Title
Ancient transcriptional regulators can easily evolve new pair-wise cooperativity

Author
Fowler, Kyle Rhey

Publication Date
2022

Peer reviewed|Thesis/dissertation

Ancient transcriptional regulators can easily evolve new pair-wise cooperativity

by
Kyle Fowler

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
James Fraser
EAF57DFF2EA94C9...

James Fraser
Chair

DocuSigned by:
Alexander Johnson
[illegible]45B...

Alexander Johnson

DocuSigned by:
Hiten Madhani
242AFD8ACF95471...

Hiten Madhani

Committee Members

Acknowledgments

Many individuals have contributed to this dissertation and influenced my scientific trajectory, all of whom deserve recognition. This body of work was made possible by numerous mentors and colleagues, friends and family. Foremost among them is my advisor, Sandy Johnson. From my first day in his lab, Sandy allowed me incredible intellectual freedom, while still providing judicious guidance. I was able to explore questions and techniques altogether new to me and the lab. This freedom helped me develop as an experimentalist and professional scientist: I learned not only how to devise and perform rigorous experiments, but also how to manage projects and mentor other researchers. Effectively, learning how to conduct a lab. I am very thankful to have had such rounded training. In addition to Sandy, the other members of my thesis committee – Jaime Fraser and Hiten Madhani – have provided very thoughtful guidance and coaching throughout my time at UCSF. I am grateful for their direction and support.

One of the most important lessons I learned at UCSF was the importance of surrounding oneself with the right people. All the members of the Johnson lab have been such a tremendous support; I admire each of them. When I joined the lab, I was incredibly fortunate to have been mentored by Candace Britton, a consummate scientist who took mentorship seriously. I did my best to emulate Candace when I then mentored another new graduate student, Francesca del Frate. Francesca is such a creative and artistic force and I will miss her perspectives and questions. It has likewise been a pleasure watching Haley Gause and Jenny Zhang, two other new graduate students, as they fearlessly brought exciting new science to the lab. The post-doctoral fellows in the lab have been equally influential. Sheena Singh-Babak was an impeccable bench-mate – perfectly balancing work and fun - and taught me a great deal about microbial genetics (and Canadian history). But I am especially thankful to have worked with and befriended Naomi Ziv. My time and training in the Johnson lab was made more complete thanks to Naomi. I am in her debt. Finally, I would be remiss if I didn't thank Matt Lohse, who transcends title and is the bedrock of the Johnson lab. UCSF is also filled with many other brilliant people, including my classmates in the Tetrad Graduate Program. This whole endeavor wouldn't have been the same without them, and I'm excited for their futures. Rachel

Greenstein in particular has been an incredible classmate, friend, and confidant. We started as roommates when I first moved to San Francisco, but grew to become more. I am so glad that we could share the graduate school journey together.

The pursuit of my PhD wouldn't have even been possible if not for my loving friends and family, who have supported me and been my greatest champions. Chief among them is my mother, Janell Fowler. She taught me to be curious and ask any question; to value the attainment of knowledge for its own sake. More importantly, she taught me to do these things out of love for others and for the world; to help leave the world better off. This has become the basis of my life and driven my scientific career. These values are evident in my siblings, Paul and Megan Fowler, who have supported me in ways they probably don't even realize. They have inspired me and lifted me up since we were children, using the struggles we each have overcome in our lives to empower one another. Finally, I am thankful to my husband, Selim Boudoukha, who in a single breath can dispel my worries, bring a smile to my face, and make me roll my eyes. I couldn't imagine this journey without you.

This body of work, and all my prior research, has been dedicated to understanding how life operates at the most fundamental of levels: the level of DNA. It has been a privilege to spend these years investigating fundamental questions. My scientific journey began well before I started graduate school and I've had many academic stewards along the way. Beginning under the tutelage of Marion Brodhagen at Western Washington University, followed by Gerry Smith at the Fred Hutchinson Cancer Research Center, and finally Sandy at UCSF. I am grateful to all these professors and universities, especially the University of California, for providing me this path forward. The next steps in my career will undoubtedly embody the drive, curiosity, and courage displayed by all the above individuals and institutions.

Finally, I dedicate this thesis to my late grandmothers, Jeane Gurwell & Alma Fowler.

The work described in Chapter 2 arose from brainstorming discussions with Candace Britton and Liron Noiman, who both helped conceive of the experimental approach. Frederick Leon, a rotation student in the TETRAD graduate program, helped assemble and test the initial mutant library and is a coauthor on the manuscript detailing this work that is being prepared for submission to a peer reviewed journal. I designed and performed all other experiments, and wrote the manuscript in collaboration with Alexander Johnson.

Abstract

Ancient transcriptional regulators can easily evolve new pair-wise cooperativity

Kyle R. Fowler

The immense diversity of life is astounding. Yet this variety is in stark contrast with life's genetic similarity. Many of the same genes and even gene sequences are shared by much of the life on Earth. Thus, perhaps what best defines a species is not its gene complement, *per se*, but how those genes are used and regulated. Much of gene regulation depends on the binding of sequence specific transcription factors to cis-regulatory sites in other gene promoters. Changes in these protein factors, or the DNA sites they bind, can alter gene expression and give rise to new traits and forms of life. Likewise, these transcription factors themselves may change, gaining (or losing) functionalities that alter the organism. For example, in some budding yeast species two ancient and unrelated transcription factors – Matα2 and MCM1 – acquired the ability to bind one another and co-regulate new genes. This facilitated the emergence of a new mode of combinatorial gene control in these species. In this work, I show that this new form of regulation likely arose easily due to the promiscuous nature of the Matα2-MCM1 interaction. This promiscuity allows for many alternative ways for Matα2 to bind MCM1 and involves widespread intramolecular epistasis within Matα2. The gain of new functional protein-protein interactions between transcriptional regulators, as explored here, is believed to be a general way that life diversifies and likely has broad implications. Understanding how and why new biological traits emerge is a fundamental question in biology.

Table of Contents

List of Figures

Chapter 2

List of Tables

Chapter 2

Chapter One

Introduction

Any genetic change has the potential to alter an organism's phenotype. Genetic changes can arise in multiple ways: *de novo* mutations, for example, generate novel DNA sequences, while recombination shuffles together novel configurations of existing genetic diversity. Geneticists have long studied the physical manifestation and consequence of mutation and recombination (Creighton and McClintock, 1931; Harris et al., 1994; Muller, 1964). Even before the discovery that DNA was the primary molecule of heritability, the study of genetics has been intimately coupled with the study of heritable phenotypes (Johannsen, 1911). Some genetic variation results in a predictable change to the organism, which can be either beneficial, detrimental, or neutral with regard to an organism's ability to reproduce and propagate that genetic change. Most genetic phenotypes, however, exhibit complicated and unpredictable patterns of inheritance that depend on other genetic factors (Bateson and Mendel, 1911; Elena and Lenski, 1997; Fisher, 1919; Phillips, 2008; Tong et al., 2004). This phenomenon of genetic interactions – termed epistasis – means that the effect of a mutation depends on the genetic background (Bateson and Mendel, 1911; Fisher, 1919). Because of this context-dependence, many mutations don't have any immediate effect on the phenotype of the population, though they will still contribute to genetic diversity. The significance of these unseen mutations – and of epistasis – wouldn't become apparent for many more years.

Genetically diverse model organisms, such as those among the Saccharomycotina fungi, have proven instrumental in the study of epistasis. This fungal lineage spans an incredible evolutionary distance – approximately equal to the diverge of humans from sea squirts – and therefore provides diverse genetic contexts in which to study mutations (Shen et al., 2018). Many epistatic interactions are even harbored amongst the standing genetic diversity found within individual species, such as among strains of the budding yeast *Saccharomyces cerevisiae*, which have dispersed around the world and are found in diverse ecological niches (Brem et al., 2005; Fournier et al., 2019; Fournier and Schacherer, 2017; Peter et al., 2018). Mutations that affect many diverse cellular processes, such as DNA mismatch repair and sugar metabolism, have been found to have strain-specific effects and depend on other genetic factors dispersed throughout the genome (Heck et al., 2006; Sato et al., 2016). Mapping these modifier sites is difficult but

often localize to other genes in the same or related pathways (Costanzo et al., 2016; Segrè et al., 2005). Indeed, many instances of intermolecular epistasis involve epistatic interactions across a molecular interface, such as between a transcription factor protein and its cis-regulatory sequence, or between a kinase and its target (Podgornaia and Laub, 2015; Starr et al., 2017). This work highlights the ubiquity of epistasis between diverse molecules and among complex genetic systems.

Intramolecular epistasis conversely involves genetic interactions between different parts of the same molecule, such as among amino acid residues in a protein. Just as epistasis emerges from complex molecular pathways, so too can it be found among the network of amino acid interactions within a protein (Miton et al., 2021; Starr and Thornton, 2016; Storz, 2018). Because protein evolution (or engineering) involves the alteration of these amino acid networks, epistasis can influence which outcomes are permissible. Given that the gene (and its encoded protein) are considered the fundamental unit of heredity, intramolecular epistasis likely has profound evolutionary implications. In this thesis, I investigate the significant influence of epistasis on the evolution of a new trait, namely combinatorial gene regulation.

References

Bateson, W., Mendel, G., 1911. Mendel's Principles of Heredity. Nature 86, 407–407. https://doi.org/10.1038/086407a0

Brem, R.B., Storey, J.D., Whittle, J., Kruglyak, L., 2005. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 436, 701–703. https://doi.org/10.1038/nature03865

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., Pelechano, V., Styles, E.B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J.S., Srikumar, T., Bahr, S., Chen, Y., Deshpande, R., Kurat, C.F., Li, S.C., Li, Z., Usaj, M.M., Okada, H., Pascoe, N., San Luis, B.-J., Sharifpoor, S., Shuteriqi, E., Simpkins, S.W., Snider, J., Suresh, H.G., Tan, Y., Zhu, H., Malod-Dognin, N., Janjic, V., Przulj, N., Troyanskaya, O.G., Stagljar, I., Xia, T., Ohya, Y., Gingras, A.-C., Raught, B., Boutros, M., Steinmetz, L.M., Moore, C.L., Rosebrock, A.P., Caudy, A.A., Myers, C.L., Andrews, B., Boone, C., 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science 353, aaf1420–aaf1420. https://doi.org/10.1126/science.aaf1420

Creighton, H.B., McClintock, B., 1931. A Correlation of Cytological and Genetical Crossing-Over in Zea Mays. Proc. Natl. Acad. Sci. U.S.A. 17, 492–497. https://doi.org/10.1073/pnas.17.8.492

Elena, S.F., Lenski, R.E., 1997. Test of synergistic interactions among deleterious mutations in bacteria. Nature 390, 395–398. https://doi.org/10.1038/37108

Fisher, R.A., 1919. The Correlation between Relatives on the Supposition of Mendelian Inheritance. Trans. R. Soc. Edinb. 52, 399–433. https://doi.org/10.1017/S0080456800012163

Fournier, T., Abou Saada, O., Hou, J., Peter, J., Caudal, E., Schacherer, J., 2019. Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. eLife 8, e49258. https://doi.org/10.7554/eLife.49258

Fournier, T., Schacherer, J., 2017. Genetic backgrounds and hidden trait complexity in natural populations. Current Opinion in Genetics & Development, Evolutionary genetics 47, 48–53. https://doi.org/10.1016/j.gde.2017.08.009

Harris, R.S., Longerich, S., Rosenberg, S.M., 1994. Recombination in Adaptive Mutation. Science 264, 258–260. https://doi.org/10.1126/science.8146657

Heck, J.A., Argueso, J.L., Gemici, Z., Reeves, R.G., Bernard, A., Aquadro, C.F., Alani, E., 2006. Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. Proc. Natl. Acad. Sci. U.S.A. 103, 3256–3261. https://doi.org/10.1073/pnas.0510998103

Johannsen, W., 1911. The Genotype Conception of Heredity. The American Naturalist 45, 129–159. https://doi.org/10.1086/279202

Miton, C.M., Buda, K., Tokuriki, N., 2021. Epistasis and intramolecular networks in protein evolution. Current Opinion in Structural Biology 69, 160–168. https://doi.org/10.1016/j.sbi.2021.04.007

Muller, H.J., 1964. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1, 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P.,

Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Phillips, P.C., 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9, 855–867. https://doi.org/10.1038/nrg2452

Podgornaia, A.I., Laub, M.T., 2015. Pervasive degeneracy and epistasis in a protein-protein interface. Science 347, 673–677. https://doi.org/10.1126/science.1257360

Sato, T.K., Tremaine, M., Parreiras, L.S., Hebert, A.S., Myers, K.S., Higbee, A.J., Sardi, M., McIlwain, S.J., Ong, I.M., Breuer, R.J., Avanasi Narasimhan, R., McGee, M.A., Dickinson, Q., La Reau, A., Xie, D., Tian, M., Reed, J.L., Zhang, Y., Coon, J.J., Hittinger, C.T., Gasch, A.P., Landick, R., 2016. Directed Evolution Reveals Unexpected Epistatic Interactions That Alter Metabolic Regulation and Enable Anaerobic Xylose Use by Saccharomyces cerevisiae. PLoS Genet 12, e1006372. https://doi.org/10.1371/journal.pgen.1006372

Segrè, D., DeLuna, A., Church, G.M., Kishony, R., 2005. Modular epistasis in yeast metabolism. Nat Genet 37, 77–83. https://doi.org/10.1038/ng1489

Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., Boudouris, J.T., Schneider, R.M., Langdon, Q.K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., Rosa, C.A., DeVirgilio, J., Hulfachor, A.B., Groenewald, M., Kurtzman, C.P., Hittinger, C.T., Rokas, A., 2018. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell 175, 1533-1545.e20. https://doi.org/10.1016/j.cell.2018.10.023

Starr, T.N., Picton, L.K., Thornton, J.W., 2017. Alternative evolutionary histories in the sequence space of an ancient protein. Nature 549, 409–413. https://doi.org/10.1038/nature23902

Starr, T.N., Thornton, J.W., 2016. Epistasis in protein evolution. Protein Science 25, 1204–1218. https://doi.org/10.1002/pro.2897

Storz, J.F., 2018. Compensatory mutations and epistasis for protein function. Curr. Opin. Struct. Biol. 50, 18–25. https://doi.org/10.1016/j.sbi.2017.10.009

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D.S., Haynes, J., Humphries, C., He, G., Hussein, S., Ke, L., Krogan, N., Li, Z., Levinson, J.N., Lu, H., Ménard, P., Munyana, C., Parsons, A.B., Ryan, O., Tonikian, R., Roberts, T., Sdicu, A.-M., Shapiro, J., Sheikh, B., Suter, B., Wong, S.L., Zhang, L.V., Zhu, H., Burd, C.G., Munro, S., Sander, C., Rine, J., Greenblatt, J., Peter, M., Bretscher, A., Bell, G., Roth, F.P., Brown, G.W., Andrews, B., Bussey, H., Boone, C., 2004. Global Mapping of the Yeast Genetic Interaction Network. Science 303, 808–813. https://doi.org/10.1126/science.1091317

Chapter Two

Ancient transcriptional regulators can easily evolve new pair-wise cooperativity

Abstract

Cells regulate gene expression by the specific binding of one or more transcription regulators to cis-regulatory sequences. Pairwise cooperativity between regulators – whereby they physically interact and bind DNA in a cooperative manner – permits complex modes of gene regulation. The combinatorial interactions of transcriptional regulators represent a major source of phenotypic novelty, facilitating the integration of new target genes or new inputs into gene networks, and resulting in increased network complexity. How functional cooperative interactions arise between regulators is poorly understood, despite abundant examples in nature of pairwise interactions. Here we explore a protein-protein interaction between two ancient transcriptional regulators in ascomycete yeasts that were recently (at least 200 million years ago) gained. This newly-formed interaction was due to the acquisition of a short amino acid sequence in the homeodomain protein Matα2 which interacts with the MADS-box regulator MCM1 and causes the two proteins to bind DNA cooperatively. By combining deep mutational scanning with a functional selection for cooperativity, we tested millions of possible alternative evolutionary solutions in the sequence space of this interaction interface. Our selection and analysis of functional alternatives reveals biochemical principals guiding and constraining the evolution of cooperativity. The newly-evolved solutions are highly degenerate, with diverse amino acid chemistries permitted at all positions but with widespread epistasis limiting functional solutions, yet ~45% of the random sequences function as well or better than the naturally evolved sequence. Permissive contexts often involve large aromatic residues, emphasizing biochemical properties that facilitate cooperativity without being strictly necessary. This work highlights how natural solutions often represent a single evolutionary path in a vast landscape of alternatives. We propose that the capacity of ancient transcriptional regulators to easily gain new combinatorial interactions contributes to the rapid evolution and diversification of transcriptional networks.

Introduction

Combinatorial control of gene expression allows for complex transcriptional networks to evolve (Johnson, 2017; Sorrells and Johnson, 2015; Wang et al., 2005). While some transcriptional regulators work on their own, most act in combination with additional regulators. Gene networks – sets of functionally related genes controlled by shared cis-regulatory sequences – are typically bound and regulated by multiple regulators simultaneously (Lee et al., 2002; Nobile et al., 2012; Sorrells and Johnson, 2015). Co-regulation is often facilitated by specific cooperative interactions between pairs of different regulators that are essential for the binding of the regulators on cis-regulatory sequences (Jolma et al., 2015; Mead et al., 1996; Monahan et al., 2017; Pilpel et al., 2001; Ravasi et al., 2010).

The rewiring of gene networks over evolutionary timescales is an important source of phenotypic novelty (Li and Johnson, 2010; Wray, 2003). Over evolutionary time, both cis-regulatory sequences and the regulatory proteins that bind them change, either neutrally or adaptively. Although transcription regulators often preserve their DNA-binding specificity over long evolutionary times, changes within regulators can give rise to new combinatorial interactions between established regulators and allow for large-scale rewirings (Sorrells et al., 2018; Tuch et al., 2008). Although extensive pleiotropy is thought to significantly constrain changes in regulatory proteins, there are well-documented instances of changes in the transcriptional regulators themselves, including the gain or loss of protein-protein interactions (Britton et al., 2020; Sorrells et al., 2018).

One well-studied gain of a protein-protein interaction is found in a particular clade of fungi where the ancient homeodomain protein Matα2 binds DNA cooperatively with the MADS box protein MCM1 (Figure 2.1A)(Vershon and Johnson, 1993). This cooperativity arose when Matα2 gained the ability to physical interact and cooperatively bind DNA with MCM1 within the clade; the interaction is not found in species outside the clade (Baker et al., 2012; Britton et al., 2020). The emergence of the interaction is due to changes only in Matα2; the MCM1 surface involved in the interaction is deeply conserved and did not change when the interaction evolved (Mead et al., 2002, 1996; Tsong et al., 2006). The novel Matα2-MCM1

interaction facilitated the formation of a new transcriptional network to repress the a-specific genes, which are involved in cell type specification. Prior to this new form of transcriptional control, the a-specific genes were instead positively regulated by the single HMG-domain protein Mata2, which was subsequently lost in many species within the clade (Butler et al., 2004; Tsong et al., 2006). Thus, the emergence of this protein interaction marked a fundamental shift in the mode of gene regulation – from positive to negative – and in the transcription regulators responsible for it.

The cooperative interaction between Matα2 and MCM1 is due to a short (approximately 7-11 amino acid) region of Matα2 (Mead et al., 1996; Tsong et al., 2006). A crystal structure of the Matα2-MCM1 complex bound to DNA shows that this region forms a short beta-strand when bound to MCM1 (Tan and Richmond, 1998). Numerous contacts between the two proteins occur over a relatively small (~20 Å) interface, including an apparent cation-pi interaction with a phenylalanine in Matα2. Prior genetic work (alanine-scanning) has shown that seven contiguous residues in this region of Matα2 (including the above phenylalanine) are essential for efficient a-specific gene repression (Mead et al., 1996).

In this work, we investigate the emergence and evolvability of combinatorial gene regulation using the Matα2-MCM1 complex as a model. We began with the pairwise interaction missing and selected from a highly diverse pool of Matα2 variants those that could efficiently repress transcription in combination with MCM1. We uncovered many distinct alternatives to the "naturally evolved" sequence, suggesting that functional interactions can arise with relative ease. Indeed, nearly half of the derivatives we assayed worked as well or better than the natural sequence. From these variants we discern rules and constraints governing the emergence of cooperativity between transcriptional regulators. This work provides a mechanistic basis for long-standing observations of transcription network plasticity and highlights the importance of epistasis in the evolution of new protein-protein interactions.

Results

Selection for cooperative gene repression from a library of Matα2 mutants

In *S. cerevisiae*, expression of the gene CAN1 in the presence of canavanine in the medium arrests cell growth (Gocke and Manney, 1979). We replaced the endogenous CAN1 promoter with a synthetic constitutive construct containing a Matα2-MCM1 cis-regulatory sequence (Figure 2.1B). When bound by Matα2-MCM1, which requires pairwise-cooperativity between the proteins, this sequence element brings about strong transcriptional repression of CAN1 and allows growth in canavanine. *S. cerevisiae* a-cells (which lack Matα2) transformed with this selection system were susceptible to canavanine even at low concentrations (MIC ~1 μg/mL) and failed to produce colonies on solid media containing canavanine. Conversely, α-cells (which produce Matα2) transformed with this system were able to grow in the presence of high concentrations of canavanine (MIC > 100 μg/mL) and produced large uniform colonies on solid media containing canavanine.

We generated two different Matα2 mutant libraries based on the *S. cerevisiae* protein. Both libraries consist of Matα2 from *S. cerevisiae* with its endogenous promoter cloned into a low-copy plasmid and mutations introduced at key positions that mediate its interaction with MCM1. The first library introduced individual amino acid changes at eleven consecutive residues (G113-M123) known to span the sequence of Matα2 that interacts with MCM1. Variants in this library possess a single amino acid change relative to the wild-type protein. Seven of these eleven positions (114-120) represent the essential "core", which interacts directly with MCM1 as observed in a crystal structure of the Matα2-MCM1 complex bound to DNA (Tan and Richmond, 1998). A wild-type Matα2 construct (Sc-α2) was also included as a control.

In the second library, we randomized the seven core residues using an NNK oligonucleotide and obtained ~1.2 million constructs with unique amino acid combinations. Variants in this library differ from the wild-type protein, on average, at all seven amino acids. With this library we sought to estimate the ubiquity of functional proteins and thereby reveal the ease with which this function could evolve. The total number of possible amino acid combinations in this library is immense (207 or ~1.3 billion possibilities)

and technically infeasible to sample completely. However, we reasoned that even a sparse sampling of this sequence space could reveal important functional trends.

To screen for functional variants of Matα2 that could interact with MCM1, we transformed the Matα2 mutant library into *S. cerevisiae* a-cells where the only expressed Matα2 protein is from the plasmid library. As a control, we spiked in cells carrying the wild-type Sc-α2 plasmid at a concentration of 0.1%. Cells with the Matα2 constructs were then grown for 24 hours in media either lacking canavanine (representing the pool of unselected variants) or at 250 μg/mL to enrich for functional variants. Following growth, the Matα2 plasmid was purified from the final populations and sequenced deeply. The frequency of any given construct among the reads from each pool correlates with its abundance in that population of cells. We can then calculate a fold-change (FC) for each construct after selection relative to either the starting population or the population grown without canavanine selection, allowing us to control for growth effects due to the vector or protein expression. For example a FC of < 1 means that the sequence was selected against in canavanine. This fold-change allows us to estimate relative fitness for each sequence examined.

In both libraries, Matα2 mutants with premature stop codons diminished in frequency in the population and therefore exhibited low FCs. Conversely, the frequency of the wild-type Sc-α2 gene – as well as those bearing synonymous mutations – remained stable after selection (Figure 2.1C & 2.1D). The fully functional Sc-α2 had a FC around 1.0 and was surpassed by numerous other constructs. This indicates that our selection regime successfully enriched for functional Matα2-MCM1 interactions.

The evolved S. cerevisiae interaction is robust to many single amino acid changes

Using the first library, we assessed the consequences of single amino acid mutations on Matα2-MCM1 function. Many amino acid substitutions were well tolerated and remained at high frequency after selection (Figure 2.1D). For example, the extant glycine at the first position (G113) could be substituted with any other amino acid without a significant fitness effect (Figure 2.1E). Previous work similarly found

G113A to be permissive for Matα2-MCM1 function, while the next seven "core" positions (L114-T120) were much more sensitive to alanine substitution. This pattern was also apparent in our screen: mutation of a core residue was more likely to be detrimental than mutation of a neighboring residue. This was especially apparent for certain amino acid replacements. For example, the core was especially sensitive to amino acids with negatively charged (D/E) or small (G/A) side chains.

Notably, not all positions were equally sensitive to mutation. The phenylalanine at position 116 is especially critical, as many substitution mutations were highly detrimental (Figure 2.1E). However, replacement with another large aromatic (e.g. F116W) or aliphatic residue (F116I) did not diminish its function, suggesting that a bulky or hydrophobic side chain at this position is critical for the naturally occurring Matα2-MCM1 interaction. Aromatic amino acids were also tolerated at other positions throughout the region: mutation to phenylalanine, tryptophan, or tyrosine at most positions had little or no effect. Together these results suggest the Matα2-MCM1 interaction depends broadly on hydropathy, as well as specific side chain chemistries at critical positions. These apparent requirements may have constrained the naturally occurring Matα2-MCM1 interaction once it had formed.

Abundant functional alternative interaction domains

Single amino acid mutations tell us which individual residues contribute to function for a given configuration (in this case, Matα2 of *S. cerevisiae*), but this approach does not reveal the broad spectrum of possible solutions. To address this question we turned to our second library where the interfacial positions of Matα2 were randomized generating a complex library of amino acid sequences. This library of random interfaces (which is not based on the naturally occurring interface) reflects the possibilities of de novo evolution of a cooperative interaction.

We used the selection scheme described above to enrich for functional Matα2 variants that had the ability to cooperate with MCM1. Following selection, these Matα2 variants exhibited a broad, largely symmetric fitness distribution with a median FC around one (Figure 2.1C). The Sc-α2 construct also

exhibited a FC close to one (0.93) and was surpassed by many variants, indicating a large dynamic range of "successful" variants, many of which perform better in the selection scheme than the naturally occurring Sc-α2.

Using the naturally occurring Sc-α2 as a basis for comparison, we estimate that ~35% of variants in our library (> 100,000 unique proteins) are at least as functional as the extant protein in our transcriptional repression assay (Figure 2.1C). We refer to these as "fit" Matα2 variants. This high frequency was reproducible between replicates and robust to read depth: when variants with a low starting frequency were excluded, the percentage of fit sequences was even higher approaching ~45%. Thus, many combinations of random amino acids result in a functional Matα2-MCM1 complex. This suggests that a sizable fraction of this sequence space is functional.

We note that it is difficult to capture and quantify low abundance variants due to stochastic losses/gains, especially if variants are unfit and decrease in frequency. Indeed, many variants exhibiting low fitness (low FC) were poorly recovered in replicate experiments and exhibited relatively noisy FC values (Supplemental figure 2.1A). Among the variants with a high starting frequency, however, we find the FC values to be highly reproducible. This further confirms that 35-45% of the variants examined here are at least as functional as Sc-α2. This result suggests that the evolution of cooperativity between these two ancient transcription factors can occur with relative ease.

Functional Matα2 variants are diverse and degenerate

What is the molecular basis for this large number of functional alternatives? Do their solutions exhibit any patterns? When we align the amino acid sequences of fit variants (those at least as functional as Sc-α2), we find that no position is strongly constrained and, instead, many amino acids are tolerated at each position (Figure 2.2A). For example, among the 100 most fit sequences, all amino acids except proline can be found at each position (Supplemental figure 2.1B), consistent with our observation that many different sequences can functionally interact with MCM1. Tabulating the amino acids from variants at least

as fit as Sc-α2 did reveal a bias for certain residues, such as for phenylalanine and leucine, with proline and charged residues being disfavored (Supplemental figure 2.1C). Because these preferences are not strongly position specific, they are not readily apparent in Figure 2.2A.

To further investigate these preferences, we assessed how individual residues and positions contributed to fitness. Normalizing amino acid frequencies to their abundances in the unselected library revealed how residues increased or decreased in frequency after selection, and therefore which residues likely promote function (Figure 2.2B). Phenylalanine showed the strongest enrichment at each position, followed by most aliphatic amino acids and tryptophan. Disfavored amino acids were primarily charged. Most striking, however, is the lack of position-specificity: all sites share the same general pattern. This is in sharp contrast to the effects of individual mutations on the extant Sc-α2 protein, which exhibited strong position-specific effects, such as the strong dependence on F116 (Figure 2.1E). Thus, the fitness landscape for this protein-protein interaction using small deviations from the naturally occurring Sc-α2 is not broadly representative and may severely underestimate the range of possibilities.

While many amino acid preferences are not position-specific in the broad sense, some positions do exhibit quantitative differences. For example, phenylalanine seems universally beneficial at all positions, yet F114 was only slightly enriched while F120 was strongly favored. Similarly, selection against disfavored residues was stronger at this latter position than at other positions. This quantitative bias largely increased with each position (reading from N- to C-terminus) with the most C-terminal position (T120) being most significant. This analysis indicates that an amino acid's contribution to fitness is largely intrinsic (e.g. phenylalanines are broadly beneficial) but the quantitative effect depends on position or context.

Testing the fitness of Matα2 variants in their natural setting

The analysis described here relied on the ability of Matα2 variants to repress transcription of a CAN1 reporter construct. We tested several variants spanning the entire range of FC values by replacing the endogenous Matα2 locus with the variant and monitoring the resulting phenotype, namely, the ability

to mate, which requires repression of multiple a-specific genes by Matα2. Using a quantitative mating assay, we measured the mating efficiencies of α-cells bearing either wild-type or mutant Matα2. Cells with wild-type Matα2 were highly proficient at mating (Table 1). In contrast, replacing the MCM1 interaction region of Matα2 with that of *C. albicans*, which diverged prior to the emergence of the interaction, strongly reduced mating efficiency. Replacement with variants from our library resulted in a range of mating efficiencies. Cells bearing Matα2 variants with low FC scores mated as poorly as the *C. albicans*-like protein. Conversely, no difference was observed between the mating efficiencies of the highest scoring variant (FC=5.2, core sequence PCLRFVF) and wild-type Matα2. However, a variant with an intermediate FC of 1.6 which indicates proficient growth in the bulk canavanine competition was mating deficient. This discrepancy likely reflects the different requirements of these assays: repression of the a-specific genes for mating involves the binding of Matα2-MCM1 to varied cis-regulatory sequences at multiple genes while growth in canavanine involves binding a single sequence at CAN1. Thus, novel Matα2 variants that are proficient at repressing our CAN1 reporter also function as well as the wild-type Matα2 in its natural setting to repress the a-specific genes and allow mating.

Fitness effects are strongly context dependent and non-additive

A single phenylalanine or leucine in the mutated sequence seems to promote fitness at all positions (Figure 2.2B), but two phenylalanines occurred less frequently together than expected (Figure 2.2C). This was not the case for leucine, with leucines co-occurring as frequently as expected assuming independence. Such non-additive interactions are indicative of intramolecular epistasis, so we next examined epistasis more systematically.

Absent epistasis, amino acids favored (or disfavored) at one position should remain favorable regardless of adjacent residues. We therefore assessed amino acid favorability in different contexts. For example, among variants with a phenylalanine at the seventh position (7F), the pattern of favored/disfavored amino acids differs from that in the general population (Figure 2.2D). This context-

dependence is amino acid and position specific. To reveal how the pattern changes in the context of 7F, we further normalized the amino acid frequencies among the 7F subset to the overall frequencies post-selection (Figure 2.2E). Surprisingly, all previously favorable amino acids (e.g., aromatic residues) were less abundant in the presence of 7F, while disfavored amino acids (e.g., charged residues) were more abundant. Even prolines, which were generally selected against and known to be disruptive to protein secondary structure, were enriched among 7F variants. This pattern was highly idiosyncratic, however, with drastically different amino acid biases exhibited by fixing different amino acids at different positions. However, each position seemed to toggle between two distinct patterns: enrichment of aromatics and aliphatics and depletion of charged residues (as in Figure 2.2B), or the inverse (as for the 7F subset in Figure 2.2E). This conclusion can be visualized by correlating the position-specific amino acid frequencies of each subset (i.e., F1, W1, etc.) to the frequencies observed overall (Figure 2.2F). Notably, in every context the 7th position was strongly influential, and the effect depended entirely on the fixed amino acid. Thus, the successful protein-protein interactions are characterized by complex epistatic interactions, with amino acid preferences across the interface being strongly contingent on the amino acids at other positions, especially at the 7th position.

Rampant intra-molecular epistasis between all positions

We further probed the extent and nature of this epistasis. Given that many of the interactions with phenylalanine were both position- and amino acid-specific, we wished to quantify pairwise epistatic interactions between all amino acid states across the seven positions. To do so, we first determined the frequency of each amino acid at each position among the fit variants (Figure 2.3A). We then calculated the expected co-occurrence of each amino acid state pair assuming independence (i.e. no epistasis). Deviations from this expectation indicate a genetic interaction between residues (Figure 2.3B).

We found numerous instances of both positive and negative epistatic interactions (Figure 2.3B). For example, the heat map illustrating the relationship between the first and seventh positions reveals that

many amino acid pairs exhibit some degree of epistasis (Figure 2.3C). This pattern was consistent between replicate selections and was not observed in the absence of selection (Figure 2.3B & supplemental figure 2.2A). In sum, 29% of all pairs of amino acid states along the interface exhibited a significant epistatic interaction, with positive and negative interactions equally represented. Consistent with the analysis discussed above, epistatic interactions were notably stronger and more numerous between position 7 and all other positions (Figure 2.3D). 65% of all amino acid state pairs involving position 7 exhibited epistasis. This epistasis likely reflects distinct structural differences between positions along this interface, and likely influences how this interaction evolves.

As a final analysis of epistasis, we asked whether the naturally occurring Matα2 sequence that interacts with MCM1 (e.g., L1, V2, F3, etc.) is subject to more or less epistasis than the sequences that arose in our experimental system. We did not find significant differences between the wild-type amino acid states and our variants in this regard. This observation suggests that the amino acids found in *S. cerevisiae* are not distinct from the numerous other functional alternatives with regard to epistatic interactions.

The rugged fitness landscape for the Matα2-MCM1 interaction

Can the Matα2 variants examined in this study help us understand how this function evolved? Fitness landscapes describe how function relates to sequence changes, and the topography of such landscapes can be smooth or rough depending in part on interactions (epistasis) between mutations. Sign epistasis – whereby the sign of a mutation's fitness effect depends on the genetic background, thus positive in some contexts and negative in others – results in rugged fitness landscapes. This is known to constrain mutational paths toward higher fitness by stalling populations at local maxima which are difficult to move away from.

An empirical determination of a complete fitness landscape for seven resides as examined here is technically not possible because the sequence space is so vast. We instead relied on our sparse but broadly sampled sequence set to assess general features of the Matα2-MCM1 interaction landscape. We took

advantage of the ~11,000 pairs of variants in our library that differ by a single amino acid. Each pair represents a single mutational step sampled randomly from across the entire fitness landscape.

Using such pairs, we can ask how often a single amino acid mutation significantly altered function. Using the fitness of the naturally occurring Sc-α2 to define the functional threshold, we found that most of the single mutations did not significantly impact function: both sequences either remain fit or unfit (Figure 2.4A). Overall only 20% of mutation pairs exhibit a functional change (i.e. one is fit and the other unfit). When these pairs were grouped by the mutation's location, we find that not all positions are equal: 35% of mutations effect function at position 7, whereas only 15% at position 1 (Figures 2.4A & 2.4B). This independent analysis confirms the conclusions described above that position 7 exhibited the most epistasis (Figure 2.3D). This analysis also allowed us to ask whether particular amino acids were more or less likely to alter fitness. After grouping mutational pairs by their sequence difference, we found that the gain of specific residues drastically altered function (Figure 2.4A). For example, mutation of position 7 to a phenylalanine (7F) from any other amino acid was highly beneficial: in 90% of contexts (n=…), the 7F mutation resulted in a functional protein (by either retaining or gaining functionality). This pattern was not observed when phenylalanine was instead substituted at position 1, again highlighting the complex interaction between amino acid and position.

We can also learn something about the fitness landscape by considering all mutational pairs with at least one fit sequence, as these represent mutational steps on or around fitness peaks. The magnitude of the fitness difference between such mutational pairs describes the 'steepness' of the peak, while the landscapes overall 'smoothness' can be estimated from the fraction of all mutation pairs that don't compromise function. Of these fitness peak-associated pairs, we find that 40% of mutations compromise the function of random fit variants (i.e. reduce fitness below that of Sc-α2). This suggests that many functional sequences exist on broad fitness peaks where more than half of potential mutations are permissible.

Is mutational robustness also characteristic of the natural occurring Matα2 protein – such as Matα2 from *S. cerevisiae*? Since our randomized and point mutant libraries of Matα2 are in the same genetic background and used the same selection regime, we can directly compare the effects of single point mutations with the above. As with the random fitness peaks, we find that Matα2 from *S. cerevisiae* is similarly sensitive to mutation using the same criteria: 43% of all point mutants lost function. This similarity suggests that extant Matα2-MCM1 interactions occupy a fitness peak similar to the others probed here, at least in terms of mutational accessibility.

Discussion

In this work we investigated how ancient transcriptional regulators acquire protein-protein interactions that result in the cooperative control of a target gene. The natural interaction between Matα2 and MCM1 found in *S. cerevisiae* provides an excellent case study. We investigated many of the evolutionary 'paths not taken' by Matα2 to uncover the constraints – or seemingly lack thereof – governing the emergence of this interaction. There apparently exist many diverse protein interfaces capable of mediating a functional, cooperative interaction with MCM1. We relied on this chemically and structurally diverse set of novel MCM1-interacting interfaces to reveal rules and constraints, including widespread intramolecular epistasis, that underlie this diversity.

Many deep mutational scans focus primarily on the consequences of many single point mutations, while our approach allows us to assess combinations of mutations (Fowler and Fields, 2014; Hietpas et al., 2011; McLaughlin Jr et al., 2012; Roscoe et al., 2012, 2013). We feel this consideration is especially important when studying the evolution or engineering of proteins. Protein interactions are a distributed property involving multiple interfacial amino acids, and the mutation of which can have unpredictable results (Aakre et al., 2015; McClune et al., 2019; Podgornaia and Laub, 2015; Starr et al., 2017). The appearance of highly idiosyncratic (epistatic) mutational effects suggests that the behavior of single mutants poorly predicts the consequences of mutations in different contexts or for combinations of mutations.

When researchers have tested combinations of mutations, epistasis has frequently emerged in other studies of protein evolution (Aakre et al., 2015; Podgornaia and Laub, 2015; Pokusaeva et al., 2019; Sarkisyan et al., 2016; Starr et al., 2018). Notably, other kinds of pair-wise protein-protein interactions have revealed instances of intramolecular epistasis. For example, particular residues in the interface between the *E. coli* PhoQ protein kinase and its substrate PhoP exhibit epistatic interactions involving a handful of residues (Podgornaia and Laub, 2015). Likewise, the evolution of bacterial toxin-antitoxin proteins appears to frequently involve intermediate antitoxins that are promiscuous, allowing the antitoxin to simultaneously recognize and inactivate an ancestral cognate toxin as well as newly evolved toxin targets (Aakre et al., 2015). Such promiscuous intermediates facilitate the evolution of new toxin-antitoxin specificities by avoiding non-functional (or less-functional) intermediates. The emergence of this promiscuity frequently depended on epistatic interactions between residues, with the promiscuous intermediate state being contingent on the order of mutations (Aakre et al., 2015). Likewise, the epistasis we observe between residues of Matα2 appear to facilitate the emergence of a new function, namely the interaction with MCM1. The degree of epistasis observed here, however, is unprecedented. We found that every position of the interface, for multiple amino acid states, exhibits epistasis. Other instances of intramolecular epistasis have involved one or a few positions, and typically between specific amino acids (Aakre et al., 2015; Podgornaia and Laub, 2015; Puchta et al., 2016; Sarkisyan et al., 2016).

The functional paths through this sequence space are thus difficult to predict due to this epistasis. Nonetheless, the preference for aromatic and hydrophobic residues is striking. Our analysis of diverse Matα2 variants that differ by a single residue suggests that the addition (or removal) of a phenylalanine dramatically affects fitness regardless of its position. This is also the case for tryptophan and tyrosine but to a lesser extent, suggesting aromatic side chains generally promote the formation of this interaction. Notably, aromatic amino acids still epistatically interact with other residues, indicating that while such residues may be broadly beneficial, the neighboring amino acids modulate the effect.

The apparent rules governing the Matα2-MCM1 interaction are reminiscent of 'fuzzy' interactions involving activation domains (ADs) of some transcription factors (Brent and Ptashne, 1985; Tompa and Fuxreiter, 2008). The ectopic expression of ADs is sufficient to stimulate transcription and involves binding the Mediator complex (Brzovic et al., 2011; Kornberg, 2005; Tuttle et al., 2018). Akin to the Matα2-MCM1 interaction, 'fuzzy' binding involves degenerate interfaces that are enriched for hydrophobic residues and favor aromatic residues (Brzovic et al., 2011; Sanborn et al., 2021; Warfield et al., 2014). Besides bulky hydrophobic residues, ADs are frequently negatively charged; the opposite is observed among Matα2 variants able to function with MCM1. This suggests AD- and Matα2-mediated interactions may have a similar basis despite the apparent binding geometries being fundamentally different. 'Fuzzy' AD binding involves multiple sites on Mediator, with ADs capable of binding in multiple discrete conformations, while Matα2 binding MCM1 likely involves a single site (Brzovic et al., 2011; Sanborn et al., 2021). Though it remains possible that some Matα2 variants identified here may bind MCM1 in a manner distinct from that of the *S. cerevisiae* protein, and may involve more than one conformation.

Why do some protein interaction interfaces exhibit epistasis and functional degeneracy while others are more constrained? It's possible the evolutionary histories of these proteins influence how these interactions emerge. The above antitoxin proteins, for example, have evolved under strong positive selection to maintain tight toxin binding even if the affinities are promiscuous (Aakre et al., 2015). Because loss of toxin protection results in a strong growth defect, this may promote the evolution of mutational robustness in antitoxin proteins and profoundly change the evolvability of the system. Conversely, the gain of a novel interaction interface in an unconstrained and highly variable protein region – as with Matα2 – may have fundamentally different constraints.

Finally, we bring up an apparent paradox raised by our results: if so many solutions exist for a functional interaction between Matα2 and MCM1, why is the naturally occurring solution preserved across a whole clade of closely related species. We consider three possibilities. (1) There is some other function for this region of Matα2 that constrains its sequence. This possibility seems unlikely given the detailed

biochemical, genetic, and structural studies of Matα2 over the past three decades, which has revealed its only role is interacting with MCM1 to promote mating by repressing the a-specific genes (Mead et al., 2002, 1996; Vershon and Johnson, 1993). The fact that one of our randomly chosen *de novo* variants (which lacked any resemblance to the naturally occurring sequence) functioned and properly recapitulated Matα2-MCM1's role in promoting mating does argue against an "unknown" function, at least one that would have a noticeable consequence on the cell. (2) The MCM1-interacting sequence of Matα2 is constrained to prevent promiscuous interaction of Matα2 with other transcriptional regulators. Over the short term, we know that the de novo functional Matα2's did not cause a noticeable fitness defect because in the control experiments (that is, in the absence of canavanine) their representation in the library did not increase or decrease. However, it is possible over evolutionary timescales that small defects caused by promiscuous Matα2's could constrain its sequence. (3) The naturally occurring sequence is precisely optimized so that even small changes which weaken or strengthen the interaction are selected against over long evolutionary times.

In conclusion, the purpose of this study was not to trace the historical evolution of an extant protein-protein interaction that allowed two ancient transcriptional regulators to cooperatively regulate gene expression. Rather, the purpose was to explore the range of possible solutions that could create such a functional interaction. Our finding that approximately 40% of random amino acid sequences can recapitulate a functional interaction between two ancient proteins suggests that new interactions between existing transcriptional regulators are sampled continuously over evolutionary time, some of which are maintained by selection. We suggest that the weak, degenerate nature of the pair-wise cooperative interaction observed here is broadly applicable and could facilitate the relatively rapid rewiring of transcription networks and the arrival of new phenotypes.

Materials and Methods

S. cerevisiae strain construction

All *S. cerevisiae* strains were in the S288C background and grown on yeast extract peptone dextrose (YEPD) media at 30°C unless otherwise indicated. Transformations were conducted using the standard lithium acetate/polyethylene glycol method (Gietz and Woods, 2002). In the S288C **a** cell, the CAN1 gene was engineered to be repressed by Matα2-MCM1 by inserting immediately upstream of the CAN1 ORF a PCR product amplified from pKF145 using oKF437 and oKF438. This PCR product contains part of the CYC1 promoter with a Matα2-MCM1 cis-regulatory sequence from STE2 inserted upstream of its transcriptional start site. The resulting strain yKF230 constitutively expresses CAN1 in the absence of Matα2 but strongly repressed by Matα2-MCM1. Deletion of the silent Matα locus (HML) in yKF230 used a NatR marker amplified from pFA6a-natMX with homology to HML and resulted in yKF231. The initial Matα2 selection screen was done in yKF230 (in which HML is intact) with all subsequent work being done in yKF231 (HMLΔ). All genetic manipulations were confirmed by PCR and DNA sequencing.

Matα2 expression plasmid construction

Matα2 and its endogenous promoter were synthesized as gBlocks (IDT) and ligated together before being inserting into the AscI site of pRNDM (Addgene), a compact CEN/ARS plasmid, to generate pKF146. This Matα2 expression plasmid was subsequently digested with NdeI and AgeI and oKF457 was inserted to generate a mutant Matα2 with a unique EcoRV site in place of the MCM1 interaction region (pKF154). This served as an efficient 'landing pad' for the subsequent insertion of various DNA sequences, which eliminates the EcoRV site in the process of regenerating either the wild-type Matα2 DNA sequence or variants containing mutations. A separate silent mutation was also introduced nearby at I124 (codon ATA to ATC) to differentiate the Matα2 construct from any chromosomal gene sequence. Regeneration of full length Matα2 was accomplished using the NEBuilder HiFi DNA Assembly master mix (New England

Biolabs) following EcoRV digestion of pKF154. The wild-type *S. cerevisiae* Matα2 protein was assembled using oKF458 and oKF459. To test if the homologous region of Matα2 from *C. albicans* was capable of interacting with MCM1, a chimeric protein was constructed using oKF460 and oKF461 which substitutes the *C. albicans* amino acid sequence SPFSNSADT in place of the *S. cerevisiae* sequence GLVFNVVTQDM.

Design of Matα2 libraries

The assembly of Matα2 mutant libraries was conducted as above using NEBuilder HiFi DNA Assembly master mix (New England Biolabs) but used degenerate oligonucleotide pools synthesized by Integrated DNA Technologies (IDT). Two different mutant libraries were generated. The first consisted of single amino acid changes at each position in the *S. cerevisiae* MCM1 interaction region. This was accomplished by annealing pairs of oligonucleotides (oKF521-542) in which each codon in the 11 amino acid MCM1 interaction region has been separately replaced by an NNK codon (where N indicates an A, C, G, or T, and K indicates a G or T). This resulted in 11 separate plasmid pools (pKF159-169) each with 32 possible DNA sequences at each NNK codon (4×4×2). In the second library, a single pair of annealed oligonucleotides (oKF462 and oKF463) containing seven consecutive NNK codons was used to randomize positions L114-T120, generating library pKF157. This second library consists of many distinct Matα2 variants, each with all seven core residues randomized, and the total possible number of combinations (327) exceeding 34×109.

Following assembly, these Matα2 mutant constructs were electroporated into 5-alpha electrocompetent *E. coli* cells (New England Biolabs) according to the manufacturer's instructions. The first library of codon point mutations were transformed separately for each codon and the cells then pooled. To maximize the number of transformants from the second (7×NNK) randomization library, ten electroporations were done in parallel and later pooled. Immediately after electroporation pre-warmed SOC media was added to the cells and they recovered for one hour at 37°C. Following recovery, we determined

the transformation efficiencies and library complexities by diluting an aliquot of cells and plating on LB media supplemented with 30 µg/mL kanamycin. The single codon, point mutation library reached saturation with every DNA mutation being represented several times over in the transformed stock. We estimate the complexity of the second 7×NNK library to be ~2.92 million unique transformants.

To select *E. coli* transformants in bulk, each library was used to inoculate 500 mL LB media supplemented with 50 µg/mL kanamycin and grown overnight shaking at 37°C. The next morning, plasmid was purified from each culture using ten QIAprep spin miniprep columns (QIAGEN). The saturated overnight *E. coli* cultures were also used to make 1 mL and 10 mL library freezer stocks stored at -80°C for later use.

The purified plasmid libraries were used to transform yKF230 or yKF231 to G418R according to a high-efficiency yeast transformation protocol (Benatuil et al., 2010). Prior to transformation, a plasmid bearing the wild-type *S. cerevisiae* Matα2 was spiked into both libraries at a frequency of 1/1000. For the more complex 7×NNK library, six transformations were done in parallel and then pooled. Transformants were selected in bulk by adding each transformation to 500 mL YEPD media supplemented with 200 µg/mL G418 sulfate and growing overnight shaking at 30°C. An aliquot of transformed cells were also diluted and spread on YEPD plates containing 200 µg/mL G418 sulfate to determine transformation efficiencies. Transformation with the point mutation library resulted in ~1.1 million transformants; the pooled 7×NNK library had ~1.88 million transformants. Following overnight growth, 5 mL of saturated yeast culture was combined with 5 mL 50% glycerol to make freezer stocks, which were subsequently stored at -80°C for later use.

Canavanine selection assay and sequencing

Frozen yeast library aliquots were thawed on ice and then added to 500 mL YEPD. Library cultures were grown overnight shaking at 30°C to allow for the equal expansion of all Matα2 variants. The following morning, cells were collected for sequencing by pelleting 45 mL of saturated culture, washing once in PBS,

and then freezing the cell pellets at -80°C for later plasmid purifications. This is the 'Pre' library and provides the starting frequencies of each variant. Selection for Matα2 variants capable of functionally interacting with MCM1 was carried out by diluting 5 mL of saturated overnight culture into 500 mL synthetic medium lacking arginine and supplemented with either 25 or 250 µg/mL L-canavanine ('low' and 'high' treatment conditions, respectively). To control for growth differences due to the plasmid or protein expression, the same media lacking L-canavanine was used to start a third culture ('zero' treatment) in which there is no selection on Matα2-MCM1 function. After 24 hours of growth in either zero, low, or high canavanine, cells were pelleted in 45 mL aliquots and the pellets frozen at -80°C. Note the low and high canavanine selections resulted in very similar results and thus we focused our analysis on the high selection condition. All figures use the high canavanine selection data and any conclusions were corroborated using the low canavanine data.

The Matα2 plasmid library was recovered by boiling the cell pellets for 5 minutes and then bead-beating in the presence of phenol and chloroform. DNA was ethanol precipitated and purified using a QIAprep miniprep kit (QIAGEN). To prepare sequencing libraries, PCR was used amplify the variable region of Matα2 from the plasmid pool using primers containing Illumina adapter sequences and sample specific barcodes. For each library, ten 50 µL PCR reactions were pooled after nine PCR cycles. Each 500 µL PCR pool was cleaned and concentrated using a MinElute PCR Purification Kit (QIAGEN) and quantified using either a Bioanalyzer or TapeStation (Agilent). Massively parallel sequencing was carried out using single end 50 base pair reads on an Illumina HiSeq 4000 sequencing system.

Data processing and analysis

Processing of sequence data was done carried out in the programming language R using Bioconductor and custom scripts. Sequence reads were first trimmed based on their Phred quality scores to remove bases distal to the point a read's quality drops below 20 (corresponding to 1% error). Trimmed sequences were then aligned to the Matα2 reference sequence and reads were eliminated if they did not

fully span the mutated region. Furthermore, reads were removed if they contained indels or any mutation outside the mutated region. Sequences passing these quality filters were tabulated and the counts normalized to the sequencing depth (i.e., reads per million). The normalized counts for each unique Matα2 sequence were then compared across all conditions. Fold-changes were calculated by taking the ratio of the normalized counts after selection over the frequency prior to selection (or without selection using the zero canavanine data). Matα2 variants were eliminated from most analyses if the variant was sequenced less than 10 times either before or after selection.

Quantitative mating assays

Quantitative mating assays were performed according to previously described methods for yeasts (Sprague, 1991). Strains to test were created by replacing the endogenous Matα2 locus with specific variants identified from our library. The chosen variants span the full range of fold-change values and were highly reproducible between replicate experiments. Specific mutations were first cloned into pKF154, the Matα2 expression plasmid used above, as previously described to generate the mutant library. Plasmids with these new variants were digested with NdeI and NruI to liberate a DNA fragment containing Matα2 and its promoter and a KanMX resistance marker. PCR primers were used to add sequence homology for the mating-type locus to the end of the KanMX gene. These DNA fragments were then used to transform yKF249, an α-cell from the W303 background with Matα2 replaced by the URA3 gene, generation strains yKF270-282. As controls, we also introduced the wild-type *S. cerevisiae* Matα2 gene (yKF266) and a variant with the *C. albicans* region (yKF268) which is unable to interact with MCM1. All genotypes were confirmed by PCR and sequencing.

The α-cells created above bearing Matα2 variants were Trp- G418R and were mated to Trp+ a-cells. For each mating, the strains were grown to mid-log phase and their OD600 measured. Cells were then combined with an a:α ratio of 10:1 and concentrated onto 0.8 μm nitrocellulose filters using a Millipore 1225 Vacuum Sampling Manifold. The filters were then placed on YEPD agar plates and incubated for

either 4 or 16 hours at 30°C to allow mating. The filters were then vortexed in 5 mL water to resuspend cells for plating. Dilutions were first plated on YEPD plates containing 200 µg/mL G418 sulfate and grown for 2 days at 30°C to select for conjugants and the limiting parental strain. These G418R colonies were counted and then replicated to SD-Trp to select for conjugants only. The Trp+ G418R colonies were counted and mating efficiencies calculated as follows: Mating efficiency = (number of Trp+ G418R colonies) / (total number of G418R colonies).

Growth competition assays

Two Matα2 variants from our library appeared to grow slowly in the absence of canavanine, suggesting that some Matα2 mutations may be detrimental to growth. Both these variants involve a single amino acid change to the wild-type *S. cerevisiae* sequence and are at the same position (N117I and N117V). To test if these specific mutations impact growth, we introduced them into the endogenous Matα2 locus as above for quantitative mating assays. We then carried out growth competitions in an attempt to measure even subtle growth differences. Each mutant was competed against an isogenic parental strain bearing wild-type Matα2 and constitutively expressing mCherry. For each competition, the mutant strain (N117I or N117V) and mCherry competitor were grown separately to saturation in liquid YEPD. Their OD600 was then measured and combined at a 1:1 ratio in a final volume of 1 mL. This mixture was then diluted to OD600 = 0.1 in synthetic (SD) media lacking arginine, which was the media condition in which the slow growth was first observed, and grown overnight at 30°C. The cultures were back diluted to OD600 = 0.1 the next morning and grown again overnight at 30°C. This repeated growth and dilution process continued for 5 days. With each daily passaging, cells were also removed and counted on a BD FACSCelesta flow cytometer. The relative growth rate of the non-fluorescent mutant strain and the mCherry wild-type strain were then determined.

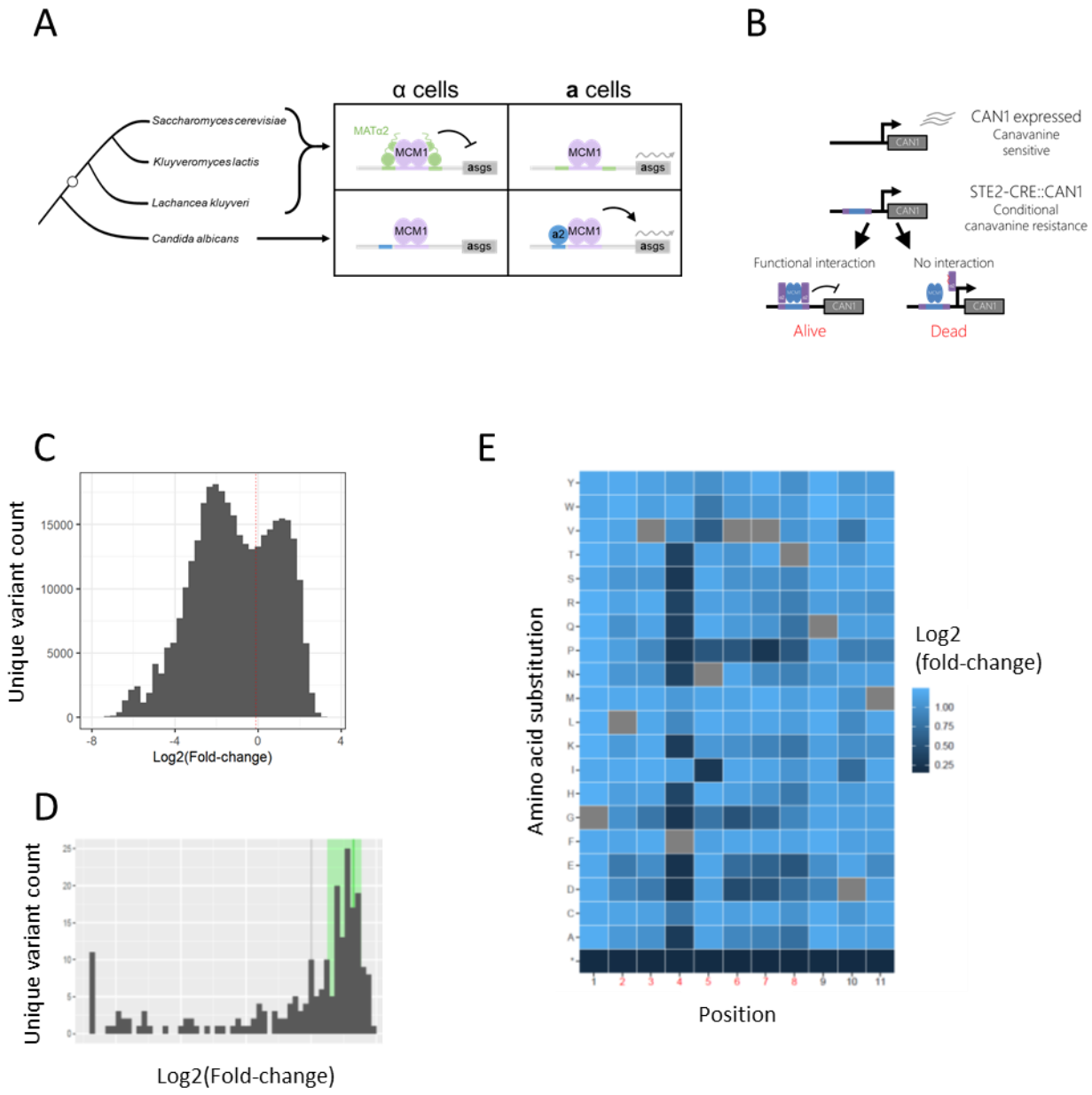


Figure 2.1 Evolution of the Matα2-MCM1 interaction and the discovery of novel functional variants

A. Representative ascomycete yeast species and their mode of a-specific gene regulation. Functional Matα2-MCM1 complexes are found in *S. cerevisiae* and related species. Outside the *L. kluyveri* to *S. cerevisiae* clade, the mode of a-specific gene regulation is different: an activator made only in a-cells induces the genes which aren't otherwise expressed.

B. Diagram of the selection for functional Matα2-MCM1 complexes. CAN1 expression in the presence of canavanine is toxic, but it's repression by Matα2-MCM1 allows for robust growth.

C. Fitness distribution of randomized variants post-selection. Histogram showing the log2 fold-changes for variants in our randomized library. Each protein variant contains seven random

residues at positions 114-120. Dashed red line indicates the fold-change of the wild-type Matα2 from *S. cerevisiae*. Those to the right of the line show greater enrichment than the wild-type sequence.

D. Fitness distribution of point mutants to the wild-type *S. cerevisiae* Matα2. Histogram showing the log2 fold-changes for all point mutants and controls. Mean fold-change of all variants is indicated by the black dashed line; fold-change of wild-type *S. cerevisiae* Matα2 indicated by green dashed line. Green shaded region designates the range of fold-change values for variants with synonymous mutations.

E. Heat-map showing log2 fold-change values for point mutants of the wild-type *S. cerevisiae* Matα2. The effect of the indicated amino acid substitution (y-axis) at each position (x-axis) is colored by its fold-change after selection compared with no selection. Premature stop codons are indicated by * (bottom row). Grey boxes denote the wild-type amino acids.

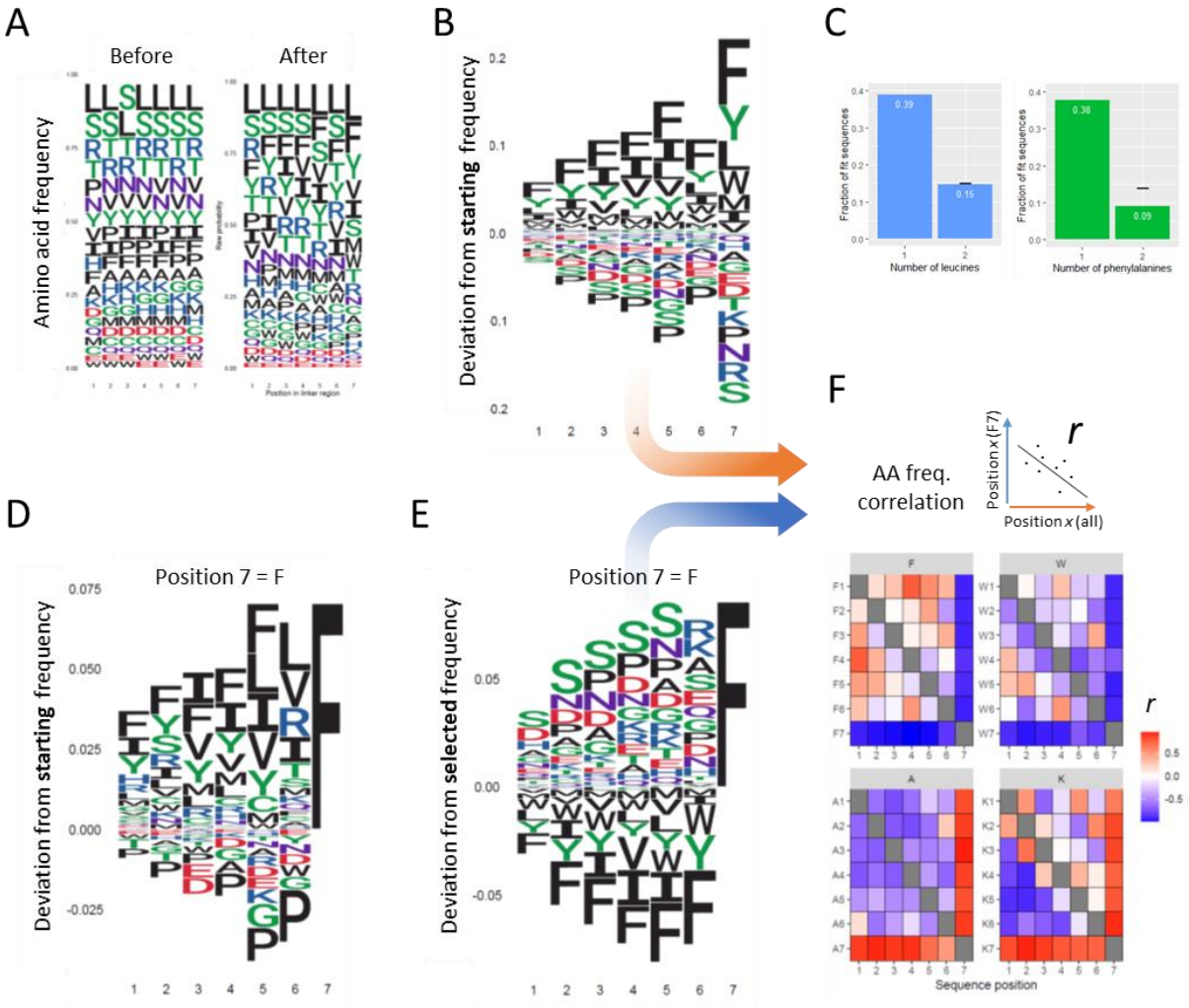


Figure 2.2 Functional Matα2 proteins with highly degenerate interaction interfaces

A. Logo plot showing the amino acid complexity of the randomized library before (left) and after (right) canavanine selection. The size of each amino acid abbreviation corresponds with its abundance in the population and are ordered from most (top) to least (bottom) abundant.

B. Normalized logo plot showing changes in amino acid abundance at each position after selection. Positive values indicate amino acids that increased in frequency, while frequency decreases are negative. The size of each letter corresponds with the magnitude of the change and are ordered as in (A).

C. The co-occurrence of some amino acids is underrepresented among the functional variants. The black horizontal line in the second column indicates the expected frequencies of double leucine (left) or double phenylalanine (right) containing variants given the frequencies of each alone (left bar). Note that while the number of functional variants containing two leucines is approximately the square of the single leucine frequency, double phenylalanine variants are less abundant than expected given independence.

D. Favorable amino acid compositions are heavily context dependent. Normalized logo plot as in (B) generated from the subset of variants with a phenylalanine at position 7 (F7).

E. Influence of F7 on adjacent amino acids. Logo plot of F7 sequences normalized to post-selection amino acid frequencies showing the effect of F7 alone. The preferred amino acid composition among these variants contrasts strongly with the overall pattern in (B).

F. Correlation between position-specific amino acid frequencies. For different sequence subsets (e.g., F1, W1, etc.), the change in amino acid frequencies (e.g., panel E) was correlated with the overall amino acid pattern (e.g., panel B) for each position separately and plotted as a heat map. Constant positions in each subset were excluded (indicated by grey boxes in the heat map).

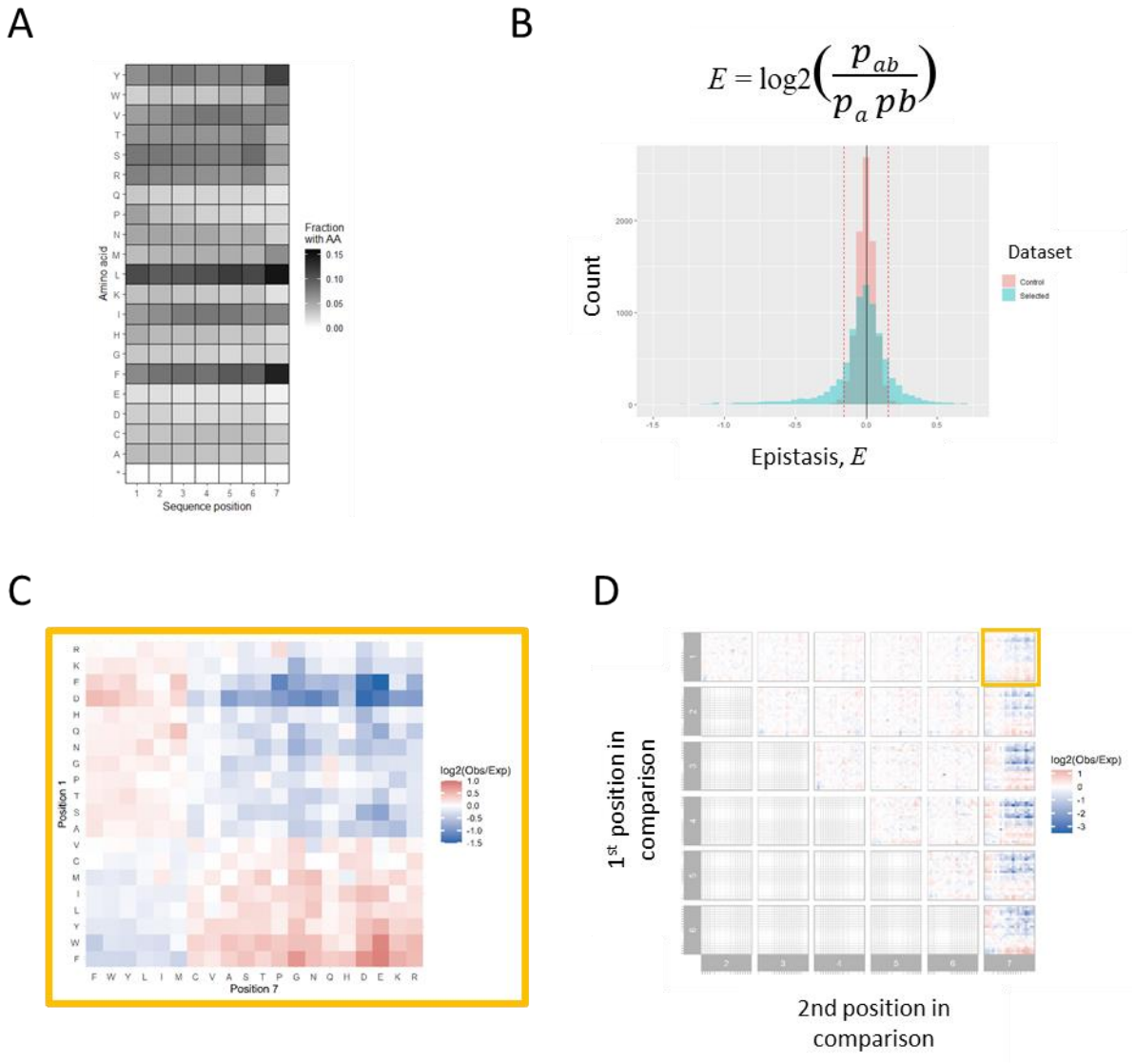


Figure 2.3 Functional Matα2-MCM1 interactions exhibit rampant intra-domain epistasis

A. The frequency of amino acids across all positions among functional variants.

B. Histogram of the effects of epistasis on the frequency of amino acid pairs across all positions relative to expected frequencies given independence. Values calculated using unselected control sequences (red) were used to define a 95% confidence interval (red dashed lines).

C. Heat map showing the frequency of amino acid pairs at positions 1 & 7 relative to the frequency given independence as in (B).

D. Heat maps as in (C) comparing all pairwise positions. Plot from (C) is indicated by an orange box.

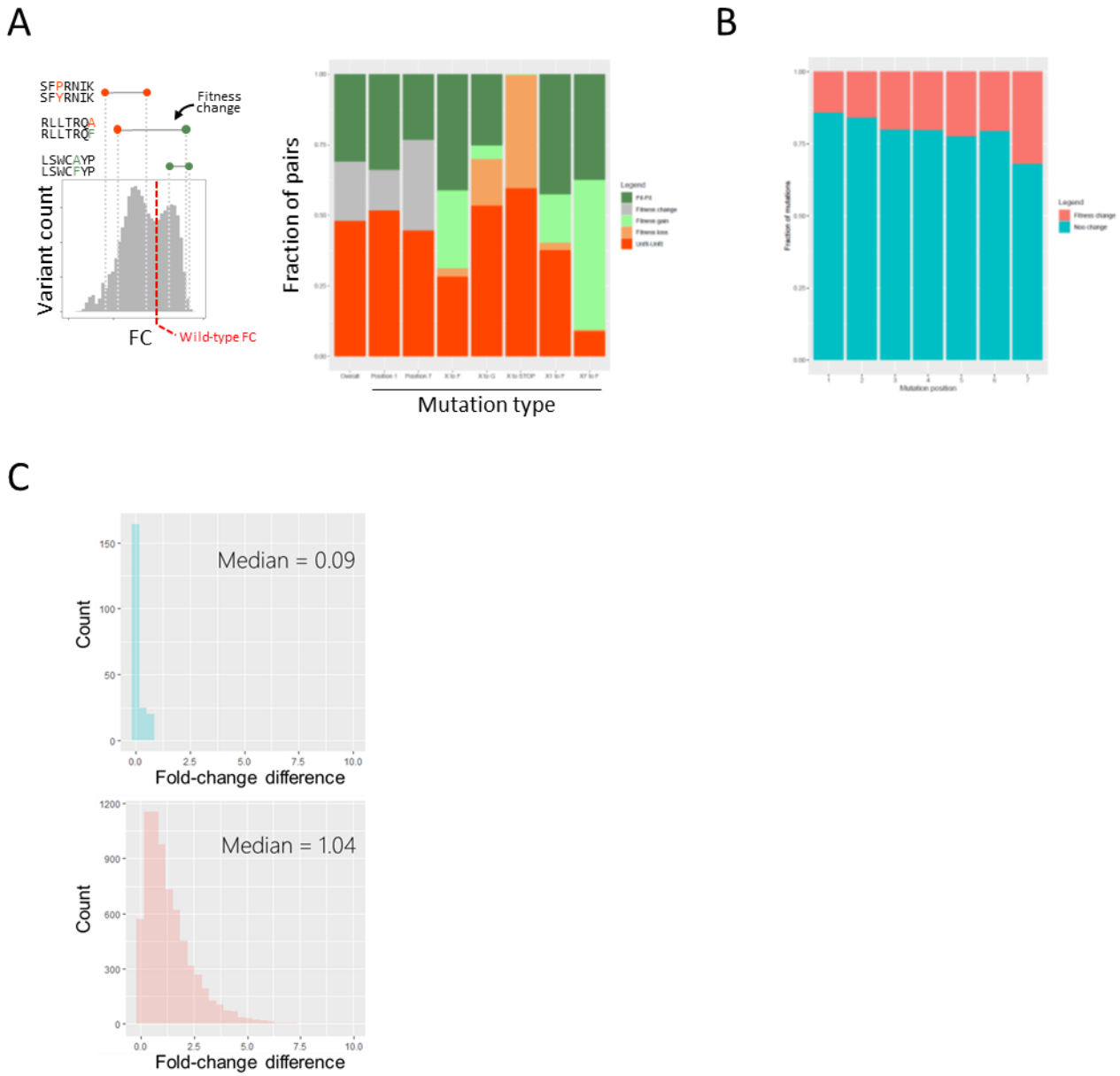
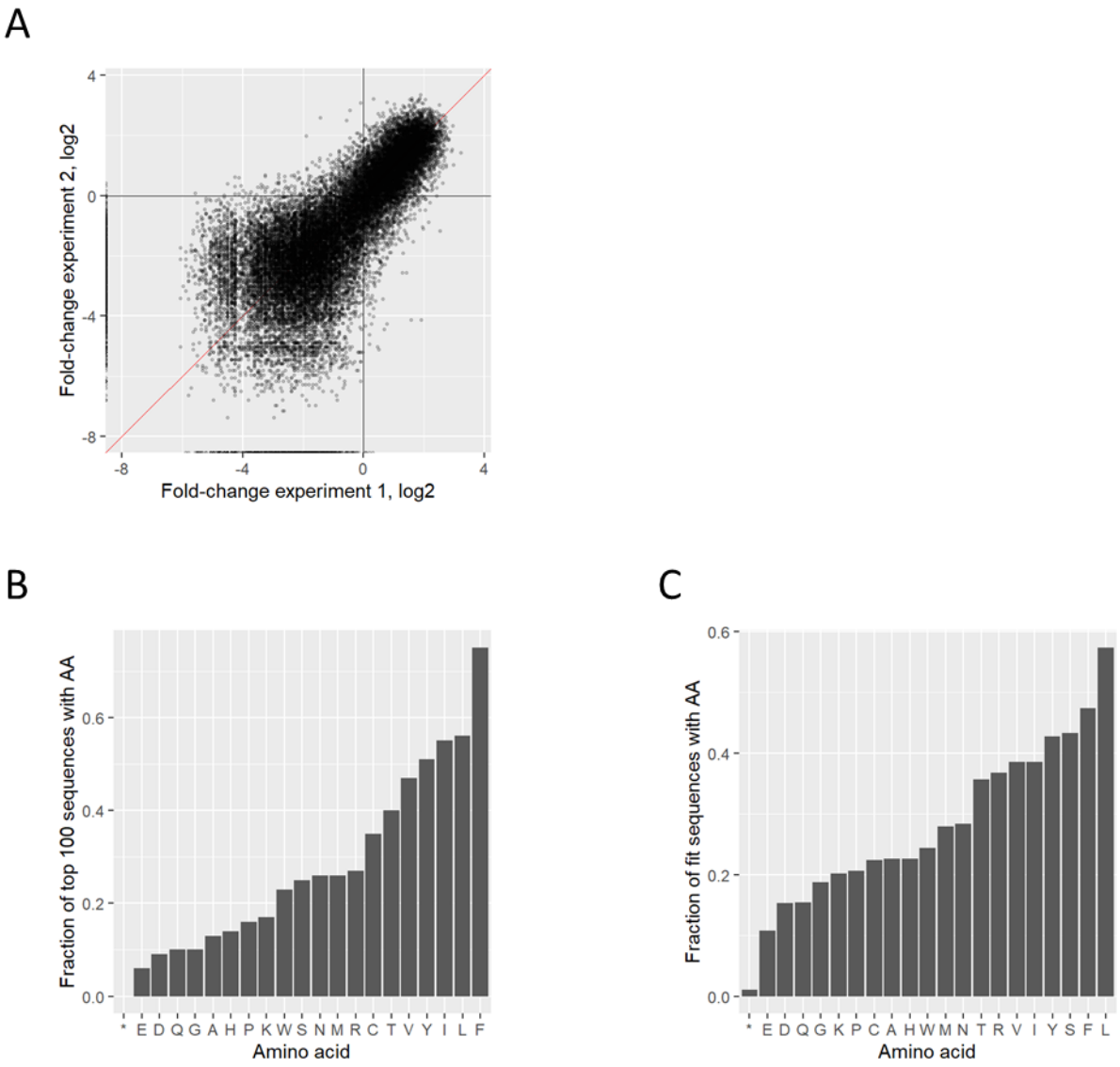


Figure 2.4 The fitness landscape of the Matα2-MCM1 interaction

A. Effects of single amino acid mutations based on variant pairs that differ by a single residue. Variant pairs are grouped based on their fitness (left) and then by mutation type (e.g., mutating to phenylalanine) and/or its position (e.g., mutation the 7th position to phenylalanine). The FC of wild-type Matα2 was used to define fit vs. unfit.

B. Fitness effects based on mutation position. As in (A), fitness change is relative to the wild-type Matα2 from *S. cerevisiae.*

C. Magnitude of the mutational effects. Point mutations to the wild-type Matα2 from *S. cerevisiae* (top) effect fitness less than among random variant pairs (bottom).

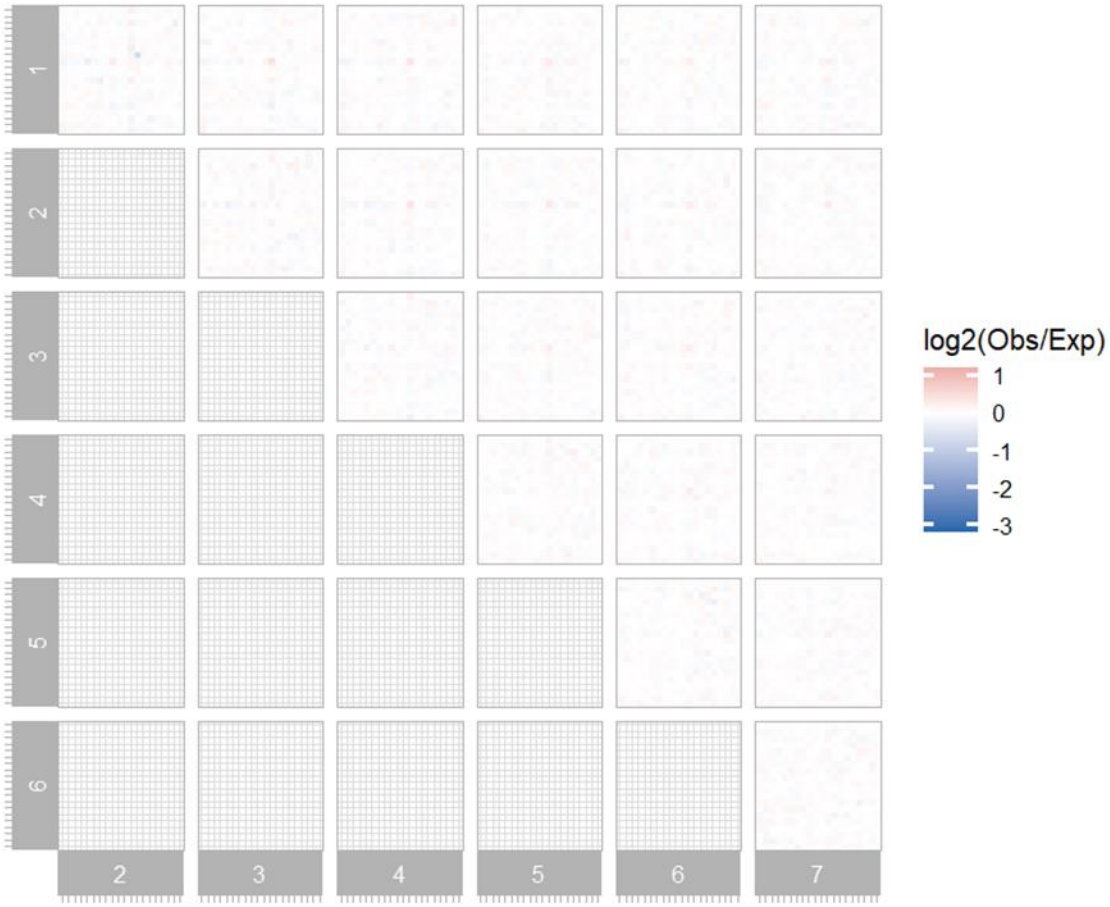


Supplementary figure 2.1

A. Reproducibility of fitness measurements of Matα2 variants. 25,000 variants from the randomized library were chosen at random and the fold-changes observed in both replicate experiments are plotted. Variants that were recovered in only one experiment are represented by points along the plot margin. The red line indicates y = x.

B. Frequency of variants with each amino acid among the 100 most fit variants. The 100 variants with the greatest fold-change were grouped and the fraction with each amino acid is indicated.

C. Frequency of variants with each amino acid among all fit variants. As in (B) but tabulating all fit variants.



Supplementary figure 2.2
Epistatic interactions depend on selection. The library was grown in media lacking canavanine and variant fold-changes were used to calculate epistatic interactions as in Figure 2.3C & 2.3D. The color scale matches that of Figure 2.3D.

Table 1. Matα2 variant mating efficiencies

Strain	Alpha2 allele	% mating	Mean % mating		Fold-change
yKF266	*S. cerevisiae*	72	71		0.93
yKF266	*S. cerevisiae*	70			
yKF268	*C. albicans*	0.1	0.1		n/a
yKF268	*C. albicans*	0.1			
yKF270	FV*T*VF	0	0		0.03
yKF270	FV*T*VF	0			
yKF272	TLMRERP	0	0		0.13
yKF272	TLMRERP	0			
yKF274	FMSISNT	0	0		0.54
yKF274	FMSISNT	0			
yKF276	ILNWFTL	0	0.05		1.6
yKF276	ILNWFTL	0.1			
yKF278	PCLRFVF	57	56.5		5.2
yKF278	PCLRFVF	56			
yKF280	N5V	45	65.5		0.6
yKF280	N5V	86			
yKF282	N5I	59	53.5		0.3
yKF282	N5I	48			

Table 2. Saccharomyces cerevisiae strains used in this chapter
All strains are in the indicated background and the indicated genotypes denote modifications to the given background. Note for yKF270-282, the amino acid sequence of the MCM1 interaction region is indicated. These variants are in the background of the *S. cerevisiae* Matα2 protein.

Strain ID	**Genotype**	**Genetic background**
yKF230	*MATa hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-CAN1*	S288C
yKF231	*MATa hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-CAN1 hmlΔ::NatMX*	S288C
yKF249	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::URA3*	W303
yKF266	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{Scer}$	W303
yKF268	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{Calb}$	W303
yKF270	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{FV*T*VF}$	W303
yKF272	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{TLMRERP}$	W303
yKF274	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{FMSISNT}$	W303
yKF276	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{ILNWFTL}$	W303
yKF278	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{PCLRFVF}$	W303
yKF280	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{N5V}$	W303
yKF282	*MATα ura3Δ::hph-*$P_{CYC1\ (Scer\ STE2\ CRE)}$*-GFP matα2Δ::MATα2*$_{N5I}$	W303

References

Aakre, C.D., Herrou, J., Phung, T.N., Perchuk, B.S., Crosson, S., Laub, M.T., 2015. Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates. Cell 163, 594–606. https://doi.org/10.1016/j.cell.2015.09.055

Baker, C.R., Booth, L.N., Sorrells, T.R., Johnson, A.D., 2012. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. Cell 151, 80–95. https://doi.org/10.1016/j.cell.2012.08.018

Benatuil, L., Perez, J.M., Belk, J., Hsieh, C.-M., 2010. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel 23, 155–159. https://doi.org/10.1093/protein/gzq002

Brent, R., Ptashne, M., 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell 43, 729–736. https://doi.org/10.1016/0092-8674(85)90246-6

Britton, C.S., Sorrells, T.R., Johnson, A.D., 2020. Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit. Science 367, 96–100. https://doi.org/10.1126/science.aax5217

Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., Hahn, S., 2011. The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex. Molecular Cell 44, 942–953. https://doi.org/10.1016/j.molcel.2011.11.008

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., Wolfe, K.H., 2004. Evolution of the *MAT* locus and its Ho endonuclease in yeast species. Proc. Natl. Acad. Sci. U.S.A. 101, 1632–1637. https://doi.org/10.1073/pnas.0304170101

Fowler, D.M., Fields, S., 2014. Deep mutational scanning: a new style of protein science. Nat Methods 11, 801–807. https://doi.org/10.1038/nmeth.3027

Gietz, R.D., Woods, R.A., 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method, in: Methods in Enzymology. Elsevier, pp. 87–96. https://doi.org/10.1016/S0076-6879(02)50957-5

Gocke, E., Manney, T.R., 1979. EXPRESSION OF RADIATION-INDUCED MUTATIONS AT THE ARGININE PERMEASE (*CAN1*) LOCUS IN *SACCHAROMYCES CEREVISIAE*. Genetics 91, 53–66. https://doi.org/10.1093/genetics/91.1.53

Hietpas, R.T., Jensen, J.D., Bolon, D.N.A., 2011. Experimental illumination of a fitness landscape. PNAS 108, 7896–7901. https://doi.org/10.1073/pnas.1016024108

Johnson, A.D., 2017. The rewiring of transcription circuits in evolution. Current Opinion in Genetics & Development 47, 121–127. https://doi.org/10.1016/j.gde.2017.09.004

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., Taipale, J., 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature 527, 384–388. https://doi.org/10.1038/nature15518

Kornberg, R.D., 2005. Mediator and the mechanism of transcriptional activation. Trends in Biochemical Sciences 30, 235–239. https://doi.org/10.1016/j.tibs.2005.03.011

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.-B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A., 2002. Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. Science 298, 799–804. https://doi.org/10.1126/science.1075090

Li, H., Johnson, A.D., 2010. Evolution of Transcription Networks — Lessons from Yeasts. Current Biology 20, R746–R753. https://doi.org/10.1016/j.cub.2010.06.056

McClune, C.J., Alvarez-Buylla, A., Voigt, C.A., Laub, M.T., 2019. Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. Nature 574, 702–706. https://doi.org/10.1038/s41586-019-1639-8

McLaughlin Jr, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S., Ranganathan, R., 2012. The spatial architecture of protein function and adaptation. Nature 491, 138–142. https://doi.org/10.1038/nature11500

Mead, J., Bruning, A.R., Gill, M.K., Steiner, A.M., Acton, T.B., Vershon, A.K., 2002. Interactions of the Mcm1 MADS Box Protein with Cofactors That Regulate Mating in Yeast. Mol Cell Biol 22, 4607–4621. https://doi.org/10.1128/MCB.22.13.4607-4621.2002

Mead, J., Zhong, H., Acton, T.B., Vershon, A.K., 1996. The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. Molecular and Cellular Biology 16, 2135–2143. https://doi.org/10.1128/MCB.16.5.2135

Monahan, K., Schieren, I., Cheung, J., Mumbey-Wafula, A., Monuki, E.S., Lomvardas, S., 2017. Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. eLife 6, e28620. https://doi.org/10.7554/eLife.28620

Nobile, C.J., Fox, E.P., Nett, J.E., Sorrells, T.R., Mitrovich, Q.M., Hernday, A.D., Tuch, B.B., Andes, D.R., Johnson, A.D., 2012. A Recently Evolved Transcriptional Network Controls Biofilm Development in Candida albicans. Cell 148, 126–138. https://doi.org/10.1016/j.cell.2011.10.048

Pilpel, Y., Sudarsanam, P., Church, G.M., 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29, 153–159. https://doi.org/10.1038/ng724

Piskacek, S., Gregor, M., Nemethova, M., Grabner, M., Kovarik, P., Piskacek, M., 2007. Nine-amino-acid transactivation domain: establishment and prediction utilities. Genomics 89, 756–768. https://doi.org/10.1016/j.ygeno.2007.02.003

Podgornaia, A.I., Laub, M.T., 2015. Pervasive degeneracy and epistasis in a protein-protein interface. Science 347, 673–677. https://doi.org/10.1126/science.1257360

Pokusaeva, V.O., Usmanova, D.R., Putintseva, E.V., Espinar, L., Sarkisyan, K.S., Mishin, A.S., Bogatyreva, N.S., Ivankov, D.N., Akopyan, A.V., Avvakumov, S.Y., Povolotskaya, I.S., Filion, G.J., Carey, L.B., Kondrashov, F.A., 2019. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLOS Genetics 15, e1008079. https://doi.org/10.1371/journal.pgen.1008079

Puchta, O., Cseke, B., Czaja, H., Tollervey, D., Sanguinetti, G., Kudla, G., 2016. Network of epistatic interactions within a yeast snoRNA. Science 352, 840–844. https://doi.org/10.1126/science.aaf0965

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C.O., Forrest, A.R.R., Gough, J., Grimmond, S., Han, J.-H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C.R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R.D., Tegnér, J., Lenhard, B., Teichmann, S.A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D.A., Ideker, T., Hayashizaki, Y., 2010. An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. Cell 140, 744–752. https://doi.org/10.1016/j.cell.2010.01.044

Roscoe, B., Bolon, D.N.A., Jiang, L., Hietpas, R., 2012. Fitness analyses of all possible point mutations for regions of genes in yeast. Nature Protocols 7, 1382. https://doi.org/10.1038/nprot.2012.069

Roscoe, B.P., Thayer, K.M., Zeldovich, K.B., Fushman, D., Bolon, D.N.A., 2013. Analyses of the Effects of All Ubiquitin Point Mutants on Yeast Growth Rate. Journal of Molecular Biology 425, 1363–1377. https://doi.org/10.1016/j.jmb.2013.01.032

Sanborn, A.L., Yeh, B.T., Feigerle, J.T., Hao, C.V., Townshend, R.J., Lieberman Aiden, E., Dror, R.O., Kornberg, R.D., 2021. Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. eLife 10, e68068. https://doi.org/10.7554/eLife.68068

Sarkisyan, K.S., Bolotin, D.A., Meer, M.V., Usmanova, D.R., Mishin, A.S., Sharonov, G.V., Ivankov, D.N., Bozhanova, N.G., Baranov, M.S., Soylemez, O., Bogatyreva, N.S., Vlasov, P.K., Egorov, E.S., Logacheva, M.D., Kondrashov, A.S., Chudakov, D.M., Putintseva, E.V., Mamedov, I.Z.,

Tawfik, D.S., Lukyanov, K.A., Kondrashov, F.A., 2016. Local fitness landscape of the green fluorescent protein. Nature 533, 397–401. https://doi.org/10.1038/nature17995

Sorrells, T.R., Johnson, A.D., 2015. Making sense of transcription networks. Cell 161, 714–723. https://doi.org/10.1016/j.cell.2015.04.014

Sorrells, T.R., Johnson, A.N., Howard, C.J., Britton, C.S., Fowler, K.R., Feigerle, J.T., Weil, P.A., Johnson, A.D., 2018. Intrinsic cooperativity potentiates parallel cis-regulatory evolution. eLife 7, e37563. https://doi.org/10.7554/eLife.37563

Sprague, G.F., 1991. Assay of yeast mating reaction, in: Methods in Enzymology. Elsevier, pp. 77–93. https://doi.org/10.1016/0076-6879(91)94008-Z

Starr, T.N., Flynn, J.M., Mishra, P., Bolon, D.N.A., Thornton, J.W., 2018. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. Proc. Natl. Acad. Sci. U.S.A. 115, 4453–4458. https://doi.org/10.1073/pnas.1718133115

Tan, S., Richmond, T.J., 1998. Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. Nature 391, 660–666. https://doi.org/10.1038/35563

Tompa, P., Fuxreiter, M., 2008. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends in Biochemical Sciences 33, 2–8. https://doi.org/10.1016/j.tibs.2007.10.003

Tsong, A.E., Tuch, B.B., Li, H., Johnson, A.D., 2006. Evolution of alternative transcriptional circuits with identical logic. Nature 443, 415–420. https://doi.org/10.1038/nature05099

Tuch, B.B., Li, H., Johnson, A.D., 2008. Evolution of Eukaryotic Transcription Circuits. Science 319, 1797–1799. https://doi.org/10.1126/science.1152398

Tuttle, L.M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., Klevit, R.E., 2018. Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. Cell Reports 22, 3251–3264. https://doi.org/10.1016/j.celrep.2018.02.097

Vershon, A.K., Johnson, A.D., 1993. A short, disordered protein region mediates interactions between the homeodomain of the yeast α2 protein and the MCM1 protein. Cell 72, 105–112. https://doi.org/10.1016/0092-8674(93)90054-T

Wang, W., Cherry, J.M., Nochomovitz, Y., Jolly, E., Botstein, D., Li, H., 2005. Inference of combinatorial regulation in yeast transcriptional networks: A case study of sporulation. Proc. Natl. Acad. Sci. U.S.A. 102, 1998–2003. https://doi.org/10.1073/pnas.0405537102

Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E., Hahn, S., 2014. A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. PNAS 111, E3506–E3513. https://doi.org/10.1073/pnas.1412088111

Wray, G.A., 2003. The Evolution of Transcriptional Regulation in Eukaryotes. Molecular Biology and Evolution 20, 1377–1419. https://doi.org/10.1093/molbev/msg140

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Kyle Fowler
C430BD63B74B45D...
Author Signature

5/31/2022
Date