just upstream of the Swi6 coding region. Swi6 proteins were purified from *E. coli* Rosetta (DE3) pLysS strains as follows. Cells were grown to OD 0.4-0.5 at 37°C in LB medium with 100 µg/mL Kanamycin. Isopropyl-β-D-thiogalactopyranoside was added to a concentration of 0.4 mM to induce protein expression, and cells were incubated overnight at 18°C. Harvested cells were resuspended in lysis buffer (1X PBS buffer pH 7.3, 300 mM NaCl, 10% glycerol, 0.1% Igepal CA-630, 7.5 mM Imidazole, and protease inhibitors). Following sonication, cell debris was removed by centrifugation at 25,000g for 20 min. Cell lysate supernatants were incubated for 1 hour at 4°C with Cobalt-NTA affinity beads (Clontech). Beads were washed with lysis buffer and proteins eluted with 25 mM HEPES pH 7.5, 100 mM KCl, 10% glycerol and 250 mM Imidazole. Proteins were further purified by size-exclusion chromatography on a Superdex 200HR 10/300 column (GE Healthcare) into a final elution buffer containing 25 mM HEPES pH 7.5, 100 mM KCl, and 10% glycerol. Tagged Swi6 was then stored at -80°C. Untagged Swi6 underwent two additional steps: (i) TEV protease was used to cleave the N-terminal 6x-His tag; and (ii) Anion exchange chromatography on a Mono Q 4.6/100 PE column (GE Healthcare) was used to separate the cleaved product from the TEV protease. Protein concentrations of all Swi6 construct samples were measured by UV absorption at 280 nm and calculated using the experimentally determined extinction coefficient ε = 34,776 M^{-1} cm^{-1}.

EDC/NHS Protein crosslinking

Samples were incubated for 2 hours at room temperature in a total volume of 20 µL with 2 mM EDC and 5 mM NHS. EDC and NHS chemistry specifically cross-links aspartate and glutamate residues to nearby lysine residues. Reactions were quenched by adding hydroxylamine to a final concentration of 10 mM, followed by SDS loading dye.

Mono- and Dinucleosome reconstitution

The mononucleosome 601 positioning sequence, containing a Pst1 site 18 bp in from the 5' end, was amplified by PCR and gel purified. The DNA fragment was assembled into mononucleosomes with recombinant *Xenopus laevis* histones by salt dialysis over 48-60 hrs (Luger et al., 1999). Reconstituted mononucleosomes were purified using a glycerol gradient. All histone octamer assemblies, nucleosome assemblies and nucleosome purifications were performed in the presence of 2 mM DTT to maintain the MLA modification.

Dinucleosome DNA templates were cloned into the pTNT vector using the Eag1 and Xho1 restriction sites. The plasmid was amplified in a dcm and dam methylation defective *E. coli* strain and the construct released by restriction enzyme digestion, then purified by native gel electrophoresis. Subsequent dinucleosome assembly steps are as described above.
Each 601 positioning sequence of the 12N arrays were separated by an EcoR1 restriction site. The array was cloned into a pCR-0Blunt backbone with EcoRV and Xho1 sites and the plasmid amplified in a dcm and dam methylation defective *E. coli* strain. The array was released by restriction digestion and purified from the digested backbone by Sephacryl S-1000 matrix gel filtration. After assembly, arrays were dialyzed into 10 mM Tris pH 8.0, 0.25 mM EDTA, 2 mM DTT and 50 mM KCl. Quality of assembly was assessed by EcoRI digestion.

Surface plasmon resonance

Matrix-free, flat, carboxymethylated gold surfaces (Sensor Chip C1, GE Healthcare) were preconditioned with five-minute pulses of 0.25% SDS and 50 mM NaOH at 25° C. Individual flow cells were prepared with the following protocol: (i) 50 µl injection of 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide / N-hydroxysuccinimide (0.5M:0.2M); (ii) 30 µL injection of 0.25 mg ml^{-1} ImmunoPure Streptavidin (Thermo Scientific) in sodium acetate buffer (pH 5.0) to a total amount of 250RU for all four flow cells; (iii) 60 µl injection of 1M ethanolamine. Nucleosomes were assembled on 147bp of 5'biotin tagged 601 sequence DNA and immobilized by injecting 2µl of 25nM H3Kc9me3 or H3K9 mononucleosomes onto active flow cells.

Prior to each SPR experiment, protein samples were dialyzed against 25 mM HEPES, 150 mM KCl, and 2 mM DTT. Following dialysis, Tween 20 detergent (Sigma-Aldrich) was added to the protein solution and the dialysis buffer to achieve a final concentration of 0.005% (v/v). In order to minimize refractive index differences between sample and buffer, the dialysis buffer was then used as the assay buffer for the SPR measurements. Swi6 concentration was determined by absorbance at 280 nm. Dose response titrations were prepared by manual 0.6 fold serial dilutions of the highest concentration into assay buffer. Individual sample cycles consisted of a 30 second buffer injection followed by a 200 second sample injection at a flow rate of 20 µl min^{-1}.

Fluorescence polarization binding measurements

Peptide polarization assays were conducted in buffer containing 1 mg/ml BSA, 50 mM HEPES pH 7.5, 100 mM KCl and 10% glycerol. Peptide concentrations were held at 50-100 nM in a reaction containing variable amounts of Swi6 protein. The binding reaction was incubated 30 min at RT and fluorescence polarization was measured using a Molecular Devices HT Analyst (λ_{ex} =480nm, λ_{em} =530nm). The following binding modelwas used to derive $K_{1/2}$'s for peptide binding from the polarization data:

$$FP_{obs} = \frac{[Swi6]^n * FP_{max} + K_{1/2}{}^n * FP_{min}}{[Swi6]^n + K_{1/2}{}^n}$$

FP_{obs} was then converted to fraction bound (Fr. bound) with the following equation:

$$\text{Fr. bound} = \frac{FP_{obs} - FP_{min}}{FP_{max} - FP_{min}}$$

Fr. bound was then plotted versus [Swi6] to obtain the K_d using the following equation:

$$\text{Fr. bound} = \frac{[Swi6]^n}{[Swi6]^n + K_{1/2}{}^n}$$

where FP_{min} is the polarization signal for the probe alone, FP_{max} is the polarization signal at saturating [Swi6] and n = Hill coefficient.

The DNA to assemble fluorescent nucleosomes was labeled on one end by amplifying the sequence using PCR with a primer (sequences available upon request) covalently linked to 6-carboxyfluorescein by a 6-carbon linker (IDT). Nucleosome polarization assays (Chin et al., 2004) were conducted in buffer containing 10% glycerol, 80 mM KCl, 20 mM HEPES buffer, 4 mM Tris, and 0.2 mM EDTA, pH 7.5. We predicted that binding by Swi6 to the H3 tail would cause local restriction of the fluorescein dye and increase fluorescence polarization. Consistent with this hypothesis, we observed an increase in fluorescence polarization with increasing concentrations of Swi6 (Figure 1d). Each anisotropy sample contained a final nucleosome concentration of 3 nM. Untagged Swi6 was serially diluted by 0.6-fold from a maximum concentration of 20 μM. Data points from three independent Swi6 dilution curves were

averaged and standard errors calculated. Polarization was measured on an Analyst AD platereader (Molecular Devices).

Native gel mobility shift assay systems

Each 20 µL sample contained buffer with 20 mM HEPES pH 7.5, 4 mM Tris pH 7.5, 80 mM KCl, 0.1% Igepal CA-630, 0.2 mM EDTA, 2 mM DTT and 4-10% glycerol. Mononucleosome samples were incubated for 45 minutes at room temperature (1 hr for dinucleosomes), then loaded on a 0.5X TBE 6% 29:1 acrylamide:bis-acrylamide gel (0.4X TBE 4% 37.5:1 acrylamide:bis-acrylamide for dinucleosomes) and run for 5 hours at room temperature at 75 V (3 hours for dinucleosomes). Array gel shift samples were loaded on 1X Tris-Acetate 1.15% agarose gels and separated for 4 hrs at 2.5 V/cm.

Binding curves were fit with the equation, fraction bound = $[Swi6]^n/([Swi6]^n + K_{1/2}^n)$. A description of the $K_{1/2}$ value can be found in (Ackers et al., 1982).

Isothermal titration calorimetry

Experiments for both WT and L315D Swi6 were performed in 25 mM sodium phosphate pH 7.5, 100 mM KCl at 15°C. Data were analyzed with the Microcal Origin software using a monomer-dimer model. The final dilution point, 17nM, in the experiment shown in Figure 2b (left panel) is near the recommended dilution limit for the instrument used. While this concentration is close to the recommended lower concentration limit for the calorimeter, it has been shown previously that heat release from homo-oligomer dissociations can be robustly monitored in this low concentration regime for interactions of K_ds >100nM (Luke et al., 2005). Therefore, we believe that the lack of any measureable heat release for WT Swi6 CSD in the 17-100nM range suggests that the K_d for CSD self-association is below 17nM, extending previous estimates (Brasher et al., 2000).

Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS/UV/RI)

The SEC-MALS/UV/RI (Arakawa, 2001; Wyatt, 1993) system was equilibrated in 25 mM HEPES or sodium phosphate, pH 7.5 and 100 mM KCl at a flow rate of 0.35 ml/min. Molar mass determination was performed using the ASTRA software. Cross-linked sample reactions were quenched and diluted to 100 µL, and the buffer exchanged to the SEC-MALS buffer using MicroSpin S-400 HR Columns X 25 (Amersham Biosciences) before injection onto the SEC-MALS system. Fractions of crosslinked Swi6 eluting from the SEC-MALS/UV/RI system were denatured and separated on 4-12% NuPAGE gradient gels (Invitrogen). Gels were stained with Sypro red and visualized on a Typhoon scanner.

Sedimentation Velocity Analytical Ultracentrifugation

H3Kc9me3 mononucleosomes and Swi6 proteins were individually dialyzed into 25 mM Tris 7.5, 80 mM KCl, 10% glycerol, and 2 mM DTT. Mononucleosomes were quantified by ethidium bromide staining and Swi6 proteins by UV absorption at 280 nm. The samples were prepared in 400 µl with an overall final OD of approximately 1 at 260 nm. Samples were incubated for 45 minutes at 24°C, then placed in an AUC chamber pre-equilibrated at 24°C and kept at zero-rpm at 24°C for 1 hour under vacuum. Runs were performed at 24°C at a rotor speed of 36K rpm for 4 hours. Scans were collected at 260 nm, with a radial step size of 0.003 cm and continuous scanning mode at approximately one minute intervals.

Western blotting

Strains were grown to log phase (OD 0.5), washed in 1XTBS, and total proteins extracted under denaturing conditions (Knop et al., 1999). Aliquots representing 0.4 OD units were separated on a 4-12% NuPAGE gradient gel (Invitrogen), blotted onto PVDF membranes and

probed with monoclonal anti-ϒ-tubulin (Sigma) or polyclonal anti-Swi6 antisera (Nakayama et al., 2000), and fluorescent secondary antibodies. Blots were scanned and quantified on a LiCor Odyssey scanner.

Chromatin immunoprecipitation (ChIP)

Strain growth and chromatin immunoprecipitation was performed as previously described (Rougemaille et al., 2008), with some modifications. Cell lysis was performed in 5 beat-beating cycles and Lysis buffer was supplemented with 1 mM PMSF, 3 µg/ml leupeptin, 2 µg/ml aprotinin and 1 µg/ml pepstatin. Chromatin fractions were purified and sonicated as described (Rougmaille et al). The lysate was brought up to 1.3 ml in extraction buffer and 0.1 ml set aside as the input fraction. The rest of the lysate was agitated on a nutator overnight at 4°C with either 1.4 µl anti-H3K9me2 antibody (Abcam ab1220) or 2 µl anti-Swi6 polyclonal antisera (Nakayama et al., 2000). 30 µl of a 50% slurry of protein A-coated magnetic beads (Dynal-Invitrogen) equilibrated in Lysis buffer was added to the lysate and incubated on a nutator for 90 min at 4°C. Beads were washed 2x 5 min in Lysis buffer, 2x 5 min in High Salt Lysis buffer, 2x 2.5 min in Wash Buffer and 1x 5 min in TE (buffers as in (Rougemaille et al., 2008)). DNA was eluted and de-crosslinked as described (Rougemaille et al., 2008). DNA eluates were quantified by RT-qPCR using Dynazyme II DNA polymerase (Finnzymes). Primers available upon request. As normalization controls for ChIP experiment containing the silencing cassette with Fr A, we used a Fr A^- control that contains all the features of the insertion cassette shown in Figure 7a, including the NatR marker, and only lacks the centromeric insert (Fr A). This ensures that in the comparison between strains, the only genomic difference is the presence or absence of Fr A.

DNA-nucleosome competition assay

In this assay, binding of Swi6 by nucleosomes reduces the fraction of DNA bound by Swi6 and results in a decrease in fluorescence polarization. Measurement of the decrease in fluorescence polarization as a function of nucleosome concentration is used to obtain a dissociation constant for nucleosome binding. An 80 bp DNA fragment was amplified from a plasmid containing the 601 positioning site with the Pst18 site using forward and reverse primers both 5' labeled with 5,6 carboxy-fluorescein (F80F). The K_d and B_{max} (maximal polarization value at saturation) for the F80F interaction with Swi6 were determined by fluorescence polarization assays as described above. The F80F probe was incubated with Swi6 in 25 mM HEPES pH 7.5, 100 mM KCl, 10% glycerol and 0.1 mg/ml BSA, with [F80F] = 10 nM and [Swi6] = 190 nM. This regime results in ~ 3.5% of the maximal polarization signal for the F80F-Swi6 interaction and is at ~ $K_{1/2}$ as assayed by gel shift. Duplicate, 2-fold dilutions of unmodified or H3Kc9me3 core nucleosomes were titrated into the F80F-Swi6 reactions and incubated 45 min at RT. Fluorescence polarization was measured as above. The K_i for unmodified or H3Kc9me3 nucleosomes was determined by fitting the data to the following equation:

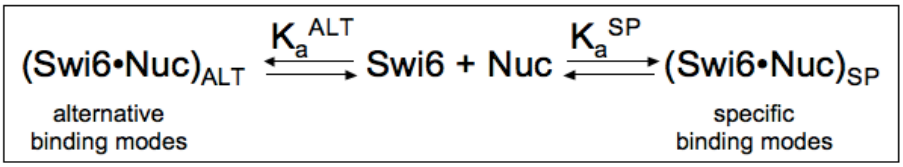
$$FP_{obs} = \frac{K_i^{NUC}(FP_{max} + [Swi6]) + FP_{min}K_d^{DNA}[nuc]_t}{K_i^{NUC}(K_d^{DNA} + [Swi6]) + K_d^{DNA}[nuc]_t}$$

where FP_{max} is the FP value for Swi6-F80F binding at saturating Swi6, FP_{min} is the FP value for F80F alone; K_i^{NUC} is the K_i for H3Kc9me3 or H3K9 nucleosomes and K_d^{DNA} is the K_d for the F80F-Swi6 interaction.

Supplementary discussion

Estimation of the fraction of Swi6 molecules bound in H3K9 recognizing orientations

We have used the simple model depicted below to estimate the proportions of Swi6 molecules bound in orientations that recognize the H3K9 residue (specific) vs. orientations that do not recognize the H3K9 residue (alternative).



We make the following initial simplifying assumptions:

(i) Swi6 binds with the same molecularity in the specific and alternative orientations.

(ii) Binding in the specific and alternative orientations is mutually exclusive.

(iii) The alternative orientations are the same in the context of H3K9 and H3Kc9me3 nucleosomes.

We can then describe the observed association constant for binding in the context of H3Kc9me3 nucleosomes, $K_a^{obs(M)}$ as follows:

$$K_a^{obs(M)} = K_a^{SP(M)} + K_a^{ALT} \quad (1)$$

Where $K_a^{SP(M)}$ is the association constant for binding in the specific mode for the H3Kc9me3 nucleosomes and K_a^{ALT} is the sum of all association constants for binding in the different alternative modes.

The observed association constant for binding to H3K9 nucleosomes, $K_a^{obs(U)}$ is:

$$K_a^{obs(U)} = K_a^{SP(U)} + K_a^{ALT} \quad (2)$$

Where $K_a^{SP(U)}$ is the association constant for binding in the specific mode for the H3K9 nucleosomes.

We then make the further assumption that the methyl mark contributes the same amount of free energy when bound by Swi6 on the H3 tail peptide and when bound by Swi6 in the specific orientation on the nucleosome. Mathematically, this can be expressed by the equation:

$$K_a^{SP(M)} / K_a^{SP(U)} = \text{the specificity of binding to H3 tail peptides}$$

Using the above formalisms and our experimental measurements, we obtain the following:

$$K_a^{SP(M)} / K_a^{SP(U)} = 130 \quad (3)$$

The observed specificity on nucleosomes is obtained from the reciprocal ratio of $K_{1/2}$ values, which represent aggregate dissociation constants, to give:

$$K_a^{obs(M)} / K_a^{obs(U)} = 4.6 \quad (4)$$

Using eq. 3, we obtain:

$$K_a^{SP(M)} = 130 K_a^{SP(U)} \quad (5)$$

Using eq. 5 in eq. 1 and eq. 2 we obtain the following from eq. 4:

$(130K_a^{SP(U)} + K_a^{ALT})/(K_a^{SP(U)} + K_a^{ALT}) = 4.6$

Simplifying, $K_a^{SP(U)} = 0.03\ K_a^{ALT}$

From eq. 5: $K_a^{SP(M)} = 130 \times 0.03\ K_a^{ALT} = 3.9\ K_a^{ALT}$

Using these relationships between the specific and alternative mode association constants, we can then derive the fraction of molecules that are bound in the specific mode.

The fraction of molecules bound in the specific modes for H3K9 nucleosomes is:

$Fr(sp)^{U} = K_a^{SP(U)} / (K_a^{SP(U)} + K_a^{ALT}) = 0.03/(0.03+1) = 0.03$, or **3%**

The fraction of molecules bound in the specific modes for H3Kc9me3 nucleosomes is:

$Fr(sp)^{M} = K_a^{SP(M)} / (K_a^{SP(M)} + K_a^{ALT}) = 3.9/(3.9+1) = 0.8$, or **80%**

Detailed description of model for the role of CD-CD interactions and the effects of altering flanking DNA

We hypothesize that Swi6 oligomerization mediated by CSD-CSD and CD-CD interactions promotes recognition of the H3K9 methyl mark in three ways (Figure 6f): (i) it correctly orients Swi6 on a nucleosome to recognize the H3K9 residues as implied by our results with mononucleosomes, (ii) it further promotes the H3K9 specific orientations via bridging interactions with nearby nucleosomes and (iii) it sterically and energetically disfavors binding in alternative orientations, which include binding to linker DNA. The above formulation allows a better understanding the origin of the observed specificity decrease concomitant with increased flanking DNA. Increasing flanking DNA in the context of the arrays is expected to: (i) reduce the ability of Swi6 CDs to bridge nearby nucleosomes and (ii) increase non-specific orientations by increasing available free DNA sites. In contrast, decreasing flanking DNA in the context of a

mononucleosome substrate will only result in a decrease in non-specific orientations without affecting bridging, as there are no adjacent nucleosomes. The model thus predicts that increasing flanking DNA in the context of arrays will cause a greater reduction in specificity than in the context of mononucleosomes. Consistent with this prediction, we observe that Swi6 displays ~5-fold decreased specificity when the 15 bp linker DNA in arrays is increased to 47 bp (Figure 5c&d), but only a ~1.5-fold reduction in specificity when the linker DNA in mononucleosomes is increased from zero or 10 bp on either side to 20 bp on either side (Figures 3c &S3d).

Supplementary Information References

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc Natl Acad Sci U S A *79*, 1129-1133.

Arakawa, T., Wen J. (2001). Size-exclusion chromatography with on-line light scattering. In Current Protocols in Protein Science (Hoboken, NJ, John Wiley and Sons, Inc.), pp. 1-21.

Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chrome domain dimer. Embo Journal *19*, 1587-1597.

Chin, J., Langst, G., Becker, P.B., and Widom, J. (2004). Fluorescence anisotropy assays for analysis of ISWI-DNA and ISWI-nucleosome interactions. In Chromatin and Chromatin Remodeling Enzymes, Pt B (San Diego, Academic Press Inc), pp. 3-16.

Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: More tags and improved practical routines. Yeast *15*, 963-972.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. In Chromatin (San Diego, Academic Press Inc), pp. 3-19.

Luke, K., Apiyo, D., and Wittung-Stafshede, P. (2005). Dissecting homo-heptamer thermodynamics by isothermal titration calorimetry: Entropy-driven assembly of co-chaperonin protein 10. Biophysical Journal *89*, 3332-3336.

Nakayama, J., Klar, A.J., and Grewal, S.I. (2000). A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. Cell *101*, 307-317.

Rougemaille, M., Shankar, S., Braun, S., Rowley, M., and Madhani, H.D. (2008). Ers1, a rapidly diverging protein essential for RNA interference-dependent heterochromatic silencing in Schizosaccharomyces pombe. J Biol Chem *283*, 25770-25773.

Wyatt, P.J. (1993). LIGHT-SCATTERING AND THE ABSOLUTE CHARACTERIZATION OF MACROMOLECULES. Analytica Chimica Acta *272*, 1-40.

Appendix 5:

Design of a photoswitchable cadherin

In collaboration with:

Ryan Ritterson, Michael Michalik, and Tanja Kortemme

Reprinted with permission from:

Ritterson R, Kuchenbecker KM, Michalik M, Kortemme T. Design of a Photoswitchable Cadherin. J.Am. Chem. Soc., 2013, 135 (34), pp 12516-12519.

Foreword:

For the cell-cell adhesion protein cadherin, the probability of dimer formation is represented by a two-dimensional surface of calcium and cadherin concentrations.

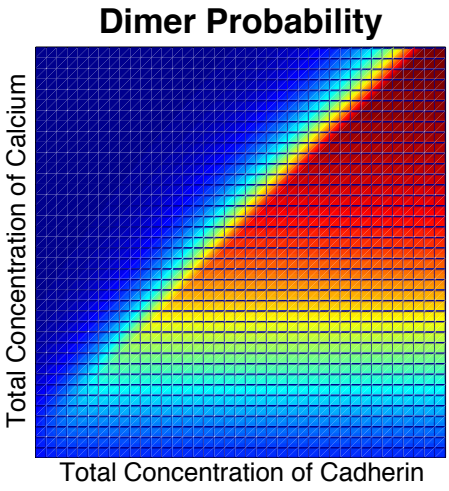


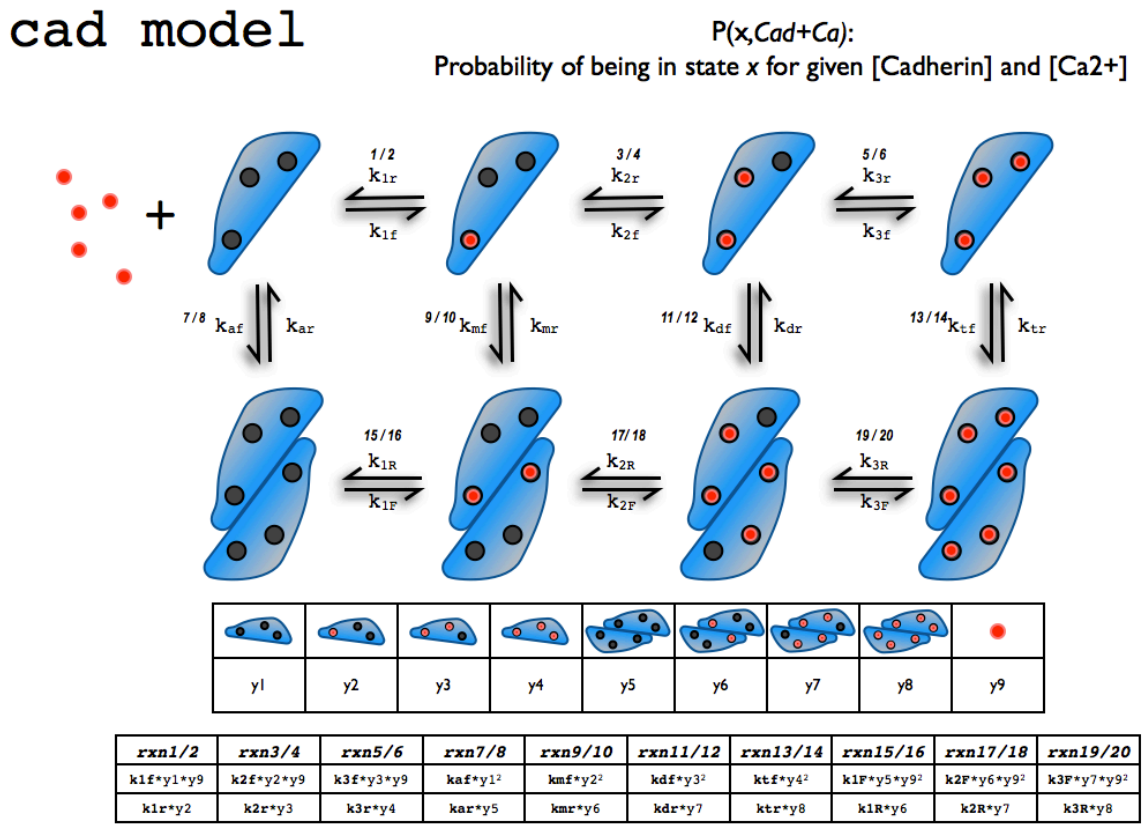
Figure f1. Two dimensional concentration dependence of cadherin dimerization.

In response to changes in extracellular calcium concentration, cadherin's self association properties modulate cell-cell adhesion processes. The transition between states is made switch like by the cooperative binding of 3 calcium ions to an individual cadherin monomer. Calcium binding acts as the allosteric promoter of dimer formation between calcium ligated monomers. The probability of dimer formation in this system is a two-dimensional surface with axes of [cadherin] and [calcium].

In an amazing feat of protein/chemical engineering, Ryan Ritterson (Kortemme Lab) covalently attached a small molecule to cadherin's structure in the region of one of the calcium binding sites. This small molecule created an artificial linkage between distant sites of the cadherin monomer. In response to UV excitation, the cis-state of the small molecule is excited and the distance between attachment points on the structure of cadherin shifts by ~5Å. Because this 5Å shift was adjacent to a calcium binding site, it was hypothesized that the cis/trans populations would have different calcium binding affinities, and, by extension, different dimerization potentials. The UV-control of the isomerization state, would then give excellent temporal and spatial control in the perturbation of the calcium / cadherin surface.

I entered this project in the design selection stage. Initial rounds of SPR data collection and construct selection proceeded via the qualitative comparison of binding pre and post illumination. However, in order to provide proof of principle for SPR measurement of dimerization, significant efforts were put towards computational models of the interplay between calcium and cadherin in the dimerization potential. For binding of three calcium ions to the three calcium binding sites of the monomer, there are eight different ligation states. These eight different species of calcium bound monomers create the potential for 36 different dimer species. This is a very complex system, and there is no analytical solution to the coupled equilibria.

Our ultimate aim was to use SPR to monitor dimer formation as a function of both calcium and cadherin concentration. In order to provide a physical basis for the observed



rxn1/2	rxn3/4	rxn5/6	rxn7/8	rxn9/10	rxn11/12	rxn13/14	rxn15/16	rxn17/18	rxn19/20
k1f*y1*y9	k2f*y2*y9	k3f*y3*y9	kaf*y1²	kmf*y2²	kdf*y3²	ktf*y4²	k1F*y5*y9²	k2F*y6*y9²	k3F*y7*y9²
k1r*y2	k2r*y3	k3r*y4	kar*y5	kmr*y6	kdr*y7	ktr*y8	k1R*y6	k2R*y7	k3R*y8

Figure f2. Simplified model of calcium dependent cadherin dimerization. For numerical solution, nine states, including free calcium ions, were considered (labeled y1 through y9).

binding, it was necessary to model the concentration of calcium liganded monomeric cadherin as a function of the calcium binding and dimerization energetics at different positions on the calcium / cadherin concentration surface. In simplification of the reaction scheme in Figure f2, I made the assumption that only monomers can bind Ca^{2+} (k1F/R = k2F/R = k3F/R = 0). Additionally, calcium binding was made strongly cooperative by depleting the singly and doubly bound intermediate species (10^{12} = k2f/r = k3f/r >> k1f/r ~ 10^{6}), and dimerization was assumed to only occur between triply liganded monomers (kaf/r = kmf/r = kdf/r = 0). To this end, we sought a model that depended on two energetic constants, a fully cooperative calcium binding process [cadherin + 3 Ca^{2+} $\rightarrow$ cadherin•$(Ca^{2+})_3$], and a dimerization process of calcium ligated cadherin monomers [cadherin•$(Ca^{2+})_3$ + cadherin•$(Ca^{2+})_3$ $\rightarrow$ (cadherin•$(Ca^{2+})_3)_2$].

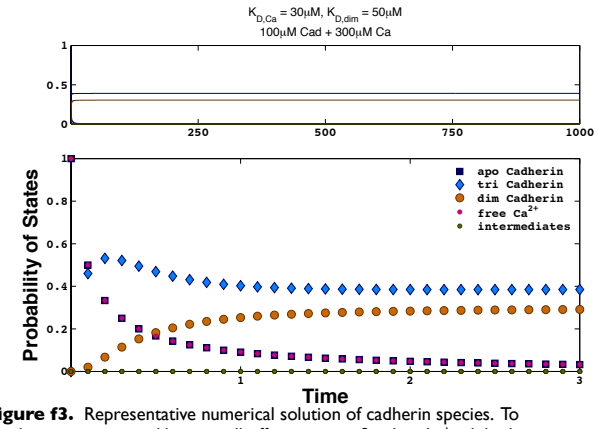


Figure f3. Representative numerical solution of cadherin species. To accelerate system equilibration, all off-rates were fixed to $1s^{-1}$, while the on-rate of calcium binding and dimerization were taken as $(K_D)^{-1}s^{-1}$. Each numerical solution was run for 1000 time steps, even though equilibration was often achieved within the first 5 seconds (note time axis of bottom plot).

Experimentally, we were able to control the absolute concentrations of calcium and cadherin. Numerical simulation was then seeded with these known concentrations and the compiled differential equations could return 1000 time steps in ~10ms, yielding absolute concentrations of monomeric trivalent cadherin for different parameter combinations. This provided a rapid method for prediction of the absolute concentration of available monomers for

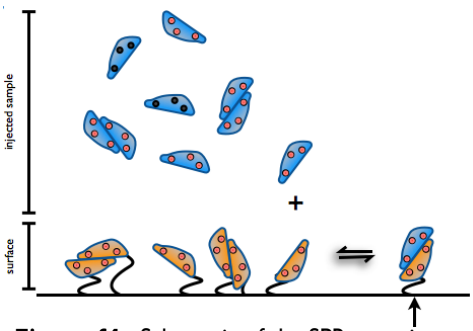


Figure f4. Schematic of the SPR experiment. Conceptually, the differential equations were run prior to injection in order to determine the available free monomer which could then interact in a standard Langmuir model with the free immobilized monomeric cadherin. Arrow indicates source of SPR signal.

interaction with immobilized cadherin.

In this way, the numerical solution of the differential equations was run at experimental concentrations for different parameter combinations and a concentration of binding competent monomer was extracted for Langmuir interaction with available monomers. Assumptions for the measurement of interaction with the immobilized cadherin included the surface existing as a mixture of fully Ca^{2+} ligated monomers and dimers, with the dimer population not participating in binding events. The computed mixture of injected sample species was assumed to be at equilibrium with no perturbation of surface bound or injected samples upon interaction. The equilibrium bound response could then be treated as a Langmuir process (Response = B_{max}*[Cadherin•$(Ca^{2+})_3$]$_{injected}$ / ([Cadherin•$(Ca^{2+})_3$]$_{injected}$ + $K_{dimerization}$). Importantly, as the $K_{dimerization}$ for numerical simulation was the same parameter used in the Langmuir model, this left only three parameters for minimization (K_{dim}, K_{Ca}, and B_{max}). By assaying different concentration vectors on the surface of the isotherm and fitting the steady steady state response, one dimensional isotherms can be obtained as in figure f5. While this data is unpublished, we were able to map much of this surface. Interestingly, the predicted

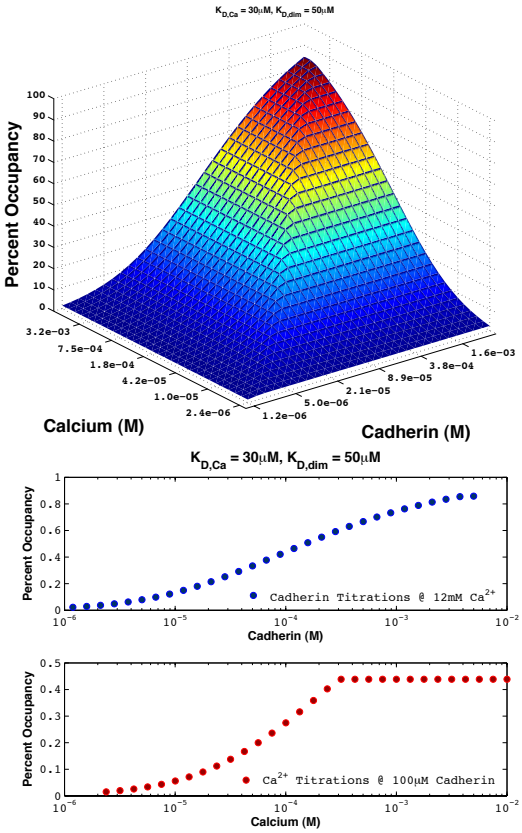


Figure f5. Model of binding to the surface for a calcium binding affinity of 30μM and dimerization affinity of 50μM. Experimentally, steady state response points in calcium and cadherin concentration space are measured, and the surface is adjusted by least squares minimization of the two binding constants. For individual vectors, protein titrations should appear non-cooperative, while calcium titrations should rise sharply before an abrupt saturation.

abrupt plateau for a calcium titration at constant cadherin was not observed for the WT protein. In order to fit the measured binding surface, it was necessary to include the potential for two dimerization binding events, strongly suggesting that cadherin dimerization occurs between sub-saturated cadherin monomers at a weaker affinity than the fully ligated monomers.

While this system warrants further investigation, proof of the design principle was made possible by fitting of the Hill equation and straightforward comparison.

Design of a Photoswitchable Cadherin

Ryan S. Ritterson[1,2,*], Kristopher M. Kuchenbecker[1], Michael Michalik[2,†], and Tanja Kortemme[1,2].

[1]Graduate Group in Biophysics & [2]California Institute for Quantitative Biomedical Research and Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California 94158

Corresponding Author: ryan.ritterson@cal.berkeley.edu

Abstract:

There is a growing interest in engineering proteins whose function can be controlled with the spatial and temporal precision of light. Here, we present a novel example of a functional light-triggered switch in the calcium-dependent cell-cell adhesion protein E-cadherin, created using a mechanism-based design strategy. We report an 18-fold change in apparent calcium binding affinity upon illumination. Our results include a detailed examination of functional switching via linked changes in Ca^{2+} binding and cadherin dimerization. This design opens avenues towards controllable tools that could be applied to many long-standing questions about cadherin's biological function in cell–cell adhesion and downstream signaling.

There has been considerable interest in light-based control of biological systems,[1] and successful applications include light-modulation of neuronal ion channels[2], light-switchable signaling proteins[3] and light-controlled protein targeting[4]. Light-based methods offer titratable, precise spatial and temporal regulation that has been demonstrated *in vitro*[5], in cell culture[4,6], and in whole animals.[7] Most examples of light-based control fall into one of two categories: (a) those that are genetically encoded using a recombinantly-produced protein borrowed from nature,[4] and (b) those created *via* targeted insertion of amino acids into a protein sequence and subsequent reaction with them of an exogenously introduced photoisomerizable small molecule, typically azobenzene based.[8] Azobenzene and related molecules undergo a reversible *cis–trans* isomerization when exposed to specific wavelengths of light, and this change in molecular shape can be coupled to changes in protein function. While in (a) the functional design is already provided naturally, one is both limited by the function (e.g., modulation of protein–protein binding, tuning fluorescence intensity) already encoded by the natural gene, and by the requirement to fuse the natural protein with the protein to be modulated. In contrast, the designs in (b) allow many types of functional modulation, such as changes in agonist binding, protein–protein binding, and protein folding. In this work, we used a new strategy where changes in protein–ion affinity couple to protein dimerization, in the cell–cell adhesion protein cadherin (**Figure 1A**).

Cadherins are a key family of calcium-dependent cell-cell adhesion proteins, and are divided into several subtypes, including the most commonly studied subtype, the classical cadherins. Classical cadherins, which include E-, N-, P-, R-, and C-cadherin[9], are composed of an intracellular domain, a transmembrane helix, and five, repeated, extracellular domains labeled EC1 (N-terminal, membrane distal) to EC5 (C-terminal, membrane proximal), along with three

calcium binding sites present in the loops at each extracellular domain boundary.[10,11] Calcium binding is required for cadherin function, as depletion of calcium disrupts cadherin-mediated cell adhesion;[12] the presence of calcium is suggested to rigidify the cadherin structure, allowing it to multimerize.[13] Depletion of cadherin significantly slows cell-cell adhesion,[14] and in a classic experiment, cadherin-free, non-adherent cells transfected with cadherin acquire morphological similarities to naturally adherent cells.[15]

Our approach to creating a light-switchable cadherin aimed to modulate its calcium-binding affinity. Because calcium binding is essential for cadherin multimerization, we reasoned that reversibly changing calcium binding affinity would be an effective way to also modulate cadherin adhesive function (**Figure 1A**). We designed cysteine residues into the protein to serve as conjugation sites for an azobenzene-based photoisomerizable chromophore, BSBCA (**Figure 1B)**. BSBCA has been used in previous applications,[5,16] demonstrating reversible switching between the *cis* and *trans* states when exposed to 370 nm (near UV) and 550 nm (green) light, respectively.[5] Our strategy involved conjugating both ends of the chromophore to the calcium-binding loops between cadherin domains EC1 & EC2, as these calcium sites have previously been shown to be most critical for function.[11] In addition, because the calcium binding sites are located in loop regions, and bind calcium with relatively weak affinities near 20 μM[17], we reasoned it would be easier to induce conformational changes affecting calcium binding there than in more rigid secondary structural elements or well-packed core regions of the protein.

Because BSBCA spontaneously crosslinks cysteine residues,[18] the design challenge presented here can be generalized as the problem of finding the best pair of residues to mutate to cysteine. In practice, however, an enormous number of pairs are possible, the overwhelming

majority of which are likely to be non-functional. We took a sequential and computational approach to identifying likely functional pairs (**Figure S1 and S1**). First, we used the program Rosetta[19] to computationally mutate all residues in four representative E-cadherin structures (PDB identifiers: 1FF5, 1EDH, 2O72 and 1Q1P) to alanine (the simplest mutation) and then calculated the predicted change in fold stability using a protocol we developed previously.[20,21] Residues with predicted destabilization >1 kcal/mol were removed from consideration, as they were presumed disruptive. Next, we narrowed the pairs to those that would be geometrically compatible with the small molecule. We calculated pair-wise C_β-C_β distances between the residues remaining using the 1FF5 structure, and kept those pairs that fell in the range 17-20 Å (appropriate for the BSBCA *trans* isomer). Finally, the remaining pairs were ranked for an additional set of structural and geometric constraints **(Figure 1C, Supporting Information).** Eleven high-ranking pairs (**Table S1**) were cloned, expressed, and tested for conjugability (**Figure S2**), photoswitchability, and functionality (**Supporting Information**). One pair, K129C/D138C (**Figure 1C-red residues**) showed the best switchability and functionality, and was further characterized in detail.

We focused on, and expressed, the first two domains of E-cadherin (EC12), because they contain the homodimeric binding interfaces[22] and they are the specificity determining domains,[23] making them most principally responsible for cadherin's function. Additionally, the shortened EC12 construct can be readily produced in high yields in *E.coli*. EC12 contains a single native cysteine residue, which we mutated to alanine (C9A), previously shown not to affect cadherin function.[24]

We first sought to show that our EC12 constructs conjugated with *trans* BSBCA could undergo isomerization to *cis*. Unconjugated BSBCA in the *trans* state has an absorbance peak

near 370 nm that decreases when illuminated at this wavelength, resulting in a population that is 80-90% *cis*; subsequent illumination at 500-550 nm will reverse the isomerization and produce a population that is >90% *trans*. Illuminating our conjugated protein (X-EC12) at 365 nm with a hand-held LED showed the characteristic reduction in absorbance of the *trans* state. The reverse isomerization (pan-visual illumination, including 500-550 nm bands) also behaved as expected, leading to a reappearance of the absorbance band of the *trans* state. We illuminated X-EC12 for 10 complete UV-green illumination cycles without any apparent loss of absorbance or switchability (**Figure 1D**); the switchability was also titratable via shorter illumination times (**Figure S3**).

We next tested whether isomerization changes calcium binding affinity. To do so, we used a previously described mass spectrometry-based assay[17] to directly measure calcium binding affinity of WT C9A as well as *trans* and *cis* X-EC12 (**Figures 2A & S4;** these assays used a cadherin concentration of 2 μM, significantly below a previously measured homodimeric K_d of WT cadherin (98.6 ± 15.5 μM)),[22] to avoid potential complications due to cadherin dimerization). If isomerization alters calcium binding, the *cis* X-EC12 should have weaker affinity than *trans*. In addition, because EC12 binds 3 calcium ions, any of which could be interfered with, a decrease in apparent cooperativity as measured by the Hill coefficient (N_h) would be expected. WT C9A cadherin specifically bound three calcium ions with a dissociation constant K_d = 28.5 ± 1.9 μM (throughout the text, errors are the boundaries of a 95% confidence interval unless otherwise indicated) and extensive cooperativity with N_h = 2.85 ± 0.47, close to previously reported numbers[17] of K_d = 20 ± 0.7 μM and N_h = 2.6 ± 0.2. *Trans* X-EC12 showed 2-fold weaker affinity and less cooperativity, with K_d = 55.2 ± 5.8 μM and N_h = 1.80 ± 0.35, but also bound three calcium ions. In contrast, calcium binding to *cis* X-EC12 was dominated by

non-specific binding. By using quadruple and higher calcium-bound states from *trans* X-EC12 (**Figure S5**) as a reference for non-specific binding, we subtracted the estimated contribution of non-specific calcium binding from the measured average calcium occupancy for *cis* X-EC12 (**Supporting Information**). The resulting line shows significantly reduced binding compared to *trans*. (A quantitative fit was not possible due to required calcium concentrations being higher than the dynamic range of the assay.)

To more directly measure the calcium binding of *cis* X-EC12, we turned to a different assay that determined the *cis* half-life as a function of calcium concentration. One general caveat inherant to azobenzene-based strategies is that switching to the *cis* state is generally incomplete, i.e. the *cis* state always contains a minor *trans* population.[8,18] However, the entirety of any change observed in half-life experiments is due only to the *cis* subpopulation, allowing measurement of pure *cis* properties unaffected by the small fraction that remains *trans*. Therefore, if chromophore isomerization significantly affects calcium binding in our conjugated constructs, with stronger binding of calcium to *trans* X-EC12, by thermodynamic coupling we would expect to see a change in *cis* X-EC12 half-life with calcium (**Figure 2B**). The *cis* state is thermodynamically unstable and *cis* BSBCA relaxes back to the stable *trans* state in the dark with a half-life of approximately 20 minutes at room temperature,[18] although conjugation to proteins can alter chromophore half-lives.[5,25] By observing the increase in absorbance at 370nm during relaxation of our conjugated constructs back to *trans*, one can compute the half-life of the process (**Figures S6, S7,** and **Supporting Information;** these assays used a protein concentration of 12 μM to minimize a potential change in half-life due to protein dimerization). The half-life should decrease with increasing calcium concentration as *trans* X-EC12 becomes stabilized by calcium binding. As expected, we observed a half-life decrease from approximately

72 minutes to 28 minutes, with an EC_{50} of 996 ± 135 μM calcium. This represents a nearly 18-fold change in apparent calcium binding affinity from the 55 μM for *trans* X-EC12 (mass spectrometry analysis, **Figure 2A**).

We also observed a cooperative transition in half-life duration, with a measured Hill coefficient (N_h) of 2.4 ± 0.74. In showing interdependence between isomerization and calcium binding, these results indicate that, as expected, isomerization of the chromophore significantly weakens calcium binding. In addition, the cooperative nature observed indicates multiple calcium ions are binding simultaneously during the transition from *cis* to *trans*, hinting that the *cis* state likely weakens multiple binding sites.

After successfully demonstrating photoswitchable calcium binding in our engineered cadherin, we next asked whether the change in calcium binding affinity also results in the expected change in protein binding activity. We used surface plasmon resonance (SPR) to measure protein homodimerization as a function of calcium concentration **(Figures 3, S8, S9** and **Supporting Information**). In this assay, similar to that of Harrison and colleagues[22], biotinylated WT C9A cadherin was immobilized to the SPR chip and WT C9A, *trans* or *cis* X-EC12 were flowed over it. Direct measurements of both calcium affinity and homodimeric protein affinity in SPR are difficult due to solution homodimers competing with those on the surface, reducing the effective protein concentration; to minimize solution homodimerization, we used a protein concentration (40 μM) below the K_d for homodimerization of WT EC12 cadherin[22]. Additionally, *cis* measurements are of mixed populations due to the inability of reaching full conversion to the *cis* state and some thermal relaxation to *trans* during the experiment, limiting the observable fold change in affinity (**Supporting Information**). We observed a calcium binding EC_{50} for C9A cadherin as measured by a single Hill fit of 72.0 μM

(with mean fit values and ±2 SD error, as measured by a bootstrapping analysis of the data, of 71.2 ± 14 μM, see **Supporting Information & Figures S10, S11**) and a Hill coefficient (N_h) of 2.24 (2.45 ± 1.7). In comparison, *trans* X-EC12 has an EC_{50} of 156 μM (170 ± 33 μM), with N_h = 1.38 (1.28 ± 0.28). These EC_{50} values are higher than those measured in the mass spectrometry assay (**Figure 2A**), which is due to cadherin binding multiple calcium ions to function, causing any measured EC_{50} to necessarily be, at a minimum, a multiple of the protein concentration used, which here was 40 μM. Strikingly, *cis* X-EC12 showed substantially weakened binding (**Figure 3A**), with EC_{50} = 619 μM (611 ± 180 μM) and N_h = 0.76 (0.77 ± 0.15), demonstrating a nearly 4-fold change in calcium affinity under these conditions. (Note: non-specific protein binding to the SPR chip appeared at calcium concentrations >2 mM for WT C9A and *trans* X-EC12 – see **Supporting Information**). The change in protein-protein binding was also reversible as measured over multiple illumination cycles with 40 μM protein and 1mM Ca^{2+} (**Figure 3B**).

An alternative explanation for the observed decrease in the SPR signal upon isomerization to the *cis* state could be an increase in *cis* homodimerization in solution, effectively reducing the concentration of X–EC12 cadherin monomers available to bind to the WT cadherin immobilized on the chip. To exclude that possibility, we analyzed X-EC12 cadherin homodimerization via gel filtration. The observed decrease in the dimer/monomer ratio after UV illumination additionally confirms the expected weaker *cis* homodimerization upon illumination (**Figure S12**).

Questions remain about the structure of the functional cadherin multimers, including evidence that cadherin forms strand-swapped dimers.[22,26,27] While we cannot directly determine

the structure of the interacting species formed in our SPR experiments, each set of SPR traces for a given cadherin variant can be fit to a single off rate returning to baseline levels, even at higher calcium concentrations (**Figure S13**). This behavior is consistent with a single dimer type formed for each assayed cadherin variant.

Taken together, our results demonstrate the successful design of a reversibly photoswitchable cadherin. When illuminated with light, its calcium binding affinity changes from 56 μM to 996 μM, a nearly 18-fold change, and this change in affinity is coupled to a change in protein–protein binding. One constraint on our current design is the present inability of this chromophore to switch completely to *cis*. Several new chromophores have become available[25,28,29] that possess either more complete isomerization or longer half-lives that could allow for isolation and use of the pure *cis* state.

When applied in cell culture experiments, the light-modulatable cadherin could help answer several outstanding questions about cadherin's function. One way to introduce this engineered molecule into a cellular context is via cadherin-coated substrata. Coated substrata have been used to study cell-cell adhesion[30] and stem cell pluripotency[31]. Creation of coated surfaces allows spatial control of cadherin patterning and fine control over cadherin concentration, which could help maximize switchability of cadherin-mediated adhesion.[30]

Although interest in photoswitchable proteins has increased in recent years, relatively few examples exist in the literature, perhaps because finding good cysteine attachment points remains difficult. Compared to high-throughput and other library techniques, we were able to create a successful conjugate using a rational design strategy and a small library of constructs. In our design, we chose to focus on modulating loop structures that may have a lower activation energy barrier to transition compared to more rigid parts of protein domains. While not all

proteins could be modified in this way, we believe our combined rational/computational selection method and our focus on loops can be generalized to create other photoswitchable designs.

Figure Legends:

Figure 1. (A) A cartoon showing the basis of our design. As designed, our photoswitch reduces calcium binding affinity, which, in turn, reduces homodimer affinity. (B) BSBCA undergoes a reversible cis/trans isomerization when illuminated with specific wavelengths of light. (C) EC12 structure showing the region targeted for photoswitchability. Labels indicate the design considerations. (D) Photoswitchability and reversibility measured by absorbance after many cycles of illumination of X-EC12.

Figure 2. Characterization of photoswitchable calcium binding affinity. (A) Calcium binding as monitored by mass spectrometry. While WT and *trans* X-EC12 bound three calcium ions specifically, *cis* X-EC12 showed considerably weaker, predominantly non-specific binding (**Supporting Information** and **Figure S4**). Fits are based on a model of a single class of binding site for a maximum of three specifically bound calcium ions. (B) The half-life of the *cis* state as a function of calcium concentration, as measured by absorbance. Error bars are ±1 SD from three independent experiments.

Figure 3. Characterization of photoswitch homodimeric binding. (A) Homodimeric binding monitored in SPR as a function of calcium concentration. The data were fit to a Hill equation. Faded points contain significant non-specific binding and were not used in the fits. Responses between flow cells were scaled to minimize a least squares difference, then mean values were normalized such that the fit value at $[Ca^{2+}] = \infty$ were 1.0 (**Supporting Information**). Error bars are ±1 SD of the 3 active flow cells in the instrument after scaling and normalization. Inset shows fits at low calcium concentrations. (B) Homodimeric binding monitored in SPR at 1mM Ca^{2+}, after repeated illumination cycles. Responses between flow cells were scaled to minimize a

least squares difference. Error bars are ±1 SD of the 3 active flow cells in the instrument after scaling.

Figure 1

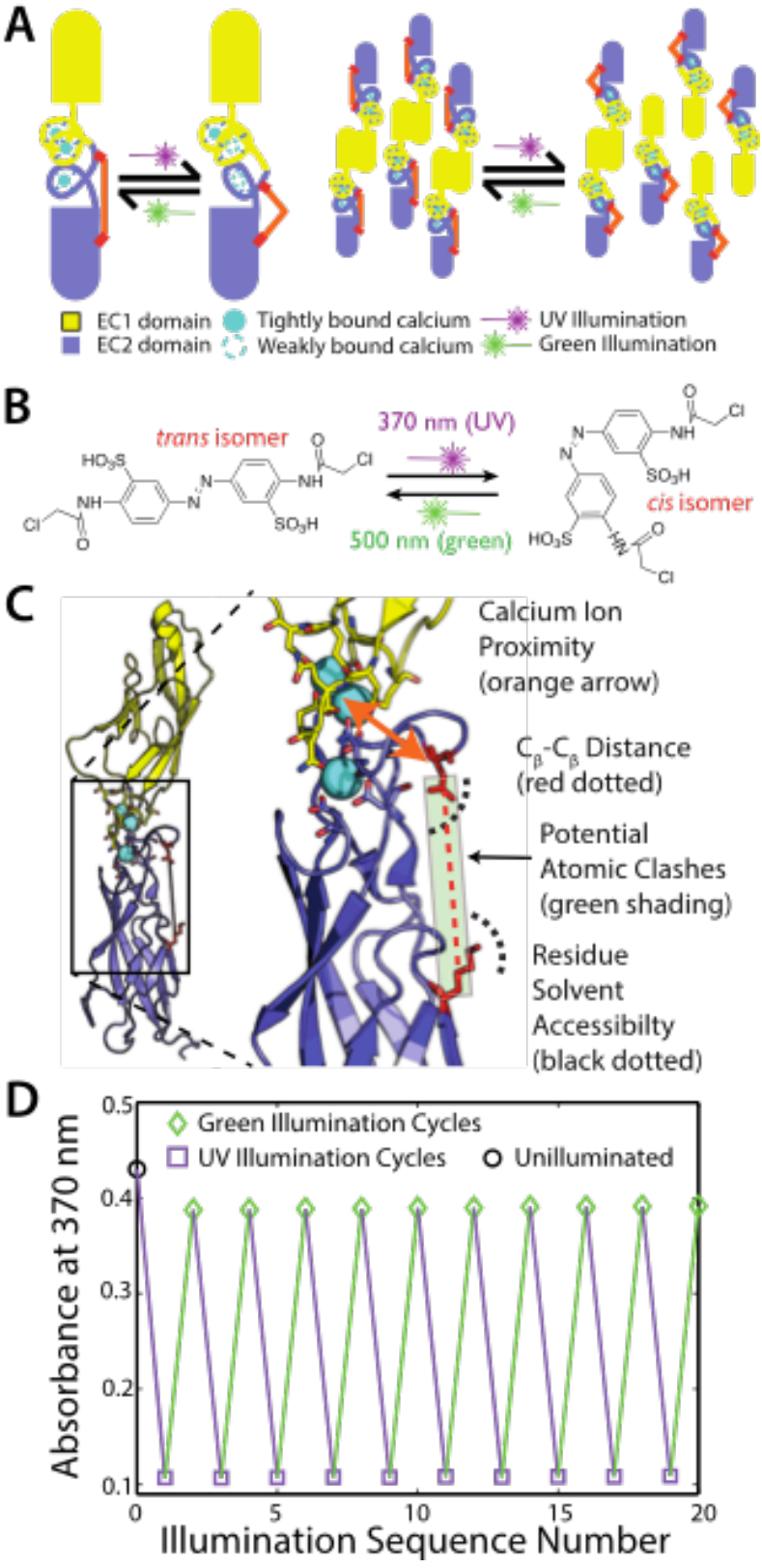

A
EC1 domain
EC2 domain
Tightly bound calcium
Weakly bound calcium
UV Illumination
Green Illumination
B
trans isomer
370 nm (UV)
500 nm (green)
cis isomer
C
Calcium Ion Proximity (orange arrow)
Cβ-Cβ Distance (red dotted)
Potential Atomic Clashes (green shading)
Residue Solvent Accessibilty (black dotted)
D
Green Illumination Cycles
UV Illumination Cycles
Unilluminated
Absorbance at 370 nm
Illumination Sequence Number

Figure 2

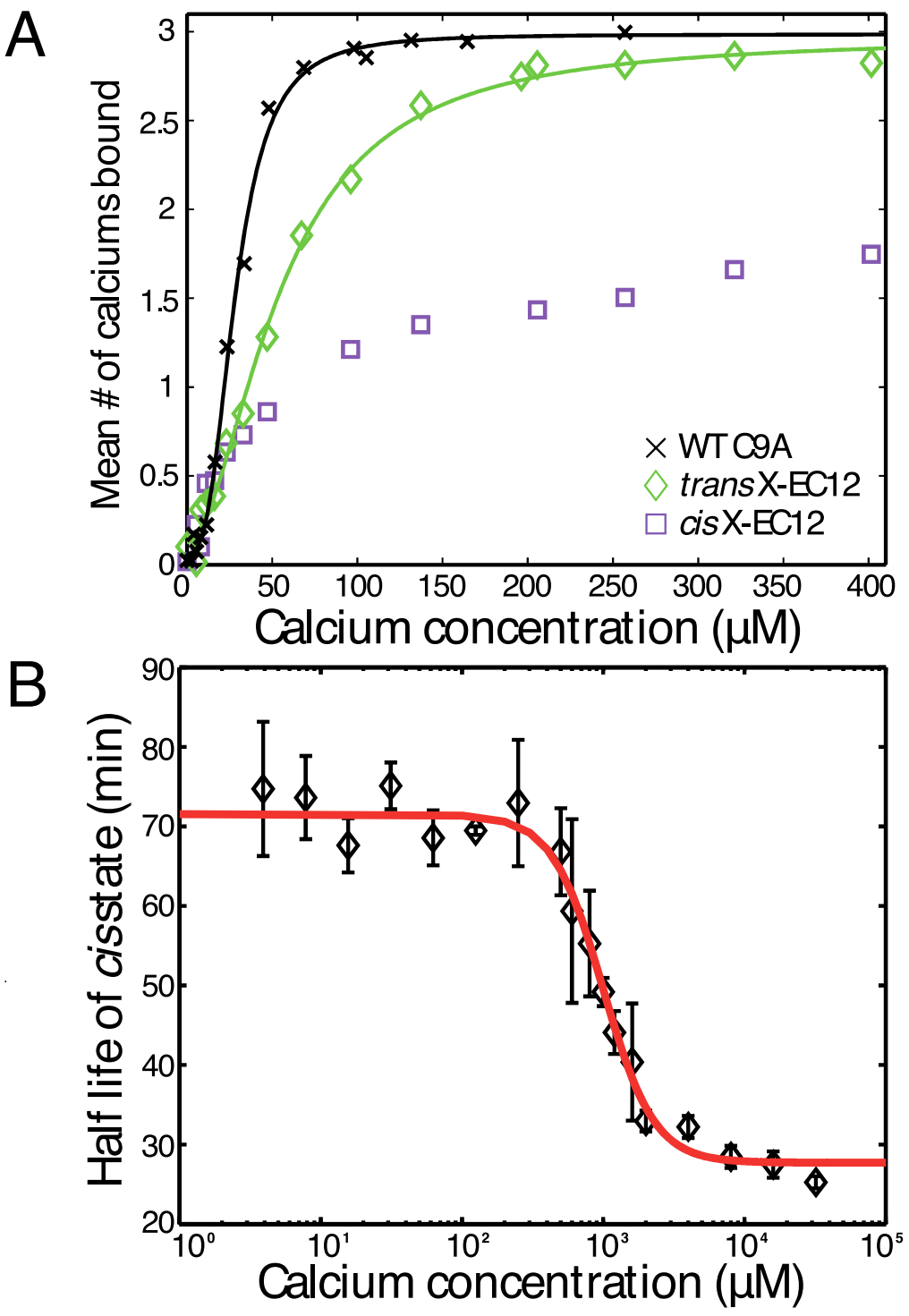
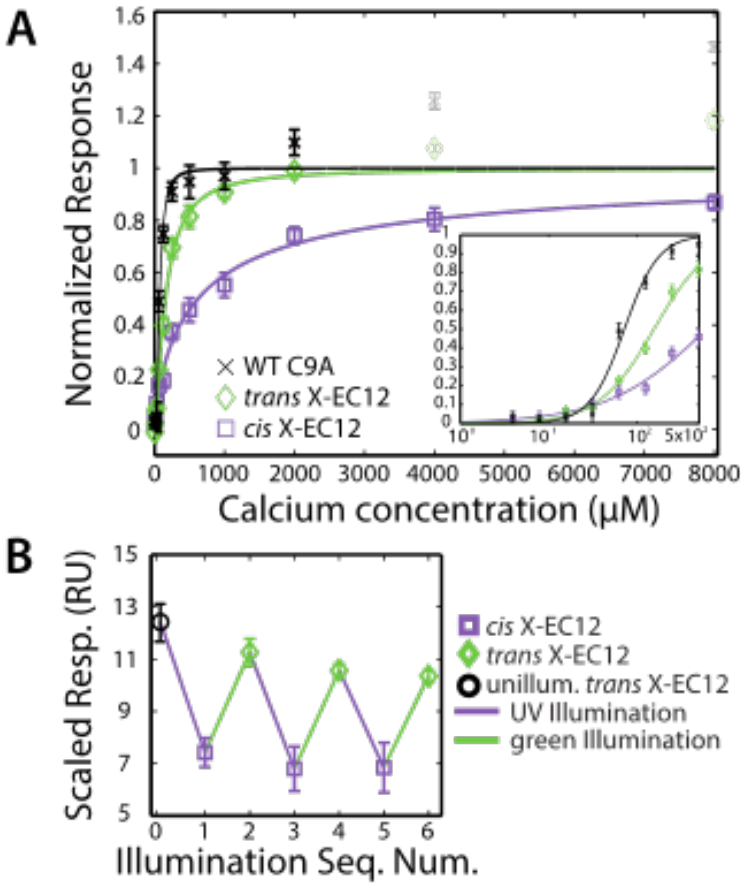

A
Mean # of calciums bound
Calcium concentration (μM)
WT C9A
trans X-EC12
cis X-EC12
B
Half life of cis state (min)
Calcium concentration (μM)

Figure 3


A
Normalized Response
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT C9A
trans X-EC12
cis X-EC12
0
1000
2000
3000
4000
5000
6000
7000
8000
Calcium concentration (μM)
B
Scaled Resp. (RU)
15
13
11
9
7
5
0
1
2
3
4
5
6
Illumination Seq. Num.
cis X-EC12
trans X-EC12
unillum. trans X-EC12
UV Illumination
green Illumination

REFERENCES

(1) Krauss, U.; Drepper, T.; Jaeger, K. E. *Chemistry* **2011**, *17*, 2552.

(2) Banghart, M.; Borges, K.; Isacoff, E.; Trauner, D.; Kramer, R. H. *Nat. Neurosci.* **2004**, *7*, 1381.

(3) Wu, Y. I.; Frey, D.; Lungu, O. I.; Jaehrig, A.; Schlichting, I.; Kuhlman, B.; Hahn, K. M. *Nature* **2009**, *461*, 104.

(4) Levskaya, A.; Weiner, O. D.; Lim, W. A.; Voigt, C. A. *Nature* **2009**, *461*, 997.

(5) Woolley, G. A.; Jaikaran, A. S. I.; Berezovski, M.; Calarco, J. P.; Krylov, S. N.; Smart, O. S.; Kumita, J. R. *Biochemistry* **2006**, *45*, 6075.

(6) Zhang, F.; Muller, K. M.; Woolley, G. A.; Arndt, K. M. *Methods Mol Biol* **2012**, *813*, 195.

(7) Wyart, C.; del Bene, F.; Warp, E.; Scott, E. K.; Trauner, D.; Baier, H.; Isacoff, E. Y. *Nature* **2009**, *461*, 407.

(8) Beharry, A. A.; Woolley, G. A. *Chem Soc Rev* **2011**, *40*, 4422.

(9) Ivanov, D. B.; Philippova, M. P.; Tkachuk, V. A. *Biokhimiya/Biochemistry* **2001**, *66*, 1174.

(10) Koch, A.; Pokutta, S.; Lustig, A.; Engel, J. *Biochemistry* **1997**, *36*, 7697.

(11) Prakasam, A.; Chien, Y.; Maruthamuthu, V.; Leckband, D. *Biochemistry* **2006**, *45*, 6930.

(12) van Roy, F.; Berx, G. *Cellular and Molecular Life Sciences* **2008**, *65*, 3756.

(13) Pokutta, S.; Herrenknecht, K.; Kemler, R.; Engel, J. *Eur. J. Biochem.* **1994**, *223*, 1019.

(14) Capaldo, C. T.; Macara, I. G. *Mol. Biol. Cell* **2007**, *18*, 189.

(15) Nose, A.; Nagafuchi, A.; Takeichi, M. *Cell* **1988**, *54*, 993.

(16) Guerrero, L.; Smart, O. S.; Woolley, G. A.; Allemann, R. K. *J. Am. Chem. Soc.* **2005**, *127*, 15624.

(17) Courjean, O.; Chevreux, G.; Perret, E.; Morel, A.; Sanglier, S.; Potier, N.; Engel, J.; van Dorsselaer, A.; Feracci, H. *Biochemistry* **2008**, *47*, 2339.

(18) Burns, D. C.; Zhang, F.; Woolley, G. A. *Nat. Protoc.* **2007**, *2*, 251.

(19) Rosetta - The premier software suite for macrommolecular modeling. http://www.rosettacommons.org (accessed May 16th, 2013)

(20) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14116.

(21) Kortemme, T.; Kim, D.; Baker, D. *Sci. Signaling* **2004**, *2004*, pl2.

(22) Harrison, O. J.; Bahna, F.; Katsamba, P. S.; Jin, X.; Brasch, J.; Vendome, J.; Ahlsen, G.; Carroll, K. J.; Price, S. R.; Honig, B.; Shapiro, L. *Nat. Struct. Mol. Biol.* **2010**.

(23) Nose, A.; Tsuji, K.; Takeichi, M. *Cell* **1990**, *61*, 147.

(24) Troyanovsky, R. B.; Sokolov, E.; Troyanovsky, S. M. *Mol. Cell. Biol.* **2003**, *23*, 7965.

(25) Samanta, S.; Woolley, G. A. *Chembiochem* **2011**, *12*, 1712.

(26) Troyanovsky, R. B.; Laur, O.; Troyanovsky, S. M. *Mol. Biol. Cell* **2007**, *18*, 4343.

(27) Häussinger, D.; Ahrens, T.; Aberle, T.; Engel, J.; Stetefeld, J.; Grzesiek, S. *EMBO J* **2004**, *23*, 1699.

(28) Blanco-Lomas, M.; Samanta, S.; Campos, P. J.; Woolley, G. A.; Sampedro, D. *J. Am. Chem. Soc.* **2012**, 120406140828000.

(29) Samanta, S.; Qin, C.; Lough, A. J.; Woolley, G. A. *Angew Chem Int Ed Engl* **2012**, *51*, 6452.

(30) Borghi, N.; Lowndes, M.; Maruthamuthu, V.; Gardel, M. L.; Nelson, W. J. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13324.

(31) Nagaoka, M.; Koshimizu, U.; Yuasa, S.; Hattori, F.; Chen, H.; Tanaka, T.; Okabe, M.; Fukuda, K.; Akaike, T. *PLoS ONE* **2006**, *1*, e15.

ACKNOWLEDGMENT

The authors would like to thank Laura Lavery and Rebeca Choy for contributing to the original idea, Elizabeth Tanner, Drs. Matthew Banghart, Andrew Woolley, James Nelson and Nicolas Borghi for helpful discussions, Alec Nielsen for assistance with assay development, and the Mass Spectrometry Resource at UCSF, NIH NIGMS P41GM103481 (A.L. Burlingame). R.R. and K.K. were partially funded by NIH grant T32 GM008284. Research funding for this project came from NSF grant 1134127.

Supporting information for: Design of a Photoswitchable Cadherin

Ryan S. Ritterson, Kristopher M. Kuchenbecker, Michael Michalik, and Tanja Kortemme.

Table S1. Characterization of library mutants

Mutant	Conjugability*	Stability*	Switchability*	Half Life (min)	Change in Ca^{2+} -dependent binding upon illumination (SPR)
129/138	**High**	**High**	**80%**	**72**	**Significant**
70/14	Poor	High	51%	36	Non-Specific binding
70/107	High	High	100%	56	No Change
5/137	High	Poor	68%	44	No Binding
6/90†	Moderate	Poor	51%	ND	ND
111/135†	High	Moderate	62%	ND	ND
16/57†	High	Poor	45%	ND	ND
70/105†	ND	Poor	ND	ND	ND
70/133†	Poor	Poor	ND	ND	ND
95/139†	Very Poor	Moderate	ND	ND	ND
70/138†	Poor	Moderate	ND	ND	ND

ND: not determined.

* Conjugability was assessed qualitatively *via* mass spectrometry by noting the fraction that remains unconjugated. Stability was assessed qualitatively by noting appearance of protein precipitates or aggregation peaks during size exclusion chromatography. Switchability was assessed *via* absorbance assays (see Methods section).

† These mutants were tested with an earlier version of the EC12 construct that contains a non-native Met residue at the N-terminus (see Constructs section in Methods).

Supplementary Materials

Figure S1. Flow chart showing the computational methodology used to create the mutant library.

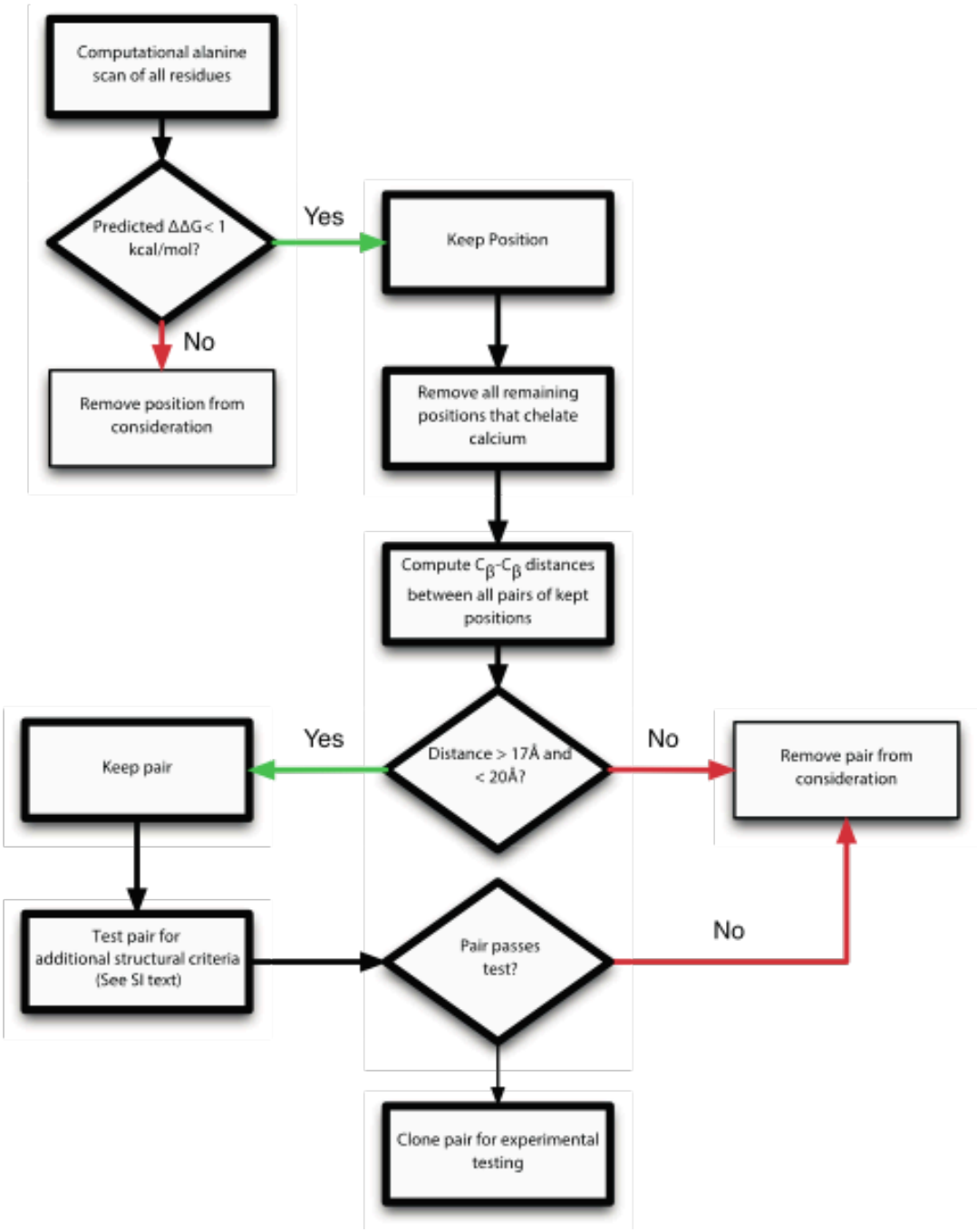


Figure S2. Mass spectra showing steps during protein production. A and C are overlays of two spectra. (A) Demonstration of mass change after cleavage. Black, uncut WT C9A protein; blue, WT C9A protein cleaved with TEV protease. Expected monoisotopic mass change was Δ1747 Da, compared to Δ1739 Da observed. (The difference could be due to imperfect protonation state prediction, such as for the 6xHis tag.) (B) Verification of Biotinylation. The majority of the protein is biotinylated. Expected and observed mass changes were both Δ226 Da. (C) Demonstration of mass change during conjugation: red, unconjugated mutant protein; green, conjugated mutant protein. Expected and observed mass changes were both Δ453 Da.

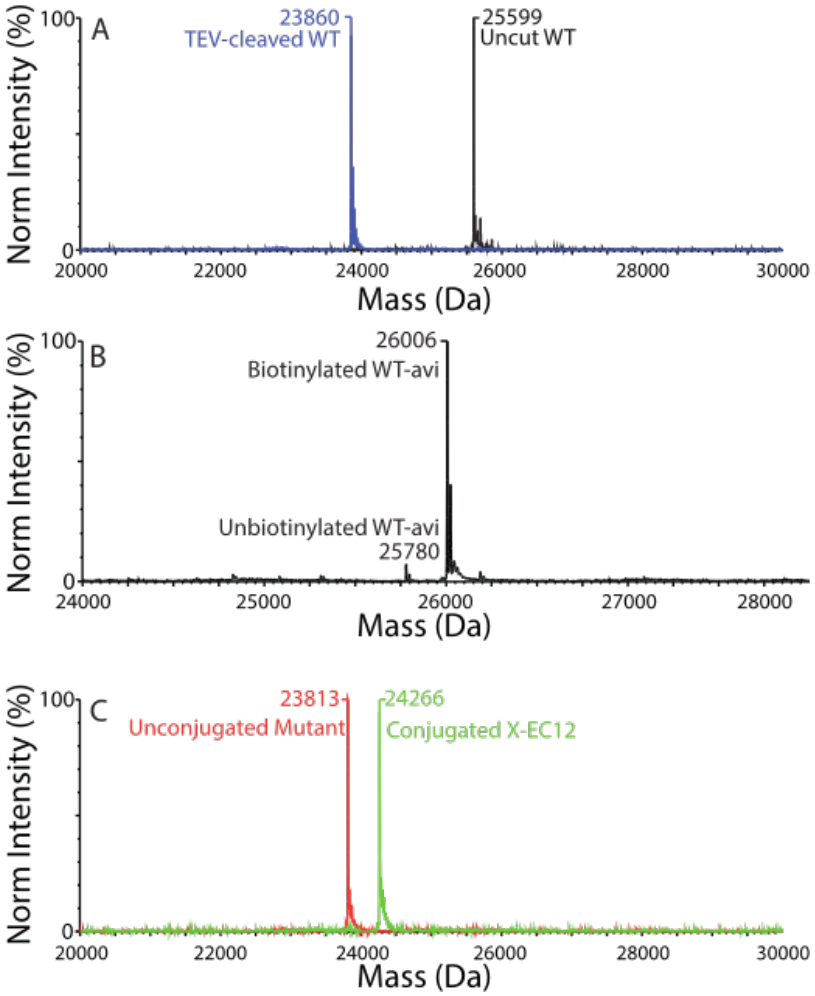


Figure S3. The extent of photoswitching is titratable. X-EC12 was illuminated with either UV light (purple lines) or green light (green lines) for various amounts of time, and the 370 nm absorbance was measured afterward. Shorter illuminations isomerize a smaller fraction of the protein.

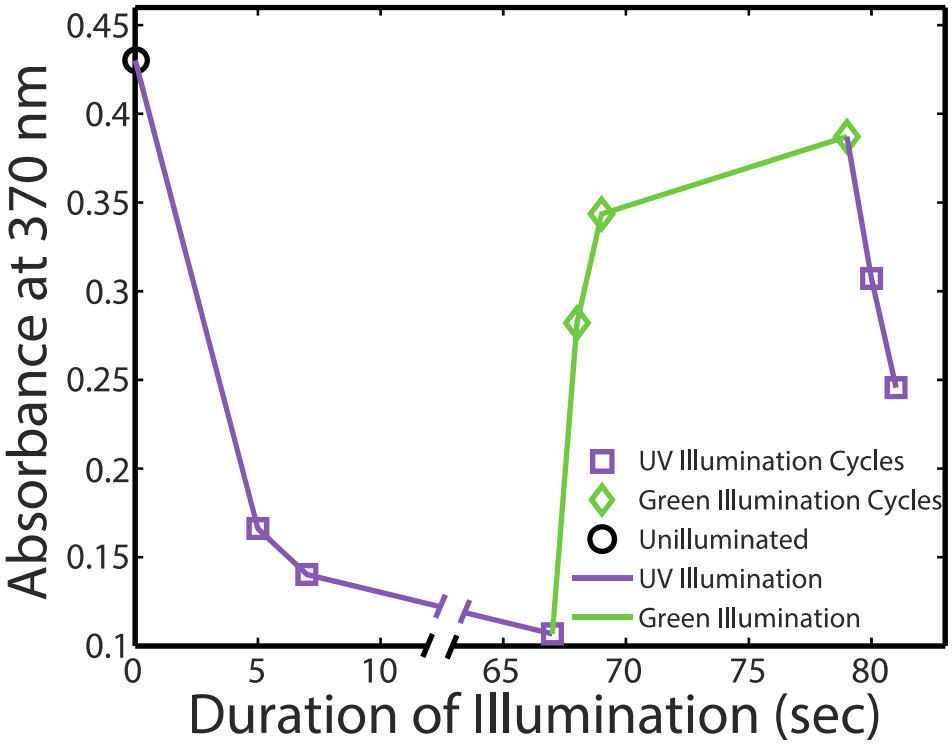


Figure S4. Calcium affinity as measured by mass spectrometry. Figure is the same as **Figure 2A**, with the addition of the *cis* state prior to subtraction of nonspecific binding (grey) and the unconjugated mutant (red). For the unconjugated mutant, K_d = 145 ± 18 μM and N_h = 1.07 ± 0.13. We speculate that the mutations increase the flexibility of the unconjugated protein, which weakens calcium binding affinity compared to WT. Subsequent addition of the *trans* crosslinker could re-rigidify the protein, allowing stronger calcium binding.

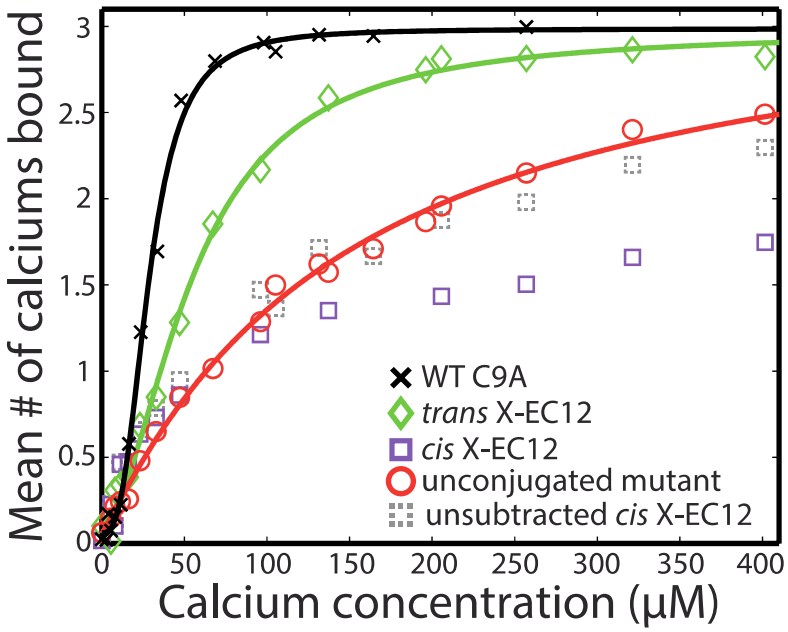


Figure S5. Example mass spectra showing calcium binding to *trans* X-EC12 at a calcium concentration of 96 μM. Unlabeled intermediate peaks are +1 Na.

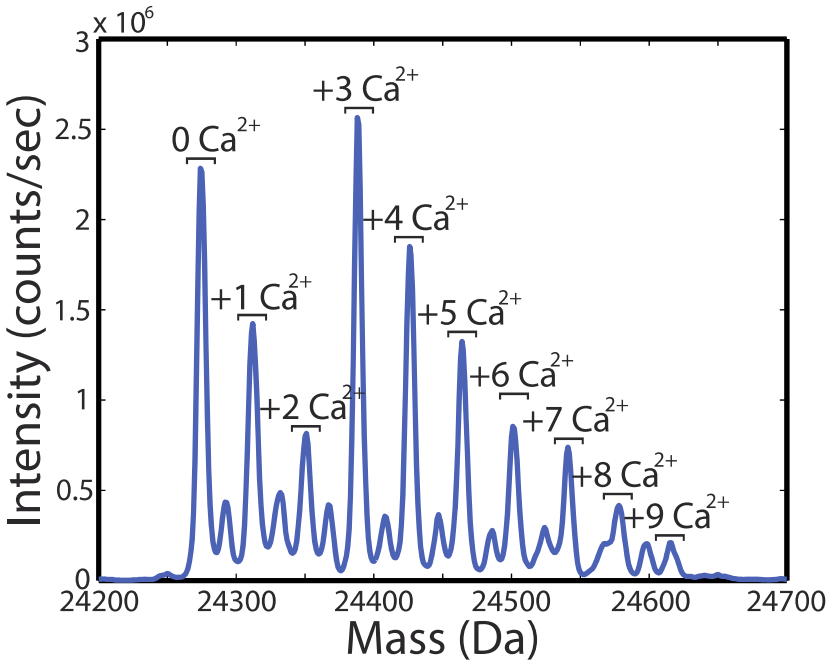


Figure S6. Illumination and relaxation of photoswitchable protein. Prior to illumination, X-EC12 is in the *trans* state (black). After illumination at 370 nm, much of the protein becomes *cis,* with a corresponding decrease in absorbance near 370 nm and a slight increase near 450 nm (darkest purple). Over time, *cis* X-EC12 relaxes to *trans*, with an increase in absorbance at 370 nm.

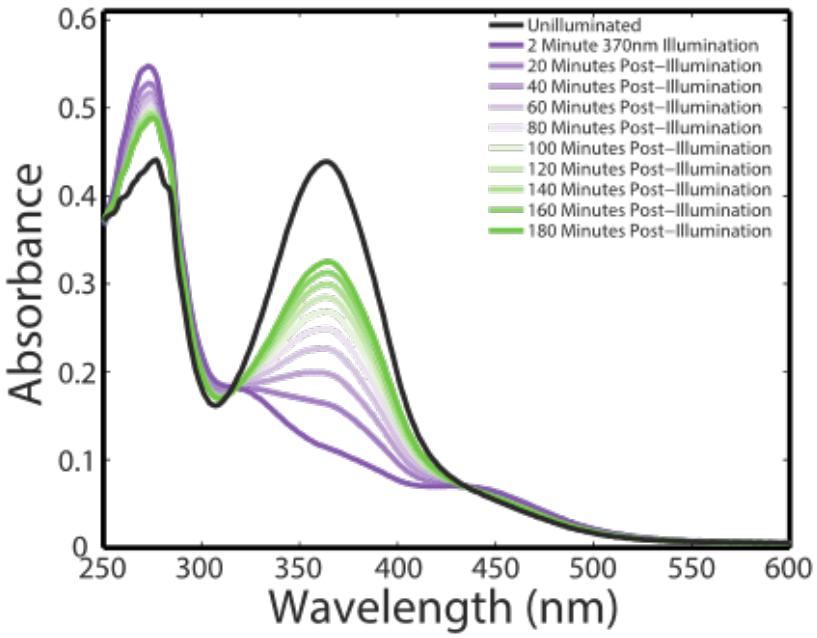


Figure S7. Example fit to single exponential of the relaxation of conjugated protein after illumination ([Ca^{2+}] = 500 μM). For this illumination, the fit gave a half-life of 63.4 min.

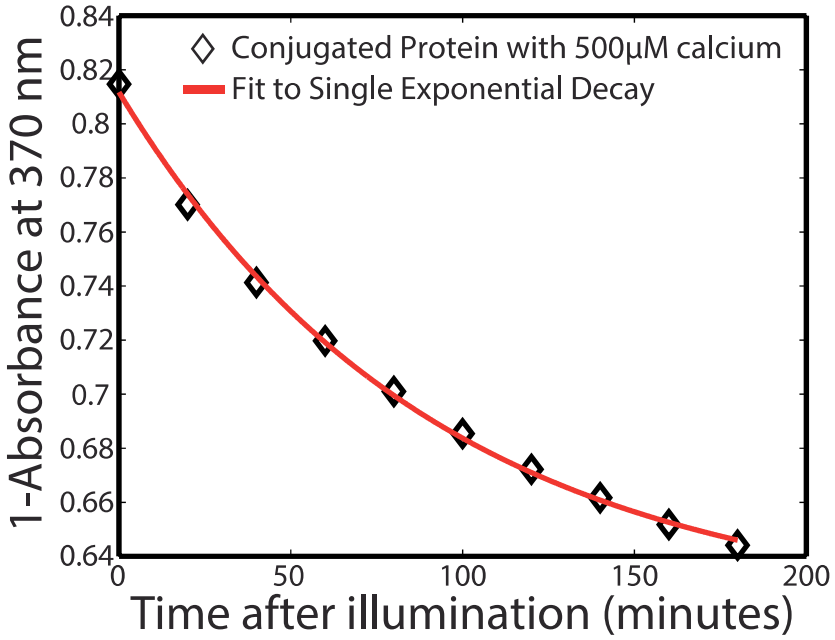


Figure S8. SPR calcium titrations as in **Figure 3A**, with the addition of the unconjugated mutant, red. For the unconjugated mutant, EC_{50} = 104 ± 27 μM and N_h = 1.37 ± 0.47.

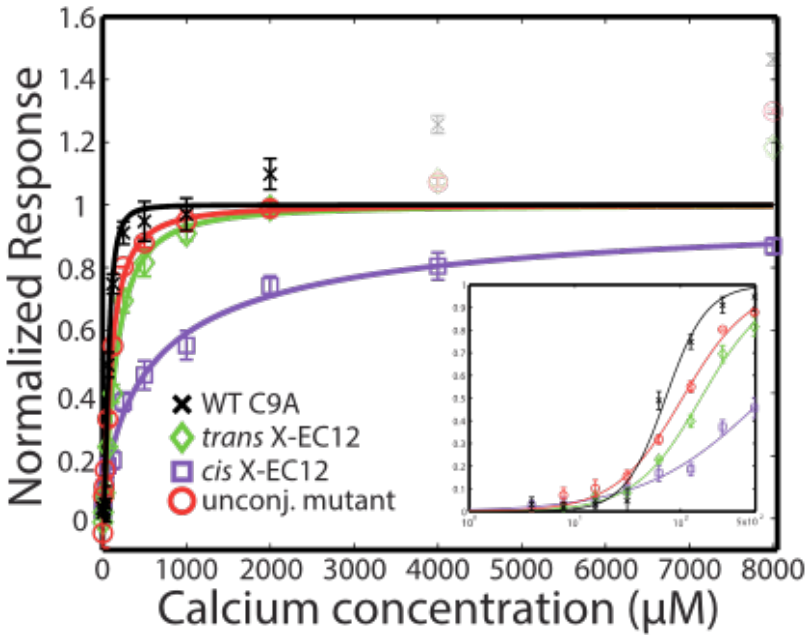


Figure S9. SPR traces from experiment in **Figure 3A**. Data are shown directly from instrument, prior to any post processing, blank subtraction, adjustment for surface response decay, scaling, or normalization. (Note: *cis* X-EC12 was collected first, leading to a higher response than *trans* X-EC12 due to surface decay)

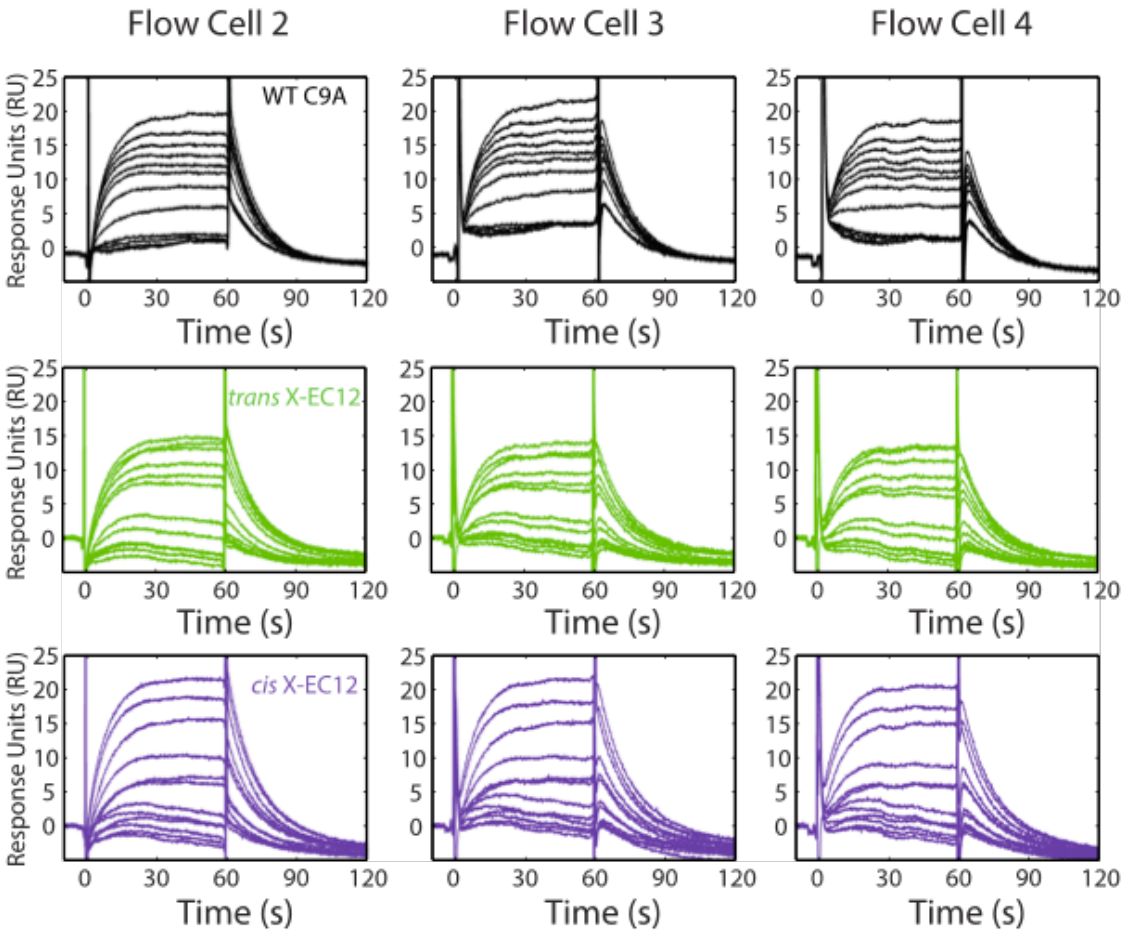


Figure S10. Plot of values resulting from the bootstrap analysis of the SPR calcium titration data. All 500 points for each protein plotted.

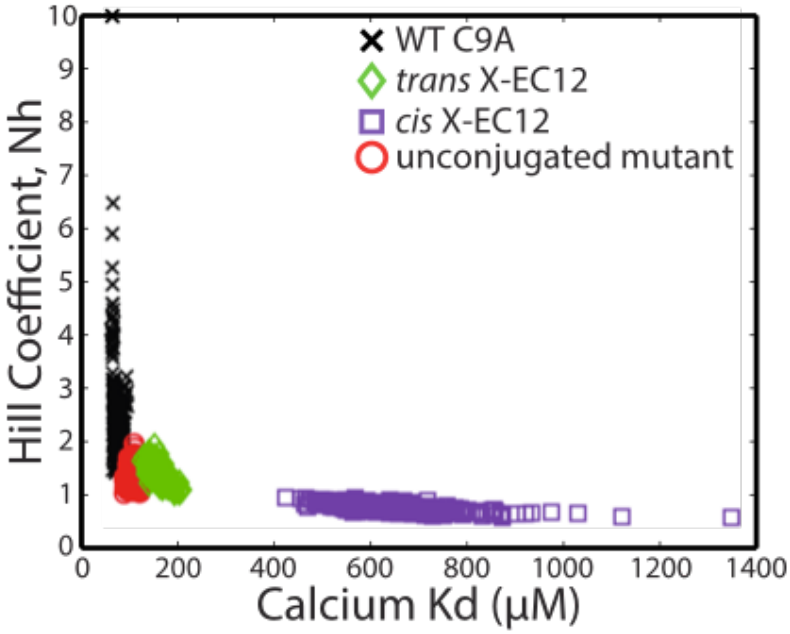


Figure S11. Plot of values resulting from the bootstrap analysis of the SPR calcium titration data. Only shown are the top 100 points by minimum sum of squared error (SSE) for each protein used to compute the mean and ±2 SD values.

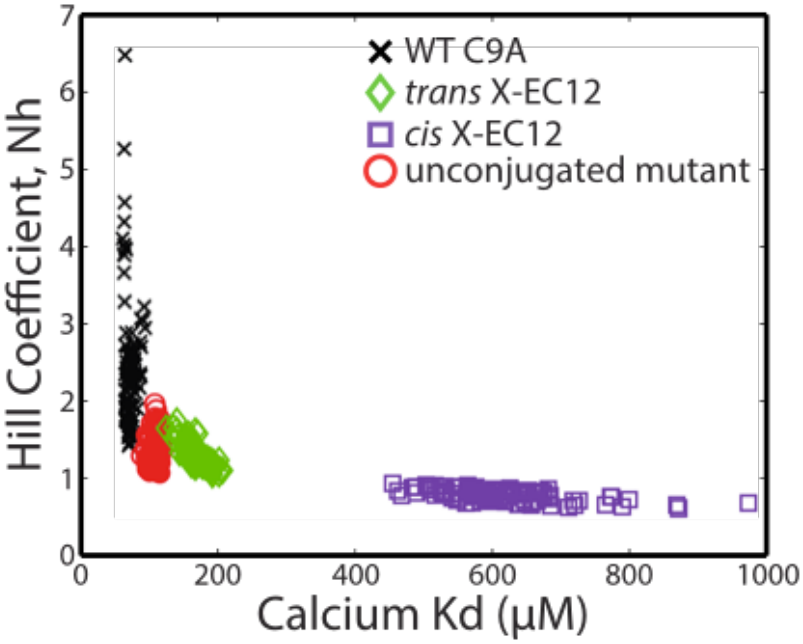


Figure S12. Size exclusion chromatography of *cis* and *trans* X-EC12. Injected samples contained 250 μM protein and 2500 μM Ca^{2+}. Chromatograms have been normalized to the height of the largest peak (the monomer) to allow easier comparison of relative peak fractions.

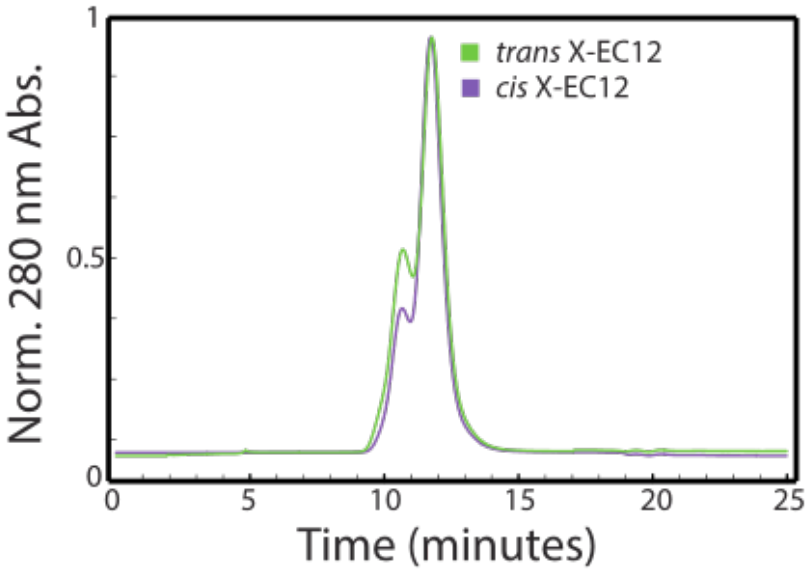
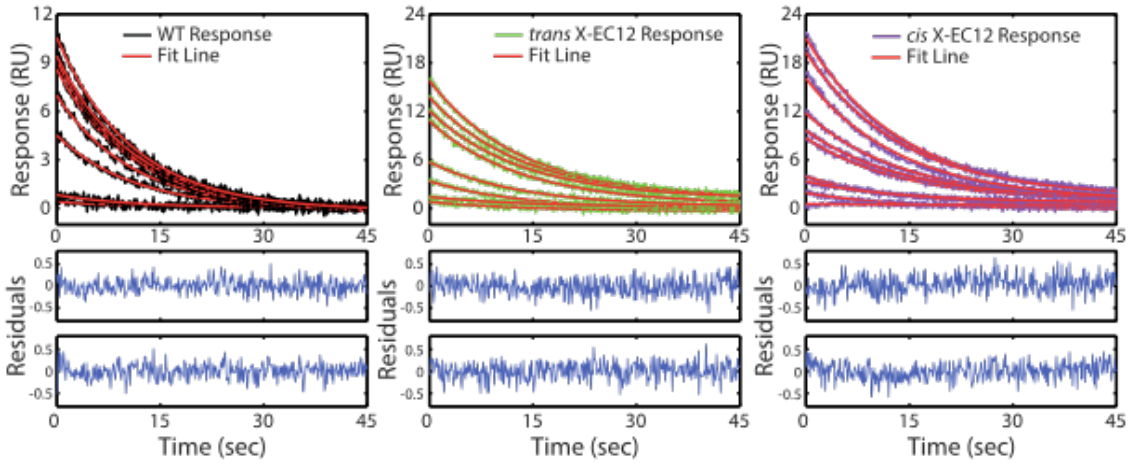


Figure S13. Dissociation of cadherin dimers from experiment in **Figure 3A**. Top panels show dissociations and fits with a common off rate to a single exponential for each protein. Bottom panels show representative residuals of fits, shown for [Ca^{2+}] = 62.5 μM (upper) and 1000 μM (lower). Reponses are from flow cell 2.



Methods

Design Methodology

As described in the main text, the design problem here can be reduced to finding the best pair of cysteine attachment sites for the chromophore, such that when conjugated, the system both maintains its native function, but also shows high switchability and functional changes when illuminated. (**Figures 1C** and **S1**) Our design process involved first finding residues likely to be mutatable using the computational design program Rosetta.[1] We used a previously described method[2,3] to predict the change in fold stability of mutating, one by one, every residue in four representative E-cadherin structures (PDB identifiers 1FF5, 1EDH, 2O72 and 1Q1P) to alanine, the simplest mutation. Residues that had an average predicted change in fold stability > 1 kcal/mol upon alanine mutation were removed from consideration. We also removed all residues that directly bound calcium from consideration, reasoning that mutating them would likely cause a significant change in cadherin's calcium affinity. Next, we calculated all pair-wise C_β-C_β distances between the residues remaining using the 1FF5 structure, and kept those pairs that fell in the range 17-20 Å (appropriate for the BSBCA *trans* isomer). These first two steps reduce the potential double mutants from >20000 pairs to 1513.

Following these initial steps, we further restricted the number of pairs via a series of additional structural criteria. First, we restricted the remaining pairs to those that had at least one endpoint within 20 Å of a calcium ion as measured by the molecular graphics program PyMol[4], reasoning that pairs with both ends distant from the calcium binding sites would be unlikely to have an effect on calcium binding. This further reduced the number of pairs to 1120. Next, to only use residues that were accessible for chromophore conjugation, we restricted

pairs to those where both endpoints were in residues that had >30 Å^2 of solvent accessible surface area (SASA) when the protein was in the monomeric state, again measured using PyMol. This reduced the set of pairs to 272. Finally, we plotted the remaining pairs on the 1FF5 PDB structure and eliminated those where the addition of the chromophore would sterically interfere or clash with the native protein structure, as estimated by using the surface representation of the protein structure and eliminating those pairs whose path of shortest distance penetrated the surface. This step eliminated a large part of the remaining set, leaving 28 pairs of conjugation sites. As a final step, we manually examined the 28 pairs and chose a set of 10, seeking to build a diverse set of conjugation sites to maximize the chances of finding a functional photoswitch. One mutant pair, P5C/D137C, did not satisfy the structural criteria; it was chosen manually in an attempt to construct a photoswitch coupled to β-strand one, based on the previously observed functional importance of the strand to cadherin dimerization.[5,6]

Constructs

Natural E-cadherin contains a pro-peptide cleaved during protein maturation, leaving an N-terminal aspartate residue instead of a methionine. This aspartate forms an important intramolecular salt bridge, resulting in recombinantly produced proteins containing an N-terminal methionine having altered function.[7] To remove both the exogenous 6x-His tag and methionine, we introduced a TEV protease cleavage site. TEV protease is compatible with the reducing buffer conditions (containing 3 mM TCEP) required to maintain free cysteines in our engineered constructs prior to conjugation, whereas other common proteases are not.

Met-EC12-6xHis was cloned into pET-22b(+) (EMD Millipore # 69744-3) as described previously.[8] 6xHis-TEV-EC12 contained a 6xHis tag and TEV cleavage site (ENLFYQ) added N-

terminally, as well as an inserted C-terminal stop codon between EC12 and the C-terminal 6xHis tag using the 5' oligonucleotide (GACCCCATATGCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGGACTGGGTC) (blue-6xHis, orange-TEV cleavage site) and the 3' oligo (TGGTGCTCGAGTCAAGGAGCGTTGTCA). For 6xHis-TEV-EC12-avitag, the same 5' oligo was used, and the 3' oligo was replaced with one containing both an AviTag[9] and stop codon (TGGTGCTCGAGTCATTCGTGCCATTCGATTTTCTGAGCCTCGAAGATGTCGTTCAGACC GCCACCAGGAGCGTTG). (green-*avitag*) The resulting PCR products were then re-inserted into pET-22b(+) using the same procedure as before.

Protein Expression and Purification

Protein Expression 6xHis-TEV-EC12 and Met-EC12-6xHis were expressed in *E. coli* BL21(DE3) cells as described previously,[8] with the substitution of 100 µg/mL carbenicillin replacing ampicillin. The 6xHis-TEV-EC12-avitag construct was co-transformed into BL21(DE3) cells with a plasmid containing biotin ligase (Gift from R. Fletterick). Protein was expressed under the same conditions as 6x-TEV-EC12 with the addition of 34 µg/mL chloramphenicol and 50 µM biotin (Thermo #29129) to LB culture medium. (BD #244610)

Lysis Cells were resuspended at 20 mL/L of culture in lysis buffer (50 mM Tris-HCl pH 8.0, 2.5 mM $MgCl_2$, 0.1% Triton X-100). After resuspension, protease inhibitor tablets (Roche #11836170001) were added according to manufacturer's instruction. After addition of 1U/mL of DNAse I (NEB #M0303) and 0.1 mg/mL lysozyme (Sigma #L6876), resuspended cells were stirred for 1 hour at room temperature. Cells were subsequently lysed using sonication, then

lysates were cleared by centrifugation at 21,000x*g* for 30 minutes at 4 °C. (*Note: In our hands, EC12 is prone to aggregation if NaCl is present in the lysis buffer prior to clearing of lysates.*)

Ni-NTA Purification After lysis, NaCl and imidazole were added to final concentrations of 300 mM and 10 mM, respectively, followed by 50% NiNTA agarose slurry (Qiagen #30230) at a volume of 5 mL/L of culture. Lysates were then nutated at 4 °C for 1-2 hours, and beads were pelleted by spinning at 3000x*g* for 10 minutes. Beads were resuspended in wash buffer (50 mM NaH_2PO_4-NaOH pH 8.0, 300 mM NaCl, 10 mM imidazole), separated *via* a gravity-flow chromatography column (BioRad #732-1010), then further washed *via* 2 column volumes (CV) of wash buffer. Protein was eluted from the beads using 1.25 CV elution buffer (50 mM NaH_2PO_4-NaOH pH 8.0, 300 mM NaCl, 250 mM imidazole). Eluted protein was injected into a HiLoad 16/600 Superdex 75 column (GE #28-9893-33) attached to an FPLC system and prequilibrated in TEV cut buffer (25 mM TrisHCl pH 8.5, 400 mM NaCl, 1 mM EDTA, 3 mM KCl, 3 mM TCEP). Monomer fractions were pooled. Typical yields of monomer ranged from 10-20 mg protein per liter of cell culture.

Protease Cleavage and Repurification

Cleavage Purified proteins were concentrated to a final concentration of 120 µM in TEV cut buffer using spin concentrators (Millipore #UFC901024) (*Note: In our hands, EC12 adsorbs to alternative protein concentrators*), after which 6xHis-TEV (Invitrogen #12575-015) was added in a mass ratio of 1:8 TEV:EC12. Protease reactions were left at 16 °C for 60 hours.

Repurification After cleavage, protease, tags, residual contaminating proteins, and residual uncleaved protein were removed using a second round of NiNTA purification. The protease reactions were desalted using a HiPrep 26/10 desalting column (GE #17-5087-01) into

repurification buffer (25 mM Tris-HCl pH 8.0, 400 mM NaCl, 30 mM imidazole, 500 µM TCEP). After desalting, Ni-NTA agarose pre-washed in repurification buffer was added at a ratio of 1mL of packed resin per 15 mg EC12 and mixtures were nutated for 1-2 hours at 4 °C. After nutating, supernatants containing cleaved protein were separated from beads using a gravity-flow chromatography column. Beads were washed using 1CV repurification buffer, and wash fractions were pooled with supernatants. Proteins were then desalted into TEV cut buffer for medium term storage. Protein purity was verified by SDS-PAGE and was estimated to be > 90% and cleavage completeness was verified *via* mass spectrometry. (**Figure S2A**) Biotinylation efficiency was also verified *via* mass spectrometry. (**Figure S2B**) Typical yields post-repurification were 66-75% compared to precleavage, with the major loss being incomplete cleavage. Biotinylation was typically >95% of overall protein.

Conjugation

3,3'-bis(sulfonato)-4,4'- bis(chloroacetamido)azobenzene (BSBCA) Handling Dry BSBCA (purchased from Linkera-osadovsk@chem.utoronto.ca) was kept at 4 °C. 16X stock solutions were created by dissolving dry BSBCA into water at a concentration of 8 mM. Stock solutions were kept at -20 °C in the dark.

Protein Conjugation Repurified and cleaved mutant protein was concentrated to a final concentration of 160 µM in TEV cut buffer, followed by addition of BSBCA from stock solutions to a final concentration of 500 µM. Conjugation reactions were kept in the dark at 25 °C for 72 hours. After conjugation, excess chromophore was removed *via* desalting using a HiPrep 26/10 desalting column (GE #17-5087-01) into TEV cut buffer. Conjugation efficiency was verified using mass spectrometry, looking for the characteristic increase of 453 Da[10], and efficiencies

were typically >99% for the mutant K129C/D138C. (**Figure S2C**) (*Note: Free chromophore has a tendency to adsorb to the resin, especially in the presence of salt. It may be removed by flushing the column with several volumes of pure water.*)

Final Purification & Storage

Prior to analysis, proteins were injected into a HiLoad 16/600 Superdex 75 column equilibrated in storage buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 500 µM TCEP) to remove any aggregates that may have developed during cleavage, repurification or conjugation. Typical monomer yields were 90% for unconjugated and 60% for conjugated protein. Proteins were stable in storage buffer for several weeks at 4 °C. For longer term storage, proteins were flash frozen in storage buffer containing 10% v/v glycerol using liquid nitrogen, then stored at -80 °C. (*Note: Conjugates had a high non-specific affinity for agarose and other common support matrices. Attempts to conjugate protein prior to cleavage and repurification led to protein aggregation. Attempts to separate TEV from conjugated EC12 using other column methods such as protease-affinity columns were similarly unsuccessful. Conjugated protein remains monomeric in the Superdex 75 and HiPrep Desalting columns used here.*)_

Protein Concentration Determination

Unconjugated protein concentrations were determined using measurements of absorbance at 280 nm using extinction coefficients predicted by the ExPASY online protein parameter tool[11], which were ε_{280}=21430 M^{-1} cm^{-1} for EC12, and ε_{280}=26430 M^{-1} cm^{-1} for EC12-avitag. For conjugated proteins, the extinction coefficient at 280 nm was computed as the sum of the predicted protein and measured free BSBCA (ε_{280}=10100 M^{-1} cm^{-1}) extinction coefficients. BSBCA's extinction coefficient was computed by measuring BSBCA's absorbance at 370 nm and

280 nm and using

$$\varepsilon_{280} = \frac{A_{280}}{A_{370}} * \varepsilon_{370} \quad (1)$$

where ε_{370}= 29000 M^{-1} cm^{-1} (see Zhang et al.[12]). BSBCA's molar extinction coefficients can vary when conjugated to protein. However, similar band intensities were observed when conjugated and unconjugated proteins of identical computed protein concentration were loaded into an SDS-PAGE gel and then stained with coomassie.

Conjugability Measurements

Post conjugation, conjugation reactions were diluted to a final protein concentration of 1 μM in pure water, and then analyzed *via* mass spectrometry using an LCT Premier (Waters). (**Figure S2C)** We observed the peak ratio between the unconjugated and conjugated (+453 Da) peaks, and, assuming equal ionizability for each species, made a qualitative determination of conjugability. (**Table S1**) Highly conjugatable mutants showed only trace remaining unconjugated proteins, whereas poorly conjugatable mutants showed as little as 10% estimated conjugation.

Switchability Measurements

For each cadherin double mutant, we determined whether the BSBCA conjugated to the protein was photoisomerizable by illuminating *trans*-relaxed X-EC12 with UV light. *Trans* BSBCA contains a characteristic absorbance peak near 370 nm, and, upon illumination at that wavelength, the peak amplitude decreases as the small molecule isomerizes into the *cis* state.[13]

Data provided to us by Dr. Andrew Woolley,[14] containing extinction coefficients for conjugated proteins separated by HPLC, containing isolated *trans* or *cis* isomers allowed us to estimate the fraction of EC12 that switched to *cis*. By assuming the observed 370 nm absorbance of any mixture of *cis* and *trans* can be described by the simple sum of the independent absorbances of the underlying *cis* and *trans* states, switching percentages can be calculated by

$$Frac = \frac{\varepsilon_{370nm,trans} - \varepsilon_{370nm,mix}}{\varepsilon_{370nm,trans} - R * \varepsilon_{370nm,trans}} \quad (2)$$

Here, R is the *cis*/*trans* extinction coefficient ratio computed from the provided data (.0541), $\varepsilon_{370nm,trans}$ is the measured extinction coefficient at 370 nm for the thermodynamically equilibrated, 100% *trans* state, and $\varepsilon_{370nm,mix}$ is the measured extinction coefficient at 370 nm for the photostable, UV illuminated state containing a mixture of *cis* and *trans*. Using this methodology, we computed a switchability for the free chromophore of 86% *cis*, close to published estimates of maximum switchability.[13]

Calcium Binding Measurements

Protein samples were prepared as previously described[15], with the modification that samples were diluted to 2 pmol/µL. Samples were injected at 5 µL/min into a QTRAP4000 instrument (Agilent). *Cis* X-EC12 was illuminated with a 1 W UV LED (Advancemart, emission maximum 365 nm) for 4 minutes immediately prior to injection. The quantity of bound calcium ions was obtained by determining Ca^{2+} binding occupancies, and assuming that calcium-free and calcium-bound molecules have the same ionizability.[16] By comparing peak areas, fractions of molecules binding 0 to 9 calcium ions were computed for each calcium concentration (**Figure S5**). For

determination of K_d, molecules binding more than three calcium ions were assumed to bind three ions specifically, and these fractions were added to the 3-ion fraction. The average number of calcium ions bound was computed, and the resulting numbers were fit to a Hill equation.[17] In order to subtract non-specific binding from *cis* X-EC12, the following equation was used for each calcium concentration, *i,* and occupancy number, *c*:

$$F_{c,i} = \hat{F}_{c,i} - \left(\sum_{a=1}^{c} F_{c,i} \times x_{a,i}\right) + \left(\sum_{a=1}^{\min(5-c,6)} \hat{F}_{c+a,i} * x_{a,i}\right) \quad (3)$$

where $F_{c,i}$ is the true fraction of the molecules binding *c* specific calciums at concentration *i*, $\hat{F}_{c,i}$ is the apparent fraction binding *c* calciums at concentration *i*, and $x_{a,i}$ is the fraction of the molecules that bind *a* non-specific calciums at concentration *i*. The first summation in the equation subtracts from the apparent fraction contributions due to non-specific binding, whereas the second term adds from the other calcium occupancy states their non-specific fractions that actually bind *c* calciums. The non-specific binding fractions, $x_{a,i}$, came from the *trans* X-EC12 calcium series using the assumption that all three calcium binding sites were occupied prior to non-specific sites, thus any fraction that appeared to bind four calciums actually bound three specific and one non-specific calcium, etc.

Isomerization Relaxation Measurements

Conjugated proteins in storage buffer at 12 µM were illuminated with the same 1 W UV LED for 4 minutes and the absorbance at 370 nm was monitored immediately after illumination and then every 20 minutes thereafter for a total of 180 minutes (**Figure S6**). The absorbances

were blank subtracted and then each absorbance was subtracted from 1.0 and the combined numbers were fit to a single exponential decay function of the form $y = A + e^{-b.x} + c$ using the curve fitting toolbox in MATLAB (The MathWorks) (**Figure S7**). R^2 values were typically >. 995. Half-lives were computed as $\lambda = {\ln(2)}/{b}$. For the calcium dependence of half-life, proteins were illuminated, and then $CaCl_2$ was added to final concentrations from 3.9 µM to 32 mM, using a 14-point 1:1 serial dilution series, plus a zero calcium point. Data points were collected five at a time on a Cary 50 Bio UV/Vis spectrometer (Varian), and the entire series was run in triplicate. The mean half-lives for each calcium concentration were plotted and fit to a Hill curve of the form

$$y = \left({bx^{N_h}}/{c^{N_h} + x^{N_h}}\right) + d \tag{4}$$

using the curve fitting toolbox in MATLAB, where N_h is the hill coefficient and *c* is the EC_{50} for calcium-dependent half-life reduction. Absorbance curves were also fit to double exponentials, without a significant increase in curve quality (Data not shown).

Isomerization Reversibility

Conjugated protein at 12 µM in storage buffer was illuminated using the same UV LED for 2 minutes and 370 nm absorbance was measured. After measurement, protein was illuminated with a 1 W LED (Sparkfun, emission maximum 455 nm, with residual intensity in the 500-550nm range) for 2 minutes, 370 nm absorbance was measured, and the process was repeated for 10 cycles.

Surface Plasmon Resonance (SPR)

Data Acquisition Matrix-free, flat, carboxymethylated gold surfaces (GE Healthcare Sensor Chip C1) were used in all SPR experiments. Individual flow cells were prepared with the following protocol: (1) 50 µl injection of 1-Ethyl-3- (3-dimethylaminopropyl)carbodiimide / N-hydroxysuccinimide (0.5M:0.2M); (2) 30 µL injection of 0.25 mg ml-1 ImmunoPure Streptavidin (Pierce #21122) in sodium acetate buffer (pH 5.0) to a total amount of ~500 response units (RU) for all four flow cells; (3) 60 µl injection of 1 M ethanolamine. Cadherin was captured on the active flow cells by manual injection of 25 nM protein to immobilization levels between 250-450 RU. No blocking procedures were performed on the reference flow cell. Prior to each SPR experiment, protein samples were dialyzed against 25 mM Tris, 150 mM NaCl, and 500 µM TCEP. Following dialysis, Tween 20 detergent (Sigma-Aldrich #P9416) was added to the protein solution and the dialysis buffer to achieve a final concentration of 0.05% (v/v). The dialysis buffer was then used as the assay buffer for the SPR measurements. Dose response titrations were prepared by serial dilutions of the highest concentration into assay buffer. While measuring *cis* X-EC12, a maximum of three data points per illumination cycle were used, to maximize *cis* fraction and minimize thermal relaxation. To minimize systematic error due to some fraction of the protein reverting to *trans* during the experiment, concentrations were injected in a random order. For all samples, injection was followed by a control injection of 20 µM WT C9A cadherin containing 1mM $CaCl_2$ to monitor degradation of chip response over time. Each sample response was subtracted by a reference response (containing no calcium), and then corrected in magnitude by the magnitude of the control injection. Raw SPR traces are shown in **Figure S9**. The responses of flow cells 2 and 3 were scaled to match flow cell 4 *via* a least squares minimization resulting in a single scalar multiplier for each flow cell. Each titration

series was then fit to an equation of the same form as eq 4, where here *c* is the EC_{50} for calcium-dependent protein binding using the curve fitting toolbox in MATLAB. Due to chip-to-chip variation, calcium binding fits were normalized such that the maximum responses at infinite concentration predicted by the fit lines were equal to one. During calcium titrations, we observed an EC_{50} -dependent, non-specific interaction at higher calcium concentrations, with a stronger (lower) EC_{50} leading to a larger magnitude effect (i.e. WT C9A had the strongest effect, *cis* X-EC12 had no apparent effect at concentrations tested). Points dominated by this effect were not used in the fits, but are shown as faded markers in **Figures 3** & **S8**. For reversibility analysis, 40 μM X-EC12 was alternatingly illuminated with UV (emission maximum 365 nm) and blue LEDs (emission maximum 455 nm, with residual intensity at 500-550 nm), 1mM Ca^{2+} was added, and responses measured. The resulting responses were subject to the same reference and control subtraction, as well as flow cell scaling used in the calcium titrations.

Bootstrap Analysis We used a bootstrapping technique in order to verify fit values were robust. Data points for each flow cell and concentration were grouped, and then data points were drawn at random, with replacement, in a number equal to the number of actual data points. These randomly drawn data points were then fit to an equation with the same form as eq 4, where here *c* is the EC_{50} for calcium-dependent protein binding and EC_{50} and N_h values were stored. This process was repeated 500 times for each protein state. The fits were then sorted by minimum sum of squared error (SSE), and the average and ±2 SD values were computed for the top 100 fits for each protein. Computed values were all near the values reported for the single best fit using the gathered data, which indicates the data values describe the system well and the fit values are robust. Clusters of all computed fit values are shown in **Figure S10** and clusters of the top 100 fits by SSE are shown in **Figure S11**.

Dissociation analysis We fit each protein variant's SPR dissociation traces to a single exponential of the form:

$$response(t) = a_x * e^{-k \cdot t} + b_x \quad (5)$$

where *response(t)* is the response as a function of time, *t*, and *k* is the off rate. For each protein, *a* and *b* were allowed to vary per response, while all responses were simultaneously fit to a single shared off rate, *k*, that minimized the sum of squared errors (SSE) of the responses. Calcium concentrations less than 10 μM were removed from the fits, as their responses were not significantly above the noise threshold of the instrument. For WT, the fit-determined off rate was k_{wt} = 0.091 sec^{-1}, for *trans* X-EC12, k_{trans} = 0.075 sec^{-1}, and for *cis* X-EC12, k_{cis} = 0.072 sec^{-1}. Because all traces for each protein could be fit to a single off rate using a single exponential fit, this indicates that we observed only one interacting species dissociating during the experiment. Dissociation traces, fits and example residuals are shown in **Figure S13.**

Size Exclusion Chromatography

We used size exclusion chromatography to show a decrease in X-EC12 homodimerization after illumination. Protein was diluted to a final concentration of 250 μM in TBS (25mM Tris pH 7.5, 150mM NaCl) at room temperature. For *trans* X-EC12, calcium was added to a final concentration of 2.5 mM and left to equilibrate for 5 minutes at room temperature, then 25 μL of protein was injected into a Superdex 75 PC 3.3/30 column (GE # 17-0771-01) equilibrated in TBS and attached to a 1200 Series HPLC (Agilent) at a flow rate of 100 μL/min. *Cis* X-EC12 was illuminated with a 1W UV LED (emission maximum 365 nm) for 6 minutes prior to the addition of calcium, after which it was treated the same as *trans*. Protein was detected by monitoring

absorbance at 280 nm. We observed the expected decrease in the fraction of protein forming dimers after illumination, indicating that the *cis* X-EC12 homodimerization is weaker than that of *trans* X-EC12. Representative traces are shown in **Figure S12**.

References for Supplementary Materials

(1) Rosetta- The premier software suite for macrommolecular modeling. http://www.rosettacommons.org (accessed May 16th 2013)

(2) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14116.

(3) Kortemme, T.; Kim, D.; Baker, D. *Sci. Signaling* **2004**, *2004*, pl2.

(4) Schrodinger, LLC 2013.

(5) Leckband, D.; Prakasam, A. *Annu Rev Biomed Eng* **2006**, *8*, 259.
(6) Harrison, O. J.; Bahna, F.; Katsamba, P. S.; Jin, X.; Brasch, J.; Vendome, J.; Ahlsen, G.; Carroll, K. J.; Price, S. R.; Honig, B.; Shapiro, L. *Nat. Struct. Mol. Biol.* **2010**.

(7) Harrison, O. J.; Corps, E. M.; Kilshaw, P. J. *J. Cell Sci.* **2005**, *118*, 4123.

(8) Koch, A.; Pokutta, S.; Lustig, A.; Engel, J. *Biochemistry* **1997**, *36*, 7697.

(9) Beckett, D.; Kovaleva, E.; Schatz, P. J. *Protein Sci.* **1999**, *8*, 921.

(10) Burns, D. C.; Zhang, F.; Woolley, G. A. *Nat. Protoc.* **2007**, *2*, 251.

(11) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. *Nucleic Acids Res* **2003**, *31*, 3784.

(12) Zhang, Z.; Burns, D.; Kumita, J.; Smart, O.; Woolley, G. *Bioconj. Chem.* **2003**.

(13) Beharry, A. A.; Woolley, G. A. *Chem Soc Rev* **2011**, *40*, 4422.

(14) Woolley, G., University of Toronto, Toronto, ON Personal Communication, 2012.

(15) Courjean, O.; Chevreux, G.; Perret, E.; Morel, A.; Sanglier, S.; Potier, N.; Engel, J.; van Dorsselaer, A.; Feracci, H. *Biochemistry* **2008**, *47*, 2339.

(16) Peschke, M.; Verkerk, U. H.; Kebarle, P. *J Am Soc Mass Spectrom* **2004**, *15*, 1424.

(17) Weiss, J. N. *FASEB J.* **1997**, *11*, 835.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature Date