2019

Knob-socket Investigation of Stability and Specificity in Alpha-helical Secondary and Quaternary Packing Structure

Taylor Renee Rabara
University of the Pacific, taylor.rabara@gmail.com

KNOB-SOCKET INVESTIGATION OF STABILITY AND SPECIFICITY IN ALPHA-HELICAL SECONDARY AND QUATERNARY PACKING STRUCTURE

By

Taylor R. Rabara

A Dissertation Submitted to the

Graduate School

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Thomas J. Long School of Pharmacy and Health Sciences
Pharmaceutical and Chemical Sciences

University of the Pacifc
Stockton, CA

2019

KNOB-SOCKET INVESTIGATION OF STABILITY AND SPECIFICITY IN ALPHA-HELICAL SECONDARY AND QUATERNARY PACKING STRUCTURE

By

Taylor R. Rabara

APPROVED BY:

Dissertation Advisor: Jerry W. Tsai, Ph.D.

Committee Member: C. Michael McCallum, Ph.D.

Committee Member: Joseph S. Harrison, Ph.D.

Committee Member: Craig A. Vierra, Ph.D

Department Chair: Jerry W. Tsai, Ph.D.

KNOB-SOCKET INVESTIGATION OF STABILITY AND SPECIFICITY IN ALPHA-HELICAL SECONDARY AND QUATERNARY PACKING STRUCTURE

By

Taylor R. Rabara

DEDICATION

I dedicate this work to my beautiful daughter, Bailey Kay. You are the best part of me, and you have motivated me to be the best person I can be on a daily basis. Everything that I do is for you, my little bug.

And to my cousins, Kayla and Patrick. I miss you both every second of every day.

in time of daffodils(who know
the goal of living is to grow)
forgetting why,remember how

in time of lilacs who proclaim
the aim of waking is to dream,
remember so(forgetting seem)

in time of roses(who amaze
our now and here with paradise)
forgetting if,remember yes

in time of all sweet things beyond
whatever mind may comprehend,
remember seek(forgetting find)

and in a mystery to be
(when time from time shall set us free)
forgetting me,remember me

e.e. cummings

ACKNOWLEDGEMENTS

Jerry- Thank you for your endless support and for giving me the opportunity to pursue this degree in your lab. You believed in me when I couldn't believe in myself, and I can never thank you enough for pushing me to be the best researcher, mentor, and version of myself that you always knew I could be. You created such an amazing opportunity for me, and I will be grateful for that for the rest of my life. I'm not sure how I ended up with the best advisor/mentor, but I feel extremely, extremely lucky. Thank you for everything.

Henry- Thank you for patience and your willingness to work with me through the data analysis and mapping for these projects. I am lucky to have had you there to help!

My committee members, Dr. McCallum and Dr. Harrison- Your support through my qualifying exams as well as the writing process has been greatly appreciated. It has been a pleasure getting to know you through my years of graduate school and working with you both through this process. Thank you for your continued support and feedback!

Dr. V- I feel extremely lucky to have gotten to continue working with you and bouncing ideas off of each other as I worked through my Ph.D. Thank you for support and continuing to be one of the greatest mentors that I could have asked for.

My lab mates: Shivarni & Melina- Thank you both for working alongside me through my highs and lows, and supporting me through this program. I wouldn't have made it without you guys.

My undergraduate researchers: Danielle, Josh, Huy, Nickraj, Roos, Cynthia, Aaron, Kara, Sona, Sarah, Isabelle, Caitlin, Chris, Michael- Having the chance to work with and teach you guys over the years and getting to see you reach your goals has been a truly amazing opportunity,

and I'm so excited to see where life takes all of you! My graduate career would not have been the same without each and every one of you.

My family and friends- This journey has been anything but easy, but your love and support helped to get me through it. Mom, your strength and perseverance through your diagnosis and treatments has motivated me to stay strong myself and to realize that together we can make it through anything. I love you.

KNOB-SOCKET INVESTIGATION OF STABILITY AND SPECIFICITY IN ALPHA-HELICAL SECONDARY AND QUATERNARY PACKING STRUCTURE

Abstract

By Taylor R. Rabara

University of the Pacific
2019

The novel knob-socket (KS) model provides a construct to interpret and analyze the direct contributions of amino acid residues to the stability in α-helical protein structures. Based on residue preferences derived from a set of protein structures, the KS construct characterizes intra- and inter-helical packing into regular patterns of simple motifs. The KS model was used in the de novo design of an α-helical homodimer, KSα1.1. Using site-directed mutagenesis, KSα1.1 point mutants were designed to selectively increase and decrease stability by relating KS propensities with changes to α-helical structure. This study suggests that the sockets from the KS Model can be used as a measure of α-helical structure and stability.

The KS model was also used to investigate coiled-coil specificity in bZIP proteins. Identifying and characterizing the interactions that determine the dimerization specificity between bZIP proteins is a crucial factor in better understanding disease formation and proliferation, as well as developing drugs or therapeutics to combat these diseases. Knob-Socket mapping methods identified Asn residues at *a* positions within the helices, and were determined to be crucial factors in coiled-coil specificity. Site-directed mutagenesis was conducted to investigate the role of the Asn residues, as well as the role played by the neighboring residues at the g and b positions. The results indicate that the Asn at the *a* position defines coiled-coil

specificity, and that the Knob-Socket model can be used to determine bZIP protein quaternary interactions.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1: INTRODUCTION

Over the last several decades, scientists have looked to solve the protein-folding problem as an in-road to protein design [3, 4]. Classically, the protein-folding problem investigated the path of folding [6], but the focus of the protein-folding problem involves being able to predict the native protein structure from a given sequence [7, 8]. Although protein primary and secondary structure is well defined, the way in which the individual amino acid residues interact to form higher ordered structures remains relatively unknown. This lack of understanding of amino acid residues interactions also extends to an inability to accurately predict changes in structure or stability based on single mutations in a protein sequence [9-12]. To make progress in this fundamental area of protein design, the underlying physical principles between amino acid residues in a protein needs to be better characterized. The properties that define residue interactions are the hydrophobic force and a continuum of electrostatic interactions: long range charge attractions between positively and negative charged residues to short range polar hydrogen bonding and charge attraction/repulsions of van der Waals forces [13-15]. However, with a wealth of sequence information from genomic initiatives, the major advancement in the field has been the development of computational methods to predict and model protein folding based primarily on statistical analyses of the large sequence and structure databases [16-18]. By utilizing massive computing resources, successful design, synthesis and testing of mini-proteins has been reported [19, 20]. Following in the same approach, a deep dive into the residue packing interactions of the protein structures deposited in the Protein Data Bank [21] discovered a simple motif that describes the packing interactions between residues called the Knob-Socket motif [1, 2, 22, 23]. The goal of the work described in this thesis is to experimentally validate the ability of the Knob-Socket model to accurately analyze, predict, and design protein structure. The

experimental work can be divided into 2 main studies; investigating α-helix stability as well as characterizing the determinants of specificity in α-helical coiled-coils.

Protein Structure

The study of protein structure stretches back almost 70 years with an initial recognition that protein structure can be generally divided into 4 general levels [24, 25]. From the backbone to the structure and chemistry of each of the 20 coded amino acids as well as post-translational modifications, the primary structure of proteins is very well understood. At this level of structure, a single protein polypeptide is composed of a linear chain of amino acids linked together through peptide bonds. This peptide linkage between an amino group of one residue and a carboxyl group of another residue occurs through a dehydration reaction. Biologically, these bonds are formed during the process of translation from a mRNA transcript, which produces an expressed protein product from the genetic code. The peptide bond between each amino acid residue was recognized as possessing a unique planar structure [26, 27] which constrains the backbone and provides a structural basis for the next level of protein structure. Because the planarity locks the peptide bond's ω torsion angle at 0° or 180°, each residue possesses 2 torsion angles ϕ and ψ that are free to rotate, and the set of residue ϕ and ψ torsion angles can be plotted to provide an indication of protein backbone conformation formally named a Ramachandran plot [28]. Early on, the hydrophobic effect was recognized as the dominant force driving a protein to a folded state [15, 29-32]. In addition, the order of amino acids and therefore, the chemical interplay between the sidechain residues was shown to determine a polypeptide to a single structure [33, 34]. These interactions determine the remaining 3 levels of protein structure; however, the rules governing higher order protein structure still remain to be

defined, even though the forces involved in protein folding have been well known for a number of years [15].

Because of this lack of knowledge, the higher levels of proteins structure are less well characterized, and predictions at these levels of structure rely heavily on statistical analyses of sequence and structural databases. The hydrogen bonding pattern of the main chain polar groups making up the peptide bond define secondary structure of proteins. Well before the first protein structures were solved [35-37], the different types of secondary structure had already been modeled based on hydrogen bond configurations [38-40]. Essentially, protein secondary structure can be divided into 2 general classes or regular and irregular secondary structure. The regular secondary structure of α-helices and β-sheets exhibit patterns of hydrogen bonding as well as defined backbone dihedral angles. Right-handed α-helices are the most well characterized secondary structures, and are defined by the repeated hydrogen bonding pattern between the backbone carbonyl of an amino acid, *i*, with the amino group of the *i*+4 residue [41]. Extended backbone strands make hydrogen bonds with each other to form parallel and anti-parallel β-sheets[42]. Additionally, irregular secondary structure is called random coil, which includes everything not classified as regular. Often, random coils are split into the subsets of turns which are short (3 to 4 residue) hydrogen bonded segments between regular secondary structure, and loops which are the longer irregular coils. While a great deal of work has been invested in understanding secondary structure [43-45], the best methods for predicting protein secondary structure remain statistical methods based on deep sequence alignments and matching [46, 47]. One of the reasons behind this difficulty is the need for a better understanding of how individual residues contribute to protein secondary structure.

However, protein secondary structure is intimately entangled with the next level of protein structure called tertiary structure, where these secondary structure segments pack to form higher ordered structures. While there are several intramolecular forces at play [15], tertiary structure is primarily driven by the hydrophobic effect [48, 49], where non-polar groups bury in the core to allow an increase in water entropy. As with secondary structure, protein tertiary structure has been extensively studied and characterized [50-54], however, the best methods have been computational knowledge-based methods for predicting and understanding protein tertiary structure [16, 55, 56]. At the final level of protein structure, protein-protein interaction between two or more folded proteins is described as protein quaternary structure [15, 57-59]. These interactions are composed of the same non-covalent forces used to hold the individual peptide chains together in the lower ordered structures such as the hydrophobic effect, but it has been found that specificity is mediated by certain polar side-chains [60-62]. Prediction and modeling at this level has required the help of experimental data [63] as well as computational modelling [64, 65].

To make progress in the protein structure field, an underlying need in understanding the higher levels of protein structure is to relate protein sequence to structural arrangements in space. In many ways, this requirement is obvious yet has remained the crux of the problem in determining structure from a protein sequence. Current successful main-chain centric methods of structure prediction build models by segmenting the protein sequence into a library of statistically likely backbone fragments and guiding large-scale, stochastic sampling of fragment combinations with knowledge-based scores of common protein features [66, 67]. The packing of side-chain residues is usually the last step in the prediction process, although it is the chemistry between the side-chain residues that determine a protein's fold [10]. While correlations between

pairs of protein residues have been found, finding consistent packing patterns has been challenging for pair-wise approaches. The main problem in studying protein side-chain packing has been finding a way to identify order from interactions that largely result from non-specific van der Waals forces between side chains. One of the goals of this thesis work is to provide better experimental characterization of protein packing for protein prediction and design.

Stability Studies in Protein Structure: α-Helices

For years, methods of studying and predicting secondary structure stability have been highly sought after. The structure of α-helical peptides was first described by Pauling et al, which began a search for the key to protein stability [41]. The stability of α-helices has long been an important subject in protein folding problem since α-helix provide simplest regular secondary structural unit of protein. Helical peptide stability has since been studied extensively. Initially, helical formation was modeled using polymer theory to assign helix-coil transition parameters by Zimm-Bragg [68] and Lifson-Roig [69]. These helix-coil transition theories were developed based on experimental data to describe α-helix formation in homopolypeptide chains, and then parameterized to the different amino acids residues. The approach has been to analyze and correlate the energetics involved in the formation of α-helices with their respective stabilities and free energy measurements, and also account for side chain interactions [70-72]. The original helix-coil theories ignore certain aspects of helices such as charged residue-helix macro-dipole interactions [73, 74] and the side chain interactions [75]. Modifications to original theories have been made to address the terminal capping phenomena as well as these shortfalls [71, 76-86]. Initially, the Cα atoms of individual residues are classified as helical or non-helical conformation and the free energies of helical conformations relative to non-helical conformations are described by nucleation and propagation parameters. The free energies of helix formation were calculated

by statistically weighing the residue conformation distributions in the peptide. The side chain charge–helix dipole interaction, interactions between side chains of residues, and the N-terminal capping effects of amino acids are included in modified versions of helix-coil transition theories. The helix propensities of amino acid helix propensities are experimentally determined in several different conditions [77, 79, 87-92]. However, the values are not consistent, depending on the reference states, experimental conditions, and model peptides.

Computational and experimental methods are commonly used together, to select for the most promising mutants and therefore decreasing library size. A new approach includes deep sequencing technologies combined with experimental selection methods. This methodology has been used to characterize protein fitness landscapes as well as to determine binding affinities [93-96]. Programs have since been developed and written to predict the ΔΔG of protein folding or for mutant variants of a single protein [97]. The better-known programs include Rosetta, FoldX, CC/PBSA, and Eris [98-101]. A study was done by Potapov et al to evaluate the precision of these programs in predicting changes in protein stability for a library of over 2000 mutants [102]. This study showed that while the correlations for the predicted values and experimentally determined values were positive, the values themselves had large margins of error. Successful approaches to stabilizing proteins is directed evolution, where random mutations are introduced into a particular enzyme sequence, and the functionality of the resulting enzymes is used as a selection method [103]. The stability of proteins can be measured using methods including DSC, pulse-chase, CD spectroscopy and fluorescence based assays [104-107]. For α-helices, stable single α-helices[108-110] were identified in naturally occurring proteins, which were used to be misidentified as coiled-coil [111]. These isolated α-helices are very rich in Arg (R), Lys (K), and Glu (E) and are stabilized by multiple salt bridges. Simple heptad

repeats of AEEEXXX (were X is K or R) motif was used to design stable long single α-helices [109]. It was found that K:R ratio is important in stability of single α-helix, and peptides with high arginine were aggregated. Arginine is more flexible in forming salt bridge than lysine to increase helical stability, and it can promote the tertiary interaction to bundle up the helices. However, all these success in creating de novo helical peptides, our understanding of natural α-helices are limited. Naturally occurring helices are not simple heptad repeats with mostly charged amino acid residues. To understand single α-helical stability of natural protein and to design peptides like natural proteins, we need a new approach deciphering residue interactions within a helix as well as between helices, which allow us to design a helix that can stay as a stable single α-helix.

Protein Design

Most current methods of protein design require the use of structure-prediction algorithms such as Rosetta [112]. Advancements in structure prediction programming have been made due to the growth of the PDB, and the development of higher quality sequence alignment search tools. One subset of computer-based structure prediction is template-based modeling which assumes similar sequences found in the PDB will assume similar native structures. Another newer subset is called fragment-based modeling, where the protein sequence is not found in the PDB. For fragment-assembly, the protein sequence is broken into smaller fragments, and a search is done on the shorter sequences. For either method, the structures and sequences are assessed using a scoring algorithm to determine the predicted structure with the lowest energy function [113-115]. Another area of research is the inverse protein-folding problem, which involves the selection of a specific protein structure, and predicting a sequence that will fold into the desired shape [116, 117]. One way to approach this problem is through the use of

combinatorial libraries [118-120]. However, because randomly generated sequences rarely become well-folded protein structures, rational design methods must be used to narrow the sequence space of the libraries. A combinatorial library has been constructed through the use of a binary code of polar and non-polar amino acids that favor particular secondary structures, while exposing the hydrophilic residues and burying the hydrophobic residues. This design technique relies on the explicit placement of polar and non-polar residues but can vary the identities of these residues to obtain a desired secondary structure. Each of these methods are largely knowledge based, and each requires extensive sampling and energy scoring for a large number of favorable conformations [114, 115]. Additionally, the field of structural biology continues to rapidly expand, increasing the need for a structure prediction method of single and more complex protein structures.

Modeling of Protein Packing

As explained above, the major difficulty in developing a useful characterization of protein tertiary structure has been in discovering an effective construct that produces order from non-specificity of packing interactions. The simplest approach has been to investigate pair-wise contacts [121-129], which has shown success in finding amino acid correlations. However, a pair-wise treatment of residue interactions is too simplistic and cannot capture the 3-dimensional complexity of packing (Fig. 3A) [130]. More elaborate analyses of protein packing, including our own, consider multi-body arrangements of residues [130-137]. While these studies have generally found side-chain interactions to be broadly regular and tetrahedral, none so far has been able to develop a coherent description of protein packing. Another approach employs graph theory to organize protein interactions in hopes of identifying some common patterns across fold types. As the graphs are quite fold specific, this strategy has difficulty in finding

common motifs across fold families [138-141] and is therefore more suited to distinguishing between protein families [142, 143]. As a new perspective on protein packing, this proposal will demonstrate that the knob-socket motif addresses the multi-body residue interactions and simplifies packing to uncomplicated pattern representations. Even in the well-studied system of side-chain interactions between α-helices [144-147], this thesis extends the classic analyses of α-helical packing: Crick's knobs-into-holes [148] and Chothia et al's ridges-into-grooves [149]. Similar to the analysis of tertiary structure discussed above, recent investigations of α-helix packing have characterized amino acid propensities [121, 122, 137, 142] and energetics [150, 151], but have not significantly advanced the insight into α-helical packing beyond the initial knob-into-hole and ridge-into-groove models. The knobs-into-holes translates to primary structure as the well-known heptad repeat [152], but this pattern is limited to helix coiled-coils [153, 154]. To describe other types of helical packing, an elegant implementation of knobs-into-holes has been developed recently that computationally assesses helical packing [155, 156]. As an alternative, the helical lattice superposition model views packing as the interlacing of side-chain Cα positions [157]. In conjunction with the helical wheel [158], these approaches have been used to dissect helix-helix packing interfaces [159-163], yet only a few examples of designed α-helices have been successful. From the pioneering work on redesigning α-helical packing [164-166] and modulating helix oligomerization state [167, 168] to more recent design of α-helix oligomers [169-173], the designed proteins in these studies have been largely built from known scaffolds and sequences. Even with such advances in design, the understanding of α-helix packing remains primarily the residue repeats indicated on a helical wheel by the canonical knob-into-hole coiled-coil or ridge-into-groove packing. The simplification of packing

by the knob-socket motif into discrete patterns presents an entirely new approach to interpret all α-helical packing and design new oligomers or even unique folds.

Quaternary Specificity of α-helical Coiled-Coils

Early on, the packing of 2 α-helices forming a coiled was recognized as a simple representative of protein packing [148]. Currently, α-helical coiled-coils have been calculated to make up 5 to 50% of the genome [162] with a wide diversity of coiled-coil based folds [174]. At the quaternary level, these coiled-coil structures have been recognized as the dimerization component in many transcription factors and in particular basic leucine zipper (bZIP) proteins [175], where the basic portion of the α-helix binds DNA and the leucine zipper forms the coiled-coil [176]. The first solved structures of coiled coils identified the pseudo-heptad repeat of *abcdefg* [177, 178], where the *a* and *d* positions are hydrophobic (*h*) residues and the remaining are polar (*p*) residues to form a heptad repeat pattern of *hpphpp* [179]. The heptad repeat sequences are commonly overlaid on a helical wheel to represent coiled-coil interactions [158, 168] and have been successfully applied in modifying and designing coiled coil sequences [180-182]. While leucine (Leu) residues have a high propensity at both the *a* and *d* positions of the heptad repeat [183], Asn residues were found only at the *a* positions, but the initial Asn to Leu mutational studies changed oligomerization state from a dimer to higher numbered bundles [167, 184, 185]. Because the coiled-coils still interacted, this change in oligomerization state was called structural specificity [186] in contrast to changing binding or sequence specificity. Investigations to stability of different amino acids at both the *a* and *d* positions generally identified that hydrophobic residues increased stability, while polar and charge groups decreased stability [187, 188]. Because the bZIP transcription factors exhibited distinct specificities for homodimerization and heterodimerization [176], and newer extended knob-into-holes model of

packing [155] that recognized the importance of the neighboring positions of *g* with *a* and *d* with *e* in packing brought about further residue characterizations of these positions' contributions to specificity and stability [189-191]. As shown in a comprehensive experimentally determined interaction map of 492 bZIP proteins [192], the coiled regions demonstrated distinct specificities between certain groups, where certain sequences homodimerized and heterodimerized and others only homodimerized. Structurally, the Asn residues were shown to form intermolecular hydrogen bonding between sidechain Asn residues from different chains [193-195]. A number of studies investigated the nature of Asn specificity to structural and sequence specificity and identified energetic [187, 196] and genetic [197] evidence for Asn specificity in a coiled coil. Yet, the exact determinants of structural versus sequence specificity remain elusive [198]. Although certain rules have been found for structural specificity [199], the successful methods to design specific coiled coil structures still rely heavily upon application of statistical analysis of structural and sequence data [200-203].

The Knob-Socket Model

Through many years of research investigating the packing in protein structure, my group has developed the knob-socket construct that characterizes specificity and simplifies the complexity of residues interactions. This model is comprehensively explained for protein 3° structure packing in three published manuscripts [1, 2, 5], with a fourth manuscript under revision [204]. The knob-socket motif was developed by building basic packing constructs from pairwise residue contacts. Voronoi polyhedral [130, 205, 206] were used to more accurately define residue contacts. From these pairwise contacts, graph theory cliques identified groups of residues that all interacted with each other [132], and contact order was used to classify residue relationships [207]. The list of packing constructs produced from this procedure was pared down

based on redundancy with each other. In the end, a single tetrahedral packing clique between two elements of 2° structure was able to account for the residue packing within proteins: the knob-socket construct (Fig. 1). As the name implies, the packing construct consists of 2 parts: a knob (B) residue from one 2° structure element packing into a surface formed by a three-residue socket from another 2° structure element (Figs. 1A-D). The challenge has been to identify the socket configuration in the different types of 2° structure, particularly the repetitive regular α-helices and β-sheets compared to the irregular coil and turn. For the regular α-helical [1] and β-sheet [2] sockets, the two neighboring residues X and Y interact with another hydrogen-bonded residue, H, to form the XY:H socket (Figs. 1E&F, respectively). For coil and turn sockets [5], the three X, Y and Z residues are consecutive in sequence to create the XYZ socket (Figs. 1 G&H, respectively).

As a basic unit of packing, the knob-socket motif reduces the degrees of freedom by defining the specificity in protein 3° structure as the pairwise interaction of individual knob residues into three residue sockets. From a structural perspective, 2° structure elements present patterns of sockets that pack single residue knobs from another 2° structure element (Fig. 2A) [204]. The socket patterns create a surface topology that indicates the preferences of knob residue packing, and moreover, a means to better understand and characterize packing of residues in protein structure. As an example, the knob-socket model characterizes and produces clear and intuitive maps of the canonical packing between elements of secondary structure (Fig. 2B-C). Even though canonical packing between two α-helices has been described in detail [148, 149, 153, 154], the knob-socket's packing surface topology maps of α-helical packing [1] provides a more precise depiction of the residues involved in the packing (Fig. 2B) beyond the

standard repetitive sequence patterns [152] or a helical wheel. In contrast, canonical packing involving β-sheets

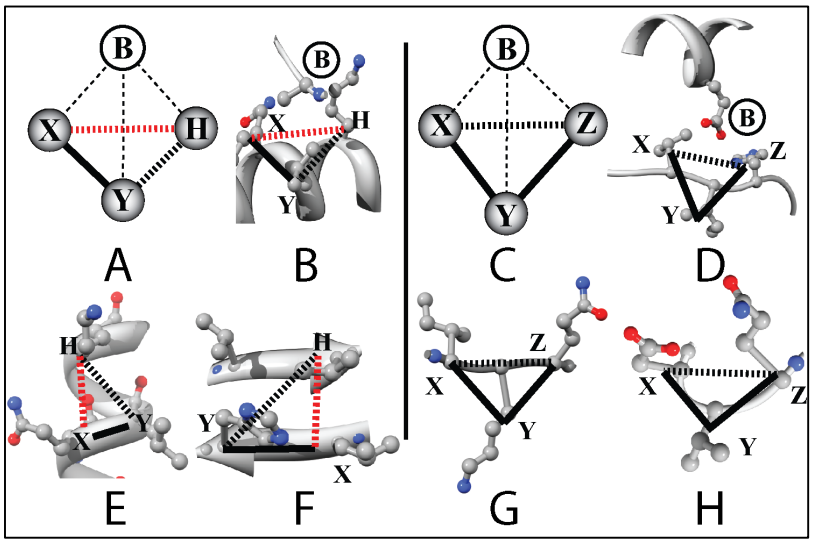


Figure 1. The Knob-Socket Motif. On the left are XY:H sockets from regular α-helix and β-sheets 2° structure, while on the right are XYZ sockets from irregular turn and coil 2° structure. Note: as these are all packing cliques, all residues' side-chains are packed with each other. **A)** The regular knob-socket tetrahedral packing clique is shown in abstract by spheres for regular 2° structure. The knob B residue from one 2° element packs via vdW interactions (thin broken lines) into the XY:H socket formed by 3 residues from the same regular 2° element. For the 3 residues in the socket, the solid line between X and Y indicates covalently bonded neighboring residues, while residue H is more distant in sequence of the same 2° structure, which is denoted by the ":". The broken red line indicates a main-chain hydrogen bond between X and H residues and the broken black line indicates only vdW packing between Y and H. **B)** An example of knob-socket packing between a coil knob with a regular XY:H helical socket.[1] **C)** The knob-socket tetrahedral packing clique is shown in abstract by spheres for irregular 2° structure.[2] Again, the knob B residues from one 2° element packs via vdW (thin broken lines) into a XYZ consecutive residue socket from an irregular 2° structure element. The solid lines indicate covalent interactions between X and Y as well as between Y and Z, while the broken line indicates only a vdW packing between X and Z. **E)** The XY:H socket of an α-helix. X and Y are ±1 from each other, while X and H are ±4.[1] **F)** The XY:H of a β-sheet. X and Y are ±2 from each other due to the alternating residue arrangement on β-strands.[2] X and H are from different β-strands, so their sequence separation is variable. **G)** & **H)** Examples of XYZ socket from irregular 2° structure of coil and turn, respectively.[5]

does not exhibit strong sequence patterns, but knob-socket packing surface topology maps exhibit the intricacies of canonical packing in β-sheets structure [5]. Most remarkably, the

packing surface topology maps reveal that packing specificity not only involves sockets filled with a knob residue, but just as important are the non-packing sockets that are free of interactions with a knob residue. Therefore, as shown in Fig. 2, the pattern of the filled sockets in the context of the free sockets defines how 2° structure will form, and further how the proteins will pack against each other.

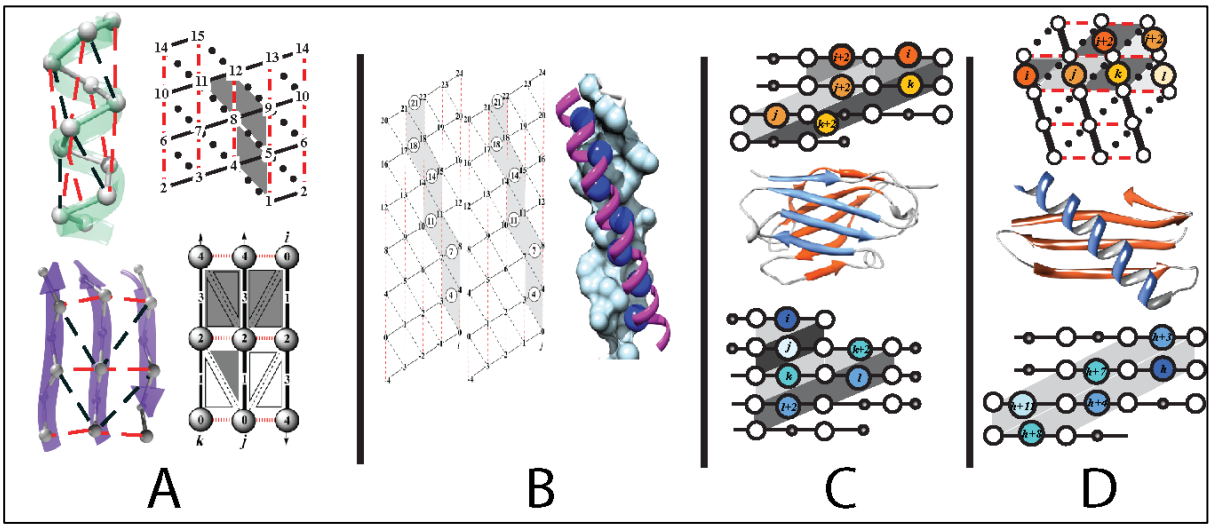


Figure 2. Packing Surface Topology Maps & Canonical Packing. In each of these maps, the grey triangles indicate filled sockets, while the white or open indicate free sockets. **A)** The repetitive nature of regular secondary structure produces a ordered lattice for an α-helix (top in green)[1] and β-sheet (bottom in purple).[2] **B)** Canonical packing of 2 parallel α-helices.[1] Left handed parallel coiled-coil pattern of helix packing with crossing angle of 25 degrees. Knobs from the other α-helix are shown by circles. **C)** Canonical β-sheet packing at parallel -30 degree & anti-parallel 150 degree crossing angles.[5] **D)** Canonical α-helix/β-sheet packing at -35 degree parallel and 140 degree anti-parallel.[5]

The existence of the filled and free sockets strongly indicates that knob-socket model can directly relate packing structure to specific amino acid sequences. For the amino acid composition of each socket, the frequency was computed for whether that socket was filled with

a knob (Fig. 3A, top) or that socket was free without a knob (Fig. 3A, bottom). While this was done for the four 2° structure socket types shown in Figs. 2E-H [1, 2, 5], Fig. 4B shows the results for the top twenty from the analysis of 8,000 α-helical sockets. Clearly, amino acid composition can dictate the state of a socket being filled or free. The top three filled α-helix XY:H sockets AL:V, AI:L, and AL:L have very high frequencies of being packed with a knob residue and are seldom found as free sockets (Fig. 3B, top histogram). The trend for free sockets is even stronger, where the top three free α-helix XY:H sockets AE:K, EE:K and EK:R are predominantly found without a knob and hardly ever pack with a knob (Fig. 4B, middle histogram). These compositions are not surprising as they contain the well-known α-helical stabilizing pairwise *i* to *i*+4 salt-bridge between X and H residues. Interestingly, there are some socket amino acid compositions like AE:L and AK:L that can are strongly both free and filled. While the top two histograms of filled and free sockets both favor α-helix conformation in the XY:H socket, the bottom histogram of Fig. 4B shows the amino acid compositions of sockets that are not prevalent in α-helices. In terms of negative design, these are sequences that disfavor α-helix formation. Taking the analysis a step further, amino acid preferences are found for both knob residues and the amino acid composition of their respective pockets. Essentially, this model provides a tetrahedral amino acid code that relates protein sequence with knob-socket structural configurations.

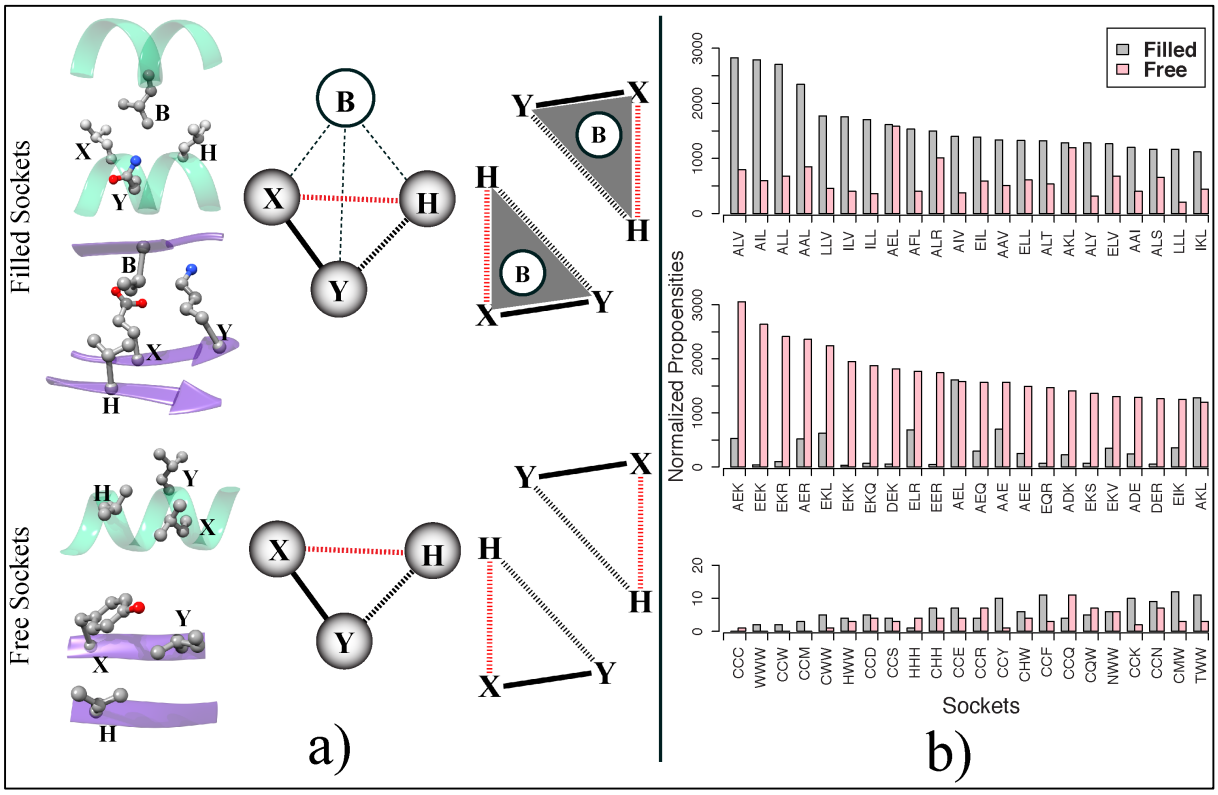


Figure 3. An Amino Acid Code for Protein Packing. a) The filled (top) and free (bottom) types of regular 2° structure XY:H sockets are shown by examples from protein structure (left), reduced ball representation (middle), and 2D mapping triangles (right). Filled sockets indicate 3° packing interactions between two elements of 2° structure: a knob B residue from one and a 3 residue XY:H socket from another. In 2D, the triangles are greyed to indicate packing. Free sockets disfavor packing with knob residues and indicate only 2° structure packing between the XY:H socket residues. In 2D, these are left white or unfilled to show no packing. **b)** Socket composition relates sequence preferences to structural arrangement. The propensity (frequency of socket composition normalized against amino acid prevalence) is shown for filled, free, and disfavored sockets in α-helices.[1] The top 2 histograms reveal that certain residue combinations favor filled sockets, while others prefer free sockets. The bottom histogram shows low-propensity (<20 counts out of ~800,000) socket compositions not containing Gly or Pro. These are considered disfavored or non-socket combinations.

Experimental Validation of the Knob-Socket Model

Despite recent advancements, understanding the nature of protein folding problem is far from satisfaction. Since protein stability depends on the protein folded structure, understanding protein folding problem is highly dependent on the understanding protein stability. As explained above, the Knob-Socket model provides a construct to rationally interrogate the packing contributions of amino acid residues. To verify the effectiveness of a Knob-Socket analysis, two experimental investigations have been pursued in modeling protein secondary structure stability and predicting protein quaternary structure specificity. The first study involves characterizing a set of single and double mutants of the designed α-helical peptide KSα1.1 [1] and correlates experimentally measured helicities and stabilities against Knob-Socket changes to propensities based on the helical lattice shown in Fig. 2A. Although not a 1 to 1 mapping, the correlation was demonstrated that the Knob-Socket analysis can indicate the direction of stability of a mutation in a helix primarily based on contacting residues. In the second study, the mapping of an α-helix coiled-coil protein quaternary interface by the Knob-Socket model (Fig. 2B) provide unique insight into the specificity of coiled-coils. Based on this analysis, a number of mutations were performed on 3 coiled-coil proteins with different specificities. The mutations were chosen to change the specificities of the coiled coils so that they would no longer dimerize with the wild-type sequence, but specifically change to heterodimerize with a different specificity. The predictions showed that the Knob-Socket analysis identified how α-helix coiled-coil determine specificity. The successful results of both experimental studies demonstrate that the Knob-Socket model provides insight into the fundamental packing determinants at both the levels of secondary and quaternary protein structure.

CHAPTER 2: MATERIALS AND METHODS

Stability Relationship of Knob-Socket Propensities Measured in the KSα1.1 α-Helix

Wild-Type KSα1.1 Construct

The KSα1.1 wild-type sequence was cloned previously into a pET-28a(+) plasmid containing an N-Terminal 6X Histidine and SUMO tag. The wild type construct was confirmed through sequencing (Sequetech), and the nucleotide and amino acid sequences are shown below in Table 1, and plasmid map shown in Figure 4.

Table 1
Nucleotide and Amino Acid Sequences of Wild-Type KSα1.1. The 6X Histidine tag is shown highlighted in magenta, the SUMO tag in black bolded text, and wild-type KSα1.1 is shown bolded in teal.

KSα1.1 Wild-Type Construct
Nucleotide Sequence (5' to 3')
CATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATG**TCG GACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCAC ATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCT TTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTC TTGTACGACGGTATTAGAATCCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAAC GATATTATTGAGGCTCACAGAGAACAGATTGGTGGA**TGGGGTGAACGCCAGGCGAAAGCGGTG GCGGATGCGCTGACCGCGCTGGAAAGCGCGATGGCGCGCATTGCGAAAGAACTG**TGA
Amino Acid Sequence
HHHHHHSSGLVPRGSHMA**SMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTP LRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG**WGERQAKAV ADALTALESAMARIAKEL

Site-Directed Mutagenesis

Mutagenesis was performed on the pET-28a(+)_KSα1.1 construct to determine the changes in secondary structure and stability with the incorporation of different amino acid point mutations. Specific amino acids were selected for site-directed

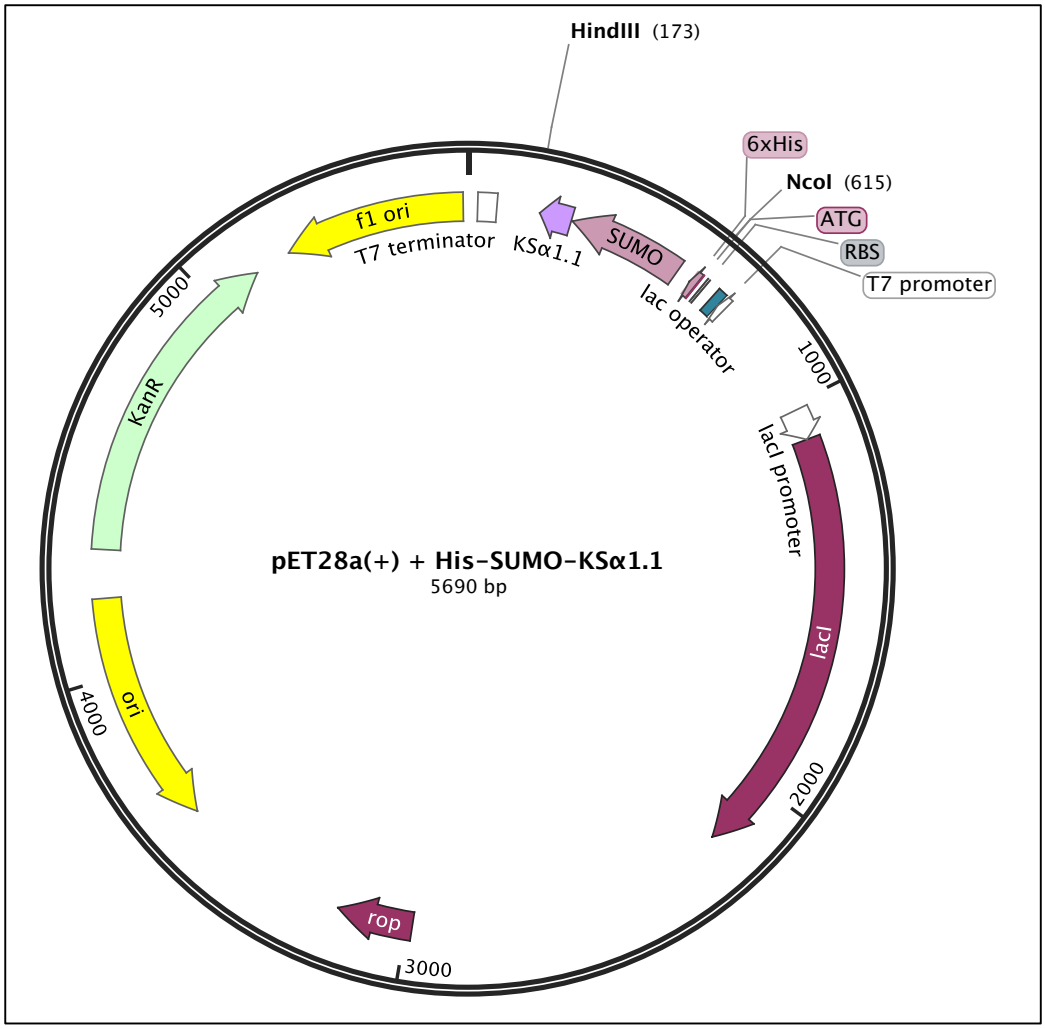


Figure 4. Plasmid map of pET28a(+) containing the 6X His-SUMO+KSα1.1 insert. Plasmid contains the 6XHis-SUMO-KSα1.1 insert. Insert is flanked by cloning sites at positions 615 (*Nco*I) and 173 (*Hind*III). Genes are preceded by a T7 promoter.

mutagenesis (SDM) based on the packing pattern of KSα1.1. Primers for point mutants were designed according to the QuikChange Lightening II Site-Directed Mutagenesis Kit (Agilent) primer design protocol and are shown below in Table 2. PCR reactions were run in a 50 µL total reaction volume containing 1X Reaction Buffer, 125 ng/primer, 10 ng template DNA, 0.5 mM

dNTP mix, 3 µL of QuikSolution, and 2.5 U of *Pfu*Ultra Polymerase. The PCR program was created with the following parameters: initial denaturation at 95°C for 50 seconds, denaturation at 95°C for 50 seconds, annealing at 54°C for 1 minute, extension at 68°C for 10 minutes and returning to the denaturation step for a total of 18 cycles. PCR reactions were then *Dpn*I digested for 1 hour at 37°C, and transformed into XL10-Gold® Ultracompetent cells that had previously been incubated with a proprietary BME mixture (Agilent) at 4°C for 30 minutes. The transformed

Table 2
Primer Sequences for Site-Directed Mutagenesis of Wild-Type KSα1.1. Primers were designed using the online Agilent primer design tool. Base pair mismatches for each primer set are bolded and highlighted in yellow.

Mutant	Primer Sequences
Q5L	**F:** 5′ CCGCTTTCGCC**A**GGCGTTCGCCC 3′
	R: 5′ GGGCGAACGCC**T**GGCGAAAGCGG 3′
A6E	**F:** 5′ GCCACCGCTTTC**T**CCTGGCGTTCGC 3′
	R: 5′ GCGAACGCCAGG**A**GAAAGCGGTGGC 3′
V9E	**F:** 5′ GCGCATCCGCC**T**CCGCTTTCGCC 3′
	R: 5′ GGCGAAAGCGG**A**GGCGGATGCGC 3′
A10R	**F:** 5′ CAGGCGAAAGCGGTG**CG**GGATGCGCTGACCGC 3′
	R: 5′ GCGGTCAGCGCATCC**CG**CACCGCTTTCGCCTG 3′
L13I	**F:** 5′ CTTTCCAGCGCGGT**TA**TCGCATCCGCCACCG 3′
	R: 5′ CGGTGGCGGATGCG**AT**AACCGCGCTGGAAAG 3′
T14V	**F:** 5′ GGCGGATGCGCTG**GT**CGCGCTGGAAAGC 3′
	R: 5′ GCTTTCCAGCGCGAC**CA**GCGCATCCGCC 3′
L16F	**F:** 5′ CCATCGCGCTTTC**G**A**A**CGCGGTCAGCGCA 3′
	R: 5′ TGCGCTGACCGCG**TT**C**G**AAAGCGCGATGG 3′
E17D	**F:** 5′ GCGCTGACCGCGCTGGA**T**AGCGCGATG 3′
	R: 5′ CATCGCGCT**A**TCCAGCGCGGTCAGCGC 3′
M20L	**F:** 5′ GCGCTGGAAAGCGCG**C**TGGCGCGCATTGCGA 3′
	R: 5′ TCGCAATGCGCGCCA**G**CGCGCTTTCCAGCGC 3′
A21E	**F:** 5′ CTTTCGCAATGCGC**T**CCATCGCGCTTTCC 3′
	R: 5′ GGAAAGCGCGATGG**A**GCGCATTGCGAAAG 3′

cells were recovered in SOC broth for 1 hour, and plated on LB agar plates containing 50 μg/mL kanamycin. After incubation at 37°C overnight, colonies were picked and purified via miniprep (Qiagen) and confirmed through sequencing (Sequetech).

Double mutant variants of KSα1.1 were also created for downstream structural and stability change analysis. Mutations were chosen in locations that were predicted to further increase the helical content and stability of the KSα1.1 single mutants. The two double mutants selected for analysis were T14V/M20L and T14V/A21E. SDM primers used to create these mutant constructs are shown in Table 3. SDM was conducted as described for the single mutants, and constructs were confirmed through sequencing (Sequetech).

Table 3
Primer Sequences for Site-Directed Mutagenesis of Double Mutants. Primers were designed using the online Agilent primer design tool. Base pair mismatches for each primer set are bolded and highlighted in yellow. The confirmed single mutant M20L and A21E plasmids were used as template for the corresponding SDM reactions.

Mutant	Primer Sequences
T14V/M20L	**F:** 5' GGCGGATGCGCTGGTCGCGCTGGAAAGC 3'
	R: 5' GCTTTCCAGCGCGACCAGCGCATCCGCC 3'
T14V/A21E	**F:** 5' GGCGGATGCGCTGGTCGCGCTGGAAAGC 3'
	R: 5' GCTTTCCAGCGCGACCAGCGCATCCGCC 3'

Large Scale Induction of Recombinant Proteins

Wild type and mutant pET28a(+)-KSα1.1 constructs were transformed individually into BL21(DE3)pLysS competent cells (Invitrogen) and grown overnight at 37°C in 250 mL pilot culture flasks of LB media containing 50 mg/mL kanamycin. The 250 mL saturated cultures

were combined with 1 L of fresh LB media containing 50 mg/mL kanamycin. The resulting 1.25 L culture was supplemented with 100 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final working concentration of 1 mM IPTG to induce protein expression. After incubation for 4 hours at 37°C with shaking, the protein inductions were pelleted in 250 mL centrifuge bottles. Pellets were resuspended and lysed in 10 mL of a 9:1 ratio of Native Lysis Buffer (NLB; 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and FastBreak Lysis Buffer (Promega) and further sonicated for at 15 second intervals for 2-4 minutes to ensure complete cell lysis. Crude lysates were clarified through centrifugation at 10,000xg for 10 minutes. Clarified lysates were collected for affinity chromatography.

Lysates were added as 20 mL fractions to a chromatography column (BioRad) along with 2 mL of Ni-NTA agarose resin (Gold Biotechnologies) and incubated for 1 hour on a rotating mixer. After incubation, the flow-through fraction was collected in a 50 mL conical tube, and the nickel resin was washed with 40 mL of Binding/Wash Buffer (20 mM NaH_2PO_4, 0.5 M NaCl, 20 mM imidazole, pH 8.0) in two 20 mL increments and was collected in four 10 mL fractions in 15 mL conical tubes. Proteins were then eluted using 15 mL of Elution Buffer (20 mM NaH_2PO_4, 0.5 M NaCl, 300 mM imidazole, pH 8.0) into three 5 mL fractions in 15 mL conical tubes. The fractions were analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis with Mini-PROTEAN® TGX gels (BioRad) to ensure the presence of proteins of correct size in the elution fractions. After confirmation, the elution fractions were combined and injected into a 15 mL 2,000 MWCO Slide-A-Lyzer (Thermo) dialysis cassette using a serological pipette. The cassette was incubated in 1 L of Dialysis Buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) at 4°C on a stir plate for 24 hours with buffer changes at 3- and 6-hour time points. After the first dialysis buffer change, purified Ubiquitin-like protease 1 (ULP-1) was

added to the dialysis cassette to remove the SUMO-His tag from the recombinant proteins. To prepare the ULP-1 samples, ULP-1 plasmids were transformed and expressed as described above. The pellets were resuspended and lysed in 20 mL of a 9:1 ratio of NLB and FastBreak Lysis Buffer and sonicated. Crude lysates were clarified through centrifugation at 10,000xg for 10 minutes. Clarified lysates were collected for purification via FPLC. Clarified lysates were purified using an ÄKTA Start chromatography system (GE Healthcare) loaded with a 1 mL HisTrap FF nickel column. The purification parameters were programed as follows: 20 mL of sample loaded flowed through column and collected as 10 mL of flow through, immobilized proteins were then washed with 20 mL of Binding/Wash Buffer and collected in four 5 mL fractions, and finally protein was eluted from column using 20 mL of Elution Buffer and collected in four 5 mL fractions. A Frac30 fraction collector was used to collect flow through, wash, and elution fractions. The elution fractions were concentrated using a 2,000 MWCO VivaSpin 15R concentrating tube (Sartorius). Aliquots of 1 mg/mL ULP-1 were stored at -80°C for future use.

After 24 hour incubation in dialysis buffer and SUMO-His tag cleavage, the protein sample was added to a clean chromatography column along with 1 mL of Ni-NTA agarose beads and incubated for 1 hour at 4°C. The flow-through, now containing a pure cleaved protein sample, was collected in a 15 mL conical tube. The resin was washed and SUMO-His tags were eluted as described previously. The fractions were analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis with Mini-PROTEAN® Tris/Tricine gels (BioRad) to ensure the presence of cleaved KSα1.1 proteins of correct size in the flow-through fraction.

Protein Visualization via SDS-PAGE Analysis

Collected fractions were prepared for TGX (Tris/Glycine) SDS-PAGE with the addition of 30 μL of Laemmli Sample Buffer (20:1 ratio of 2X Laemmli Sample Buffer to β-mercaptoethanol) to 30 μL of protein. After boiling for 5 minutes at 95°C, the samples were loaded into a 12% Mini-PROTEAN® TGX SDS-PAGE gel and electrophoresed at 200 V for 30 minutes. The gel was then removed from the plastic casing and placed into a staining container with 20-30 mL of Imperial™ Protein Stain (ThermoFisher) and stained and de-stained according to the manufacturer's protocol. De-stained gels were dried and cast using a DryEase™ Mini Cellophane casting system (ThermoFisher).

Collected fractions were prepared for Tris/Tricine SDS-PAGE with the addition of 30 μL of TTS Sample Buffer (20:1 ratio of 2X TTS Buffer to β-mercaptoethanol) to 30 μL of protein. After boiling for 5 minutes at 95°C, the samples were loaded into a 16.5% Mini-PROTEAN® Tris/Tricine SDS-PAGE gel and electrophoresed at 95 V, 35 mA for 90 minutes. The gel was then removed from the plastic casing and placed into a staining container with 20-30 mL of Imperial™ Protein Stain (ThermoFisher) and stained and de-stained according to the manufacturer's protocol. De-stained gels were dried and cast using a DryEase™ Mini Cellophane casting system (ThermoFisher).

Preparing Purified Samples for Circular Dichroism

Confirmed cleaved KSα1.1 samples were concentrated and desalted using 2,000 MWCO VivaSpin 15R concentrating tube (Sartorius). Desalting was done by adding a total of ~30 mL milliQ water to the protein sample between spins. Protein was concentrated after desalting to a final volume ranging from 0.5-2.0 mL. Concentrated samples were stored in 10 mM potassium phosphate (pH 7.0). A280 concentrations were measured for each concentrated sample using a

NanoDrop 2000c (Thermo). Samples were then diluted to 200 μM in 10 mM potassium phosphate. All samples were analyzed using a Jasco J-180 Spectrophotometer with a Jasco MPTC-490S light source with a 1 mm Quartz cuvette. Before analyzing the protein sample, the instrument was blanked using 200 μL of 10 mM potassium phosphate. The parameters for the full spectrum scan are shown in Table 4.

Table 4
Parameters used for full spectrum circular dichroism analysis. Circular dichroism was measured using a Jasco J-180 instrument along with the Spectra Measurement program within the Spectra Manager™ Suite. The same parameters were used on each 200 μM KSα1.1 variant sample.

Parameter	Value
Wavelength	260-190 nm
Data Pitch	0.5 nm
Start Mode	Immediately
Scanning Mode	Continuous
Scanning Speed	50 nm/minute
Channels	#1: CD, #2: HT
Sensitivity	Standard
D.I.T	1 second
Bandwidth	1.00 nm
Number of Accumulations	3

Circular Dichroism Data Analysis

Raw circular dichroism data was deconvoluted using an online platform called DichroWeb [208] to determine relative amounts of α-helical content in each KSα1.1 variant. Files collected from the Jasco Spectra Manager™ Suite were exported at text files (.txt) to be

used in the analysis. Concentrations in mg/mL and mean residue weights (MRW) were calculated for each protein variant. To ensure an accurate percentage of helical content, the data was subjected to three analysis programs (CDSSTR, CONTIN-LL, and SELCON-3) and averaged. All analysis programs used the same reference set, SMP180 (190-240 nm). Parameters and equations used in the data deconvolution are listed in Table 5.

Table 5

Parameters used for Dichroweb Deconvolution. The listed parameters were used for each KSα1.1 variant. MRWs were calculated using the protein variant molecular weight (MW) in Daltons and the total number of amino acids in the sequence (n). Protein concentrations were converted from 200 μM to their respective concentrations in mg/mL using the protein variant MW in Daltons. Path length refers to the width of the cuvette used in the analysis. The online DichroWeb tool can be accessed with the following link: http://dichroweb.cryst.bbk.ac.uk/html/home.shtml

Parameter	Value
File Format	Free with Preview (.txt)
Input Units	Millidegrees
Initial Wavelength	260 nm
Final Wavelength	190 nm
Lowest nm in analysis	190 nm
Wavelength Step	0.5 nm
Analysis Programs	CDSSTR, CONTIN-LL, SELCON-3
Reference Set	SMP180 (Optimized for 190-240 nm)
Scaling Factor	1.0
Output Units	Delta epsilon
Mean Residue Weight (Da)	$\frac{\text{Protein MW (Da)}}{\text{(n-1)}}$
Protein Concentration (mg/mL)	$\frac{\text{(MW)(200 μM)}}{10^6}$
Path Length	0.1 cm

Denaturation Studies

Chemical denaturation studies were done using increasing amounts of guanidine hydrochloride (GuHCl). Proteins were expressed and purified as described previously. Solutions of protein at 200 μM were prepared in 10 mM potassium phosphate buffer supplemented with denaturant ranging from 0.0 M to 3.0 M final GuHCl concentration in 0.2 M increments. Solutions were mixed gently and incubated at room temperature for 15 minutes prior to circular dichroism analysis. The change in circular dichroism at 222 nm was recorded and graphed against the respective denaturant concentration. Parameters are shown in Table 6.

Thermal denaturation studies were done by monitoring the circular dichroism change of a single protein sample at 222 nm over temperatures ranging from 10°C to 80°C. Proteins were expressed and purified as described previously. Protein samples at 200 μM concentration were prepared and kept on ice until circular dichroism analysis. The change in circular dichroism was recorded every 0.5°C, and graphed against the respective temperature. The resulting curves were analyzed further and used for ΔG calculation of each variant. Parameters are shown in Table 6.

Table 6

Parameters Used for Circular Dichroism Analysis Under Chemical and Thermal Denaturation Conditions. Circular dichroism was measured using a Jasco J-180 instrument along with the Spectra Measurement program (chemical denaturation) and Variable Temperature Measurement program (thermal denaturation) within the Spectra Manager™ Suite. The same parameters were used for each 200 μM KSα1.1 variant sample.

Parameter	Value
Chemical Denaturation	
Wavelength	223-221 nm
Data Pitch	0.5 nm
Start Mode	Immediately
Scanning Mode	Continuous
Scanning Speed	20 nm/minute
Channels	#1: CD, #2: HT
Sensitivity	Standard
D.I.T	1 second
Bandwidth	1.00 nm
Number of Accumulations	3
Thermal Denaturation	
Initial Temperature	10°C
Final Temperature	80°C
Sensitivity	Standard
D.I.T	4 seconds
Channels	#1: CD, #2: HT
Start Mode	Keep target temperature +/- 0.10°C for 5 seconds
Bandwidth	1.00 nm
Monitor Wavelength	222 nm
Data Pitch	1.0°C
Baseline Correction	None

Knob-Socket Predictions and Knob-Socket Hexagon Analysis

The Helix Knob-Socket Preference Test (http://tsailab.chem.pacific.edu/helix-socket-prediction.html) provides propensity information about protein sequences. This program will display every socket present in the inputted protein sequence, along with free and filled propensities, knob propensities, and a total helical propensity for each socket. When summed, the total socket propensity was predicted to correlate with the overall helical content and stability of a given protein. Helical lattice mapping analysis revealed that each single point mutation only affected six surrounding sockets, leading to the term Knob-Socket Hexagon. Total propensity values for each mutant KS Hexagon was calculated and compared to the wild type propensity value. Based on the differences in total propensities, mutant variants were predicted to have more or less helical content and therefore more or less stable, respectively. General trends (positive or negative changes) in total propensity were compared to the DichroWeb helical content values as well as the calculated ΔG values from thermal denaturation experiments.

Correlation Analysis

Values for total KS propensity, ΔG, and helical content were compiled for each mutant and graphed in three permutations: 1. KS propensity vs. ΔG, 2. ΔG vs. helical content, and 3. KS propensity vs. helical content. Each permutation was then analyzed via Pearson correlation for statistical significance.

Specificity of Quaternary Interactions in bZIP Coiled-Coils

bZIP protein interactions were investigated along with the amino acids within these coiled-coil proteins that dictate dimer specificity. Through KS mapping analysis, residue sequences that may play a role in coiled-coil specificity were identified, and their importance was analyzed experimentally through site-directed mutagenesis and BACTH studies. To confirm the KS analysis of bZIP specificity, a bacterial adenylate cyclase two hybrid (BACTH) system coupled with a beta galactosidase assay was used to test dimerization of coiled-coils. A BACTH system uses a protein interaction-mediated reconstitution of adenylate cyclase activity in *E. coli*. The catalytic domain of adenylate cyclase is composed of two fragments, T25 and T18. When separated, the enzyme is rendered nonfunctional. To take advantage of this, fusion proteins consisting of one of the adenylate cyclase fragments and a bZIP protein were made. If the bZIP proteins interacted during co-expression, the T25 and T18 fragments also interacted and returned functionality to the adenylate cyclase enzyme, as was measured by cAMP production.

Construct Design for pKT25 and pUT18C Expression Plasmids

Gene sequences for three bZIP proteins were taken from NCBI and mapped onto Knob-Socket helical lattices. The DNA binding region and leucine zipper regions for each transcription factor were identified, and residues were labelled using a heptad repeat pattern (*abcdefg*). The number and position of asparagine (Asn) residues along the leucine zipper region differed for each bZIP protein. A partial amino acid code including the full-length leucine zipper region and the two heptad repeats before the leucine zipper region were used for construct design. The amino acid sequence was reverse translated and codon optimized for expression. The optimized nucleotide sequence ends were modified to include a 5' *Bam*HI restriction site

along with a single adenine insertion that would ensure the proper reading frame after cloning into the BACTH expression vectors, as well as a 3' *Eco*RI restriction site preceded by an

Table 7

Modified bZIP Protein Sequences Used for BACTH Assay Construct Design. Sequences for *H. sapiens* bZIP transcription factors cJun, p21SNFT, and CREB4 were obtained from NCBI with the listed accession codes. Partial amino acid sequences shown were used based on the location within the coiled coil. Amino acid sequences were reverse translated and modified to contain a 5' *Bam*HI restriction site shown in red, a single adenine (A)base pair insert highlighted in gray, as well as a 3' engineered stop codon (TAG) highlighted in cyan, and *Eco*RI restriction site shown in blue.

bZIP Protein	**Accession Number**
cJun	BC009874
Nucleotide Sequence	
5' GGATCC A GCGGAACGCAAACGCATGCGCAACCGCATTGCGGCGAGCAAATGCCGCAAAC GCAAACTGGAACGCATTGCGCGCCTGGAAGAAAAAGTGAAAACCCTGAAAGCGCAGAACAGCGA ACTGGCGAGCACCGCGAACATGCTGCGCGAACAGGTGGCGCAGCTGAAACAGAAAGTGATGAAC CATGTGAACAGCGGCTGCCAGCTGATGCTGACCCAG TAG GAATTC 3'	
Amino Acid Sequence	
A E R K R M R N R I A A S K C R K R K L E R I A R L E E K V K T L K A Q N S E L A S T A N M L R E Q V A Q L K Q K V M N H V N S G C Q L M L T Q	
bZIP Protein	**Accession Number**
p21SNFT	NP061134
Nucleotide Sequence	
5' GGATCC A GAAAAAAACCGCGTGGCGGCGCAGCGCAGCCGCAAAAAACAGACCCAGAAAG CGGATAAACTGCATGAAGAATATGAAAGCCTGGAACAGGAAAACACCATGCTGCGCCGCGAAAT TGGCAAACTGACCGAAGAACTGAAACATCTGACCGAAGCGCTGAAAGAACATGAAAAAATGTGC CCG TAG GAATTC 3'	
Amino Acid Sequence	
E K N R V A A Q R S R K K Q T Q K A D K L H E E Y E S L E Q E N T M L R R E I G K L T E E L K H L T E A L K E H E K M C P	
bZIP Protein	**Accession Number**
CREB4	EAW53252
Nucleotide Sequence	
5' GGATCC A CGCCGCAAAATTCGCAACAAACAGAGCGCGCAGGATAGCCGCCGCCGCAAAA AAGAATATATTGATGGCCTGGAAAGCCGCGTGGCGGCGTGCAGCGCGCAGAACCAGGAACTGCA GAAAAAAGTGCAGGAACTGGAACGCCATAACATTAGCCTGGTGGCGCAGCTGCGCCAGCTGCAG ACCCTGATTGCGCAGACCAGCAACAAAGCGGCGCAGACCAGC TAG GAATTC 3'	
Amino Acid Sequence	
R R K I R N K Q S A Q D S R R R K K E Y I D G L E S R V A A C S A Q N Q E L Q K K V Q E L E R H N I S L V A Q L R Q L Q T L I A Q T S N K A A Q T S	

engineered stop codon. The bZIP sequences used for construct building are shown in Table 7. The modified bZIP sequences were synthesized and cloned into a pET-24a(+) plasmid (GenScript). Newly synthesized pET-24a(+) plasmids containing the bZIP inserts as well as empty pKT25 and pUT18C plasmids (Risser lab, University of the Pacific) were transformed and purified for cloning procedures.

In addition to the three sequences shown, a point mutant of CREB4 was also synthesized. cJun and p21SNFT proteins contained a single Asp residue at positions a3 and a2, respectively. CREB4 contained two Asp residues at positions a3 and a5. The point mutant synthesized contained the a3 Asp, but the a5 Asp was mutated to a Leucine (Leu). This point mutant was termed CREB4_N49L. This was intended to determine if specificity could be changed based on Asp number and location within the binding region.

Construct Building for pKT25 and pUT18C Expression Plasmids with bZIP Inserts

Purified plasmid stocks of the empty expression plasmids, pKT25 and pUT18C, along with purified pET-24a(+) plasmids containing the bZIP sequences were digested in 20 μL reactions containing 1X CutSmart Buffer (New England Biolabs) and 20 U of both *Eco*RI HF and *Bam*HI HF restriction enzymes. The digestions were incubated at 37°C for 15 minutes, and heat inactivated at 65°C for 20 minutes. Digested samples were electrophoresed on a 3% agarose gel containing 1 μg/mL ethidium bromide (EtBr) at 100 V for 20 minutes. Bands were visualized using a handheld UV lamp, excised and purified according to the QIAquick Gel Extraction Kit protocol (Qiagen). Purified linearized plasmids and inserts were ligated in 10 μL reactions containing 1X T4 DNA Ligase Buffer and 200 U of T4 DNA Ligase (New England Biolabs) for 10 minutes at room temperature. The ligation reactions were transformed into

XL10-Gold® Ultracompetent cells and plated onto LB agar plates containing either 50 μg/mL kanamycin (pKT25 ligations) or 100 μg/mL ampicillin (pUT18C ligations), and incubated overnight at 37°C. Resulting colonies were grown to saturation and plasmids were purified according to the QIAprep Spin Miniprep Kit protocol (Qiagen). Purified plasmids were checked for insert diagnostically via restriction digestion with EcoRI and BamHI and gel electrophoresis. Positive clones were confirmed via sequencing (Sequetech) using the primers listed in Table 8.

Table 8
Primers used for pKT25 and pUT18C Construct Sequencing.

Construct	Primer Sequences
pKT25	5′ GTTTTCCCAGTCACGAC 3′
pUT18C	5′ TATGCGGCATCAGAGCAG 3′

Site-Directed Mutagenesis

Mutagenesis was performed on the pKT25 and pUT18C constructs containing bZIP to investigate coiled-coil specificity. Specific amino acids were selected for site-directed mutagenesis (SDM) based on placement within the pseudo-heptad repeat pattern along the binding edge. Primers for point mutants were designed according to the QuikChange Lightening II Site-Directed Mutagenesis Kit (Agilent) primer design protocol and are shown below in Table 9. PCR reactions were run in a 50 μL total reaction volume containing 1X Reaction Buffer, 125 ng/primer, 10 ng template DNA, 0.5 mM dNTP mix, 3 μL of QuikSolution, and 2.5 U of *Pfu*Ultra Polymerase. The PCR program was created with the following parameters: initial

denaturation at 95°C for 50 seconds, denaturation at 95°C for 50 seconds, annealing at 54°C for 1 minute, extension at 68°C for 10 minutes and returning to the denaturation step for a total of 18 cycles. PCR reactions were then *Dpn*I digested for 1 hour at 37°C, and transformed into XL10-Gold® Ultracompetent cells that had previously been incubated with a proprietary BME mixture (Agilent) at 4°C for 30 minutes. The transformed cells were recovered in SOC broth for 1 hour, and plated on LB agar plates containing 50 µg/mL kanamycin or 100 µg/mL ampicillin. After incubation at 37°C overnight, colonies were picked and purified via miniprep (Qiagen) and confirmed through sequencing (Sequetech).

Table 9

Primer Sequences for Site-Directed Mutagenesis of bZIP Sequences in pKT25 or pUT18C Plasmids. Primers were designed using the online Agilent primer design tool. Base pair mismatches for each primer set are bolded and highlighted in yellow.

Mutant	Primer Sequences
CREB4_H48K/N49L	**F:** 5' CCACCAGGCTAATGAG**CTT**GCGTTCCAGTTCCTGC 3'
	R: 5' GCAGGAACTGGAACGC**AAG**CTCATTAGCCTGGTGG 3'
CREB4_H48A/N49L	**F:** 5' CACCAGGCTAATGAG**AGC**GCGTTCCAGTTCCTGC 3'
	R: 5' GCAGGAACTGGAACGC**GCT**CTCATTAGCCTGGTG 3'
CREB4_H48T/N49L	**F:** 5' CCACCAGGCTAATGAG**AGT**GCGTTCCAGTTCCTGCA 3'
	R: 5' TGCAGGAACTGGAACGC**ACT**CTCATTAGCCTGGTGG 3'
CREB4_H48Q/N49L	**F:** 5' CACCAGGCTAATGAG**CTG**GCGTTCCAGTTCC 3'
	R: 5' GGAACTGGAACGC**CAG**CTCATTAGCCTGGTG 3'
CREB4_N49L/I50A	**F:** 5' TGCGCCACCAGGCT**AGC**GAGATGGCGTTCCAG 3'
	R: 5' CTGGAACGCCATCTC**GCT**AGCCTGGTGGCGCA 3'
cJun_V51N	**F:** 5' CTTTCTGTTTCAGCTGCGC**ATT**CTGTTCGCGCAGCATGTTC 3'
	R: 5' GAACATGCTGCGCGAACAG**AAT**GCGCAGCTGAAACAGAAAG 3'
p21SNFT_L46N	**F:** 5' ATTGGCAAACTGACCGAAGAA**AAT**AAACATCTGACCGAAGCGCTG 3'
	R: 5' CAGCGCTTCGGTCAGATGTTT**ATT**TTCTTCGGTCAGTTTGCCAAT 3'

Preparation of BTH101 Competent Cells

A LB agar plate containing BTH101 (F-, cya-99, araD139, galE15, galK16, rpsL1 (Str r), hsdR2, mcrA1, mcrB1) cells was provided by the Risser lab (University of the Pacific). A single colony was grown to saturation in LB overnight at 37°C. The saturated culture was added to 100 mL of fresh LB and incubated until an $OD_{550} \sim 0.5$ was reached. The culture flask was then incubated on ice for 10 minutes. The cells were centrifuged in at 50 mL conical tube at 5,000 rpm for 5 minutes, and the supernatant was discarded. The pellet was resuspended in 15 mL of pre-chilled TJB1 Buffer (100 mM RbCl, 50 mM $MnCl_2 \bullet 4H_2O$, 30 mM KOAc, 10 mM $CaCl_2 \bullet 2H_2O$, 15% w/v glycerol, pH 5.8, filter sterilized). After another 5 minute incubation on ice, the cells were centrifuged again at 5,000 rpm for 5 minutes. The supernatant was removed, and the pellet was resuspended in 4 mL of ice cold TJB2 Buffer (10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2 \bullet 2H_2O$, 15% w/v glycerol, pH 7.0, filter sterilized). The resuspended solution was added to pre-chilled 1.5 mL centrifuge tubes in 50 mL aliquots. The aliquots were flash frozen in an ethanol bath and further stored at -80°C. Transformation efficiency of the cells was assessed using standard transformation protocols.

Bacterial Adenylate Cyclase Two Hybrid (BACTH) Assay

In addition to the pKT25 and pUT18C plasmids containing bZIP sequences, positive and negative controls were prepared. Positive control plasmids were obtained from the Risser Lab (University of the Pacific), which were pKT25 and pUT18C constructs containing truncated GCN4 sequences known to interact strongly. Empty pKT25 and pUT18C plasmids were prepared and used for the negative control. BACTH assay buffers were prepared before experiments were conducted, including Permeabilization Solution (100 mM Na_2HPO_4, 20 mM

KCl, 2 mM $MgSO_4$, 0.8 mg/mL CTAB, 0.4 mg/mL sodium deoxycholate), Substrate Solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, autoclaved) and Stop Solution (1 M Na_2CO_3, autoclaved).

Plasmids were co-transformed into BTH101 cells and plated on MacConkey agar containing 20% maltose, 0.5 mM IPTG, as well as 50 μg/mL kanamycin and 100 μg/mL ampicillin for selection of positive co-transformants. Interaction between the two hybrid proteins leads to cAMP production, and gives a *cya*$^+$ phenotype to cells that are non-reverting adenylate cyclase deficient (*cya*). *Cya*$^+$ colonies will appear red on MacConkey media, while *cya* colonies will appear faintly pink or white. Plates were incubated at 30°C for 48 hours. Colonies were examined after incubation for preliminary interaction data.

Plasmids were also co-transformed and grown on LB agar plates containing 1 mM IPTG, 50 μg/mL kanamycin and 100 μg/mL ampicillin for selection of positive co-transformants. Plates were grown at 30°C for 48 hours. Colonies were selected and grown in triplicate reactions in 3.0 mL of LB/IPTG/Kan/Amp media at 30°C overnight. Permeabilization Solution was supplemented with 5.4 μL/mL BME before use, and was aliquoted in 80 μL volumes for each sample to fresh 1.5 mL centrifuge tubes. Substrate Solution was supplemented with 1 mg/mL ONPG and 2.7 μL/mL BME. Co-transformed liquid cultures were removed from incubation, and 20 μL of each sample was added to one of the 80 μL Permeabilization Solution aliquots, and inverted to mix. The samples and complete Substrate Solution were incubated at 30°C for 20 minutes. After incubation, the samples were ready for substrate addition. One-by-one, 600 μL of Substrate Solution was added to each sample. Upon addition of substrate, a timer was started. The sample was capped and inverted until the appearance of faint yellow color. Simultaneously, 700 μL of Stop Solution was added while the timer was stopped. The time was recorded in minutes, and the process repeated for all remaining samples.

300 µL volumes of each BACTH reaction were added to wells of a BioLite 48-well Clear Bottom Multidish (ThermoFisher). Additionally, 300 µL volumes of each bacterial cell culture used in the assay was added to the 48-well dish. Using a Synergy HI Microplate Reader (Biotek), the A420 of the BACTH reactions and A600 of each bacterial cell culture were taken and recorded. Corresponding Miller Units for each interaction was calculated using the equation in Figure 5. Miller Units in each experiment was analyzed using Prism 7 (GraphPad). Data was analyzed for each set of triplicates via One-Way ANOVA with multiple comparisons, and Tukey's post hoc test was conducted as a correction for multiple comparisons.

$$1000\ x\ \frac{(A420)}{(A600\ x\ 0.02\ mL\ x\ reaction\ time)}$$

Figure 5. Equation for Miller Unit calculation. The equation below was used for calculation of Miller Units for each sample. A420 and A600 readings were taken from the Synergy HI Microplate reader, reaction volume was 0.02 mL, and reaction time was used in minutes.

CHAPTER 3: RESULTS

Stability Relationship of Knob-Socket Propensities Measured in the KSα1.1 α-Helix

The KSα1.1 amino acid sequence was previously designed on the Knob-Socket helical lattice by creating free sockets surrounding a pattern of filled sockets as well as the predicted knob residues that would pack in a parallel coiled-coil formation (Figure 6A). The helical lattice also indicates the spatial relationships and bonding relationships between each KSα1.1 residue in a helical conformation. As shown by Figure 6B, a residue in the middle of a helix has direct interactions on 6 residues, whose sockets form a hexagon of interaction. Point mutants were selected, and Knob-Socket analysis was conducted using the Helix Knob-Socket Preference Test online platform. This analysis defines that a single point mutant is affected by the six surrounding sockets. These residues and their sockets make up Knob-Socket Hexagon (Figure 6B).

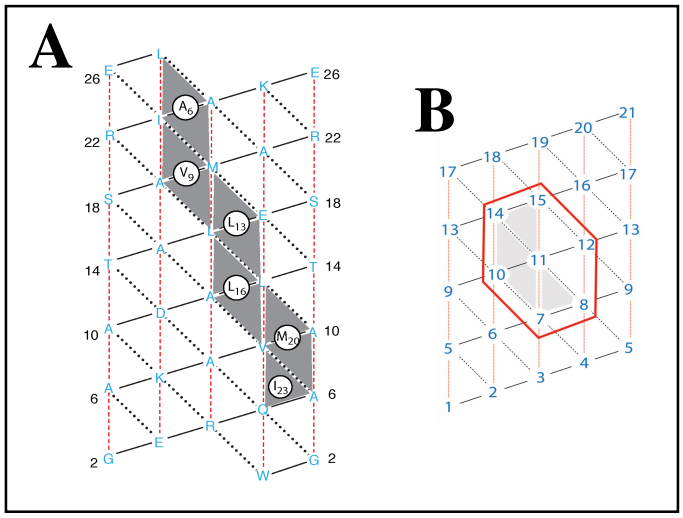


Figure 6. KSα1.1 helical lattice mapping and the Knob-Socket Hexagon. (A) The KSα1.1 peptide sequence was mapped onto a Knob-Socket helical lattice, with the residues colored blue. Covalent peptide bonds are represented by solid black lines; hydrogen bonds are represented by red dashed lines; and Van der Waals interactions are represented by dotted black lines. Free sockets are white, and filled sockets are shaded in gray. Knob residues packing into the filled sockets are shown as white circles with a packing pattern of a parallel helical coiled-coil. Numbers on either side of the lattice refer to the position of that residue within the sequence. (B) In an α-helix, each residue directly contacts the 6 surrounding residues. As an example around the 11th residue, the pattern of 6 sockets involved in a Knob-Socket Hexagon is outlined in red.

Computational Knob-Socket Hexagon analysis (Figure 7) of the point mutants led to a value termed the 'Socket Propensity Difference.' This was calculated by first totaling the propensity values for the six participating sockets affected by the particular residue being mutated in the wild type and mutant hexagons, respectively. Socket propensity values were calculated based on socket amino acid composition over a comprehensive set of protein structures [209]. To calculate the Socket Propensity Difference, the wild-type socket propensity value was subtracted from the mutant socket propensity value. The sign of the difference served as a preliminary prediction of whether the mutation would positively or negatively affect the structure and stability of the wild type KSα1.1. So, if the socket propensity value of the mutant was lower than wild-type and therefore an indication of helix destabilization by the mutant, a negative value would result for the Socket Propensity Difference. Conversely, if the socket propensity of the mutant is higher than wild-type and therefore an indication of helix stabilization, a positive Socket Propensity Difference would result. Of the 10 designed point mutants, 4 were predicted to stabilize the wild type peptide, and 6 were predicted to destabilize the wild type peptide (Table 10).

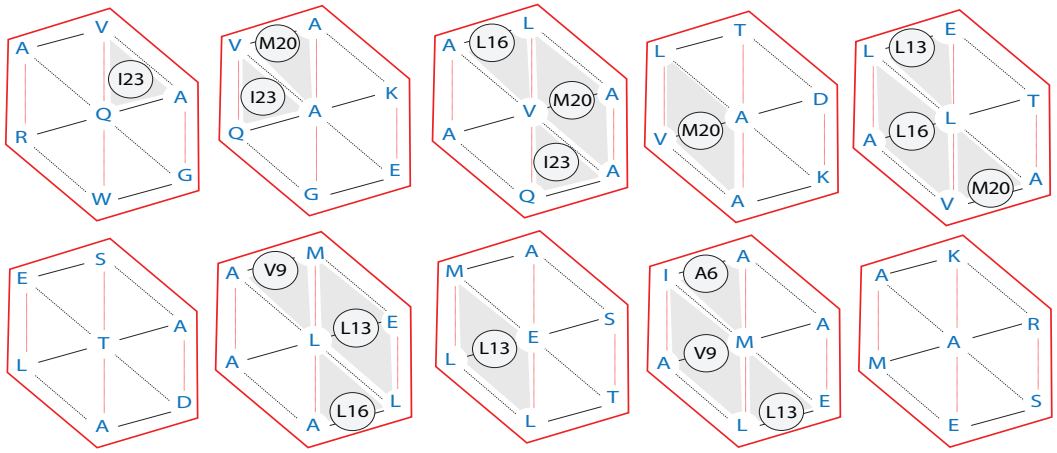


Figure 7. Knob-Socket Hexagons affected by each KSα1.1 point mutation.

Table 10

Knob-Socket Propensity Differences for Each KSα1.1 Point Mutation. The wild type and mutant sequences are shown with their respective socket propensity difference. Socket propensity differences were calculated by taking the total KS propensity of a mutant variant and subtracting the wild type total propensity value. Increases in propensity are colored blue, while decreases in propensity are colored red.

Mutant	Sequence	Socket Propensity Difference
Wild Type	**WGERQAKAVADALTALESAMARIAKEL**	-----
Q5L	WGERLAKAVADALTALESAMARIAKEL	24.690
A6E	WGERQEKAVADALTALESAMARIAKEL	-11.620
V9E	WGERQAKAEADALTALESAMARIAKEL	-4.650
A10R	WGERQAKAVRDALTALESAMARIAKEL	-21.820
L13I	WGERQAKAVADAITALESAMARIAKEL	-32.310
T14V	WGERQAKAVADALVALESAMARIAKEL	6.440
L16F	WGERQAKAVADALTAFESAMARIAKEL	-53.740
E17D	WGERQAKAVADALTALDSAMARIAKEL	-11.440
M20L	WGERQAKAVADALTALESALARIAKEL	57.370
A21E	WGERQAKAVADALTALESAMERIAKEL	0.600

KSα1.1 fusion proteins were expressed and purified via nickel affinity chromatography. The fusion proteins contained a 6X Histidine tag that was immobilized on Ni-NTA agarose resin. After washing, the fusion KSα1.1 proteins were eluted using a high concentration of imidazole. The fusion proteins were then treated with ULP-1 to remove the 6X Histidine + SUMO tags. The protease treated mixture was then subjected to another round of affinity chromatography. In this phase, the cleaved KSα1.1 protein was expected to be collected in the flow-through fraction, while the 6X Histidine + SUMO tags would be found after elution with imidazole. Fractions for each purification procedure were collected and analyzed using SDS-PAGE.

SDS-PAGE for the initial fusion protein purification of wild-type KSα1.1 is shown in Figure 8. The gel was loaded with PageRuler Protein Ladder (Lane 1), flow through (Lane 2), wash 1 (Lane 3), wash 4 (Lane 4), and elutions 1-4 (Lanes 5-8). After Coomassie staining, bands of approximately 18 kDa were detected in all elution fractions. This matched with the expected size of the KSα1.1 fusion protein (6XHis+SUMO ~ 15 kDa, KSα1.1- 2.9 kDa). This result indicated that the fusion protein was isolated successfully and could be used for the next phase of the purification procedures.

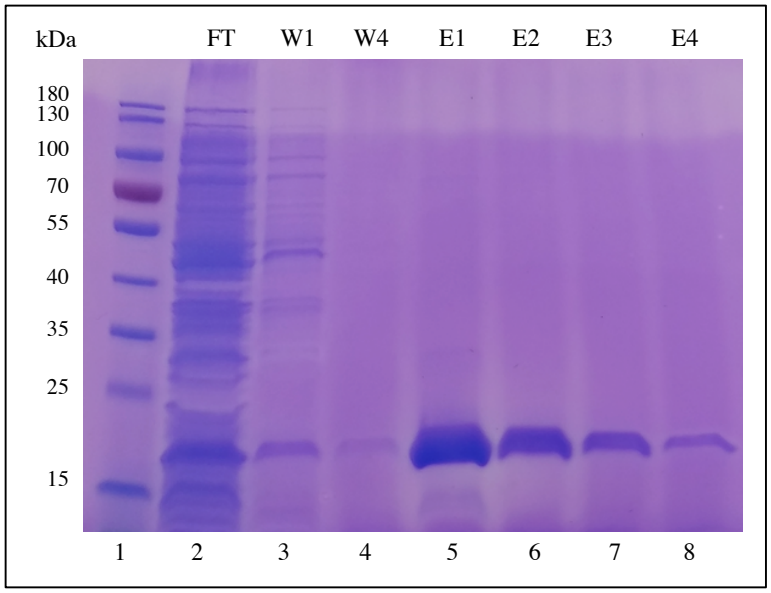


Figure 8. Purification of 6XHis-SUMO-KSα1.1 fusion proteins. The gel was loaded with PageRuler Protein Ladder (Lane 1), flow through (Lane 2), wash 1 (Lane 3), wash 4 (Lane 4), and elutions 1-4 (Lanes 5-8). A band of ~18 kDa corresponding to the fusion protein is present in all elution fractions (Lane 5-8).

After the purification of the fusion KSα1.1 protein and subsequent ULP-1 treatment, the sample was analyzed via Tris/Tricine SDS-PAGE (Figure 9). This was done to improve

resolution of the relatively smaller cleaved KSα1.1 peptide, which is outside of the typical resolving range for glycine gels. The Tris/Tricine gel was loaded identically to the glycine gels with PageRuler Protein Ladder (Lane 1), flow through (Lane 2), wash 1 (Lane 3), wash 4 (Lane 4), and elutions 1-4 (Lanes 5-8). After Coomassie staining, a band of ~ 3 kDa can be seen in the flow through fraction (Lane 2), which corresponded to cleaved KSα1.1. 6X Histine + SUMO tags can be seen in the wash and elution fractions around the expected size of ~15 kDa. This result along with the absence of any other bands in the flow through fraction suggested that protease treatment was successful, and that the cleaved KSα1.1peptide had been purified away from other proteins. The cleaved KSα1.1 peptide sample was concentrated, desalted and used downstream in circular dichroism analysis.

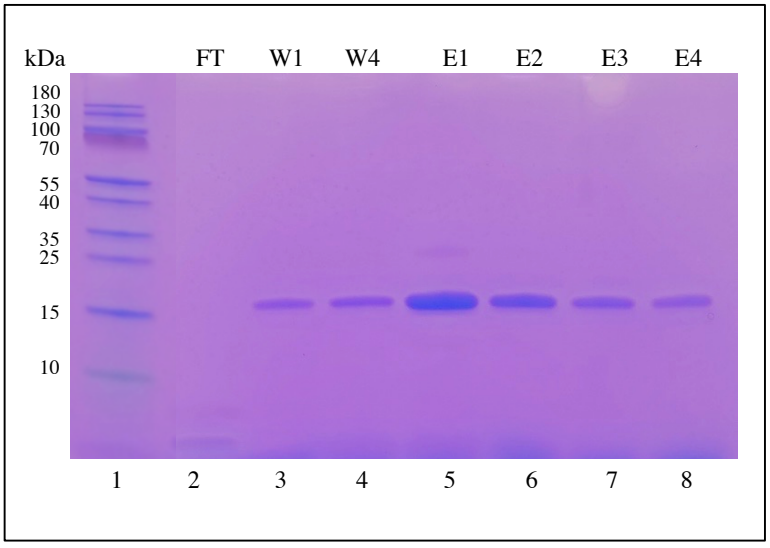


Figure 9. Purification of cleaved KSα1.1 proteins. The gel was loaded with PageRuler Protein Ladder (Lane 1), flow through (Lane 2), wash 1 (Lane 3), wash 4 (Lane 4), and elutions 1-4 (Lanes 5-8). A band of ~3 kDa corresponding to the cleaved KSα1.1 protein is present in the flow through fraction (Lane 2).

Circular dichroism (CD) was measured from 260 nm - 190 nm in 200 μM samples of each KSα1.1 variant and graphed using Plot2 (Figure 10). Each peptide signature showed features of α-helical proteins with two minima at 222 nm and 208 nm, and a maximum around 195 nm. Most mutant signatures were similar in intensity to the wild type, with the greatest deviants being point mutants Q5L and M20L. Raw CD data was used in deconvolution analysis to determine the differences in helical structure. Fractional helical percentages were collected from three analysis programs (CONTIN-LL, CDSSTR, and SELCON-3) and averaged (Table 11). Wild type KSα1.1 was shown to contain 27.8% helical secondary structure, and the point mutant variants ranged from 15.6% to 55.7% helical content. Averaged helical percentages for each mutant were compared to the wild type value as well as to the predictions made from the preliminary Knob-Socket Hexagon analysis. All ten predictions matched the DichroWeb results in terms of positive or negative changes in helical content, where the four predicted stabilizing mutations exhibited an increase in helical content and the six predicted destabilizing mutations exhibited a decrease in helical content when compared to wild type. While the changes matched in sign, the corresponding magnitudes did not appear to correlate.

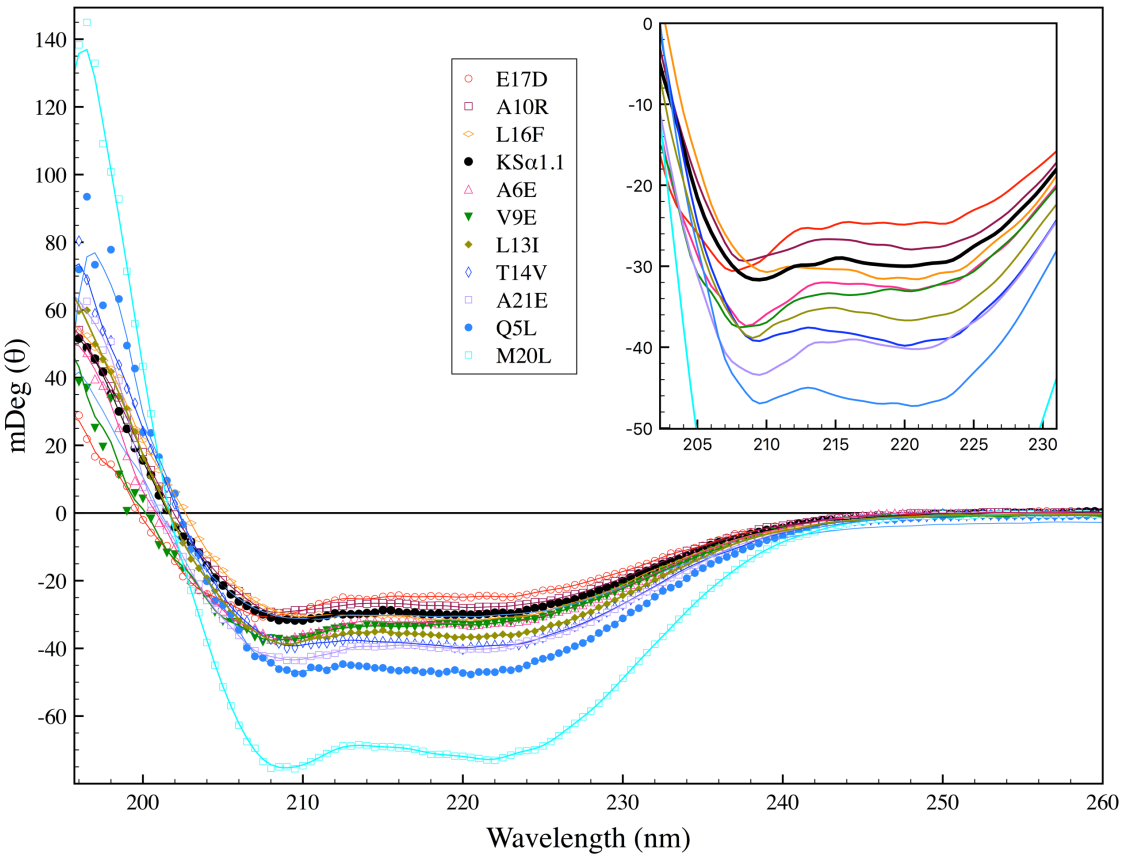


Figure 10. Secondary structure analysis of KSα1.1 protein and mutant variants via circular dichroism. Circular dichroism for each 200 µM protein sample was measured from 260-190 nm. Data was collected in millidegrees, and graphed against wavelength. A typical CD signature for alpha helices is seen for all variants. The inlaid plot is shown to resolve differences within the 230-205 nm range.

Table 11

Deconvolution of Raw Circular Dichroism Data using DichroWeb. Fractional alpha-helical percentages were calculated using the DichroWeb online program using three analysis programs: CONTIN-LL, CDSSTR, and SELCON-3. Values were averaged and colored according to an increase in alpha-helical content (blue) or a decrease in alpha-helical content (red). Values that could not be calculated are shown as dashed lines.

Mutant	WT	Q5L	A6E	V9E	A10R	L13I	T14V	L16F	E17D	M20L	A21E
CONTIN-LL	0.289	0.466	0.221	0.223	0.172	0.218	0.243	0.190	0.161	0.444	-----
CDSSTR	0.310	0.450	-----	-----	0.080	0.220	0.410	0.180	0.150	0.650	0.400
SELCON-3	0.230	0.333	0.297	0.238	-----	-----	0.299	-----	-----	0.578	0.290
AVERAGE	**0.278**	**0.416**	**0.259**	**0.231**	**0.126**	**0.219**	**0.327**	**0.185**	**0.156**	**0.557**	**0.345**

Relative stabilities of all KSα1.1 variants were also investigated through chemical and thermal denaturation studies. The circular dichroism at 222 nm was monitored over ranges of chemical denaturant or temperature ranges on 200 μM samples of each KSα1.1 variant and graphed using Plot2 (Figure 11). Chemical denaturation of KSα1.1 variants was conducted using 0.2 M increments of GuHCl, and thermal denaturation was conducted over a range of 10°C-80°C. As seen in the full CD spectra, most mutant signatures remained similar to the wild type denaturation curves, again with the exception of point mutants Q5L and M20L. However, Q5L maintained the almost linear trend seen with the wild type and other mutants, while M20L exhibited a more typical unfolding curve in both experiments.

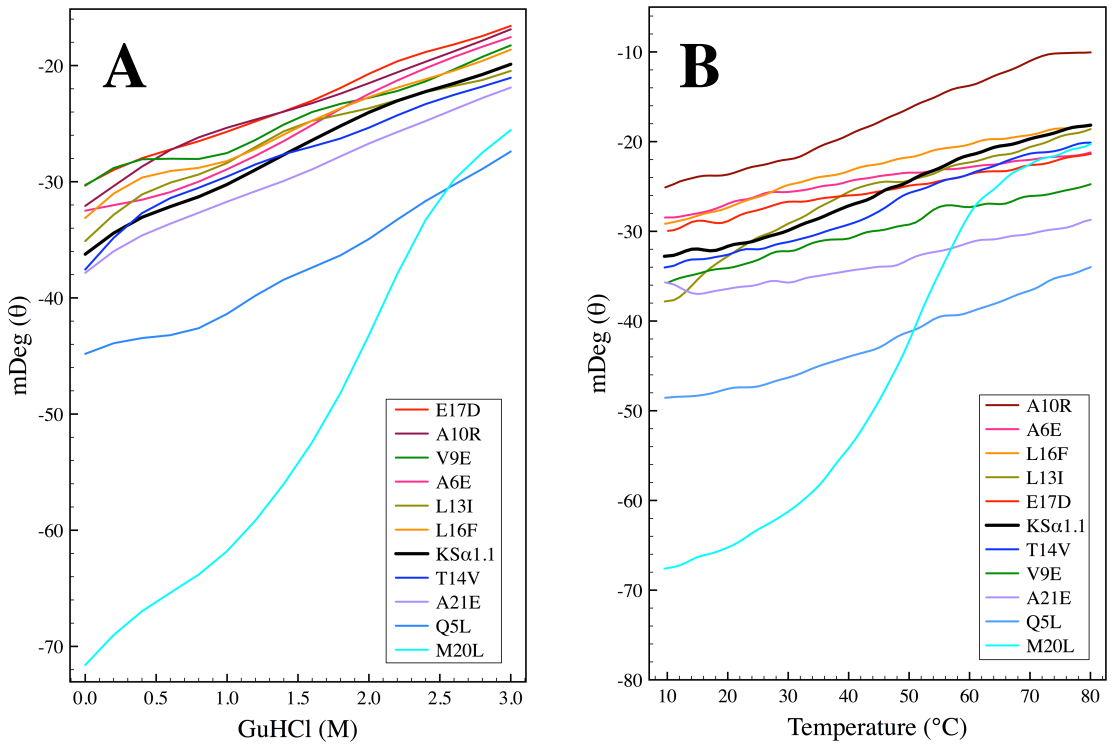


Figure 11. Denaturation studies of KSα1.1 protein and mutant variants. The circular dichroism at 222 nm was monitored over ranges of chemical denaturant (A) or temperature (B) ranges on 200 μM samples of each KSα1.1 variant.

Next, Knob-Socket Hexagon analysis and total propensities were calculated for the two designed double point mutants, T14V/M20L and T14V/A21E. In both cases, the mutant hexagons do not overlap with each other and therefore the resulting effects are additive. For example, with the T14V/M20L double mutant, the single mutant Socket Propensity Differences are 6.440 (T14V) and 57.370 (M20L) could simply be added to give a difference of 63.810 for the corresponding double mutant (Table 12).

Table 12

Knob-Socket Propensity Differences for Each KSα1.1 Double Point Mutant and Corresponding Single Mutants. The wild type and mutant sequences are shown with their respective socket propensity difference. Socket propensity differences were calculated by taking the total KS propensity of a mutant variant and subtracting the wild type total propensity value. All resulting socket propensity differences are positive and colored blue.

Mutant	Sequence	Socket Propensity Difference
Wild Type	**WGERQAKAVADALTALESAMARIAKEL**	-----
T14V	WGERQAKAVADALVALESAMARIAKEL	6.440
M20L	WGERQAKAVADALTALESALARIAKEL	57.370
A21E	WGERQAKAVADALTALESAMERIAKEL	0.600
T14V/M20L	WGERQAKAVADALVALESALARIAKEL	7.040
T14V/A21E	WGERQAKAVADALVALESAMERIAKEL	63.810

Circular dichroism (CD) was measured from 260 nm - 190 nm on 200 μM samples of each KSα1.1 double mutant and graphed along with wild type and the appropriate point mutants using Plot2 (Figure 12). As expected, each double point mutant peptide signature also showed features of α-helical proteins with two minima at 222 nm and 208 nm, and a maximum around 195 nm. Additionally, each double point mutant signature displayed a stronger intensity than the single point mutants. Raw CD data was used in deconvolution analysis to determine the

differences in helical structure. Fractional helical percentages were collected from three analysis programs (CONTIN-LL, CDSSTR, and SELCON-3) and averaged (Table 13).

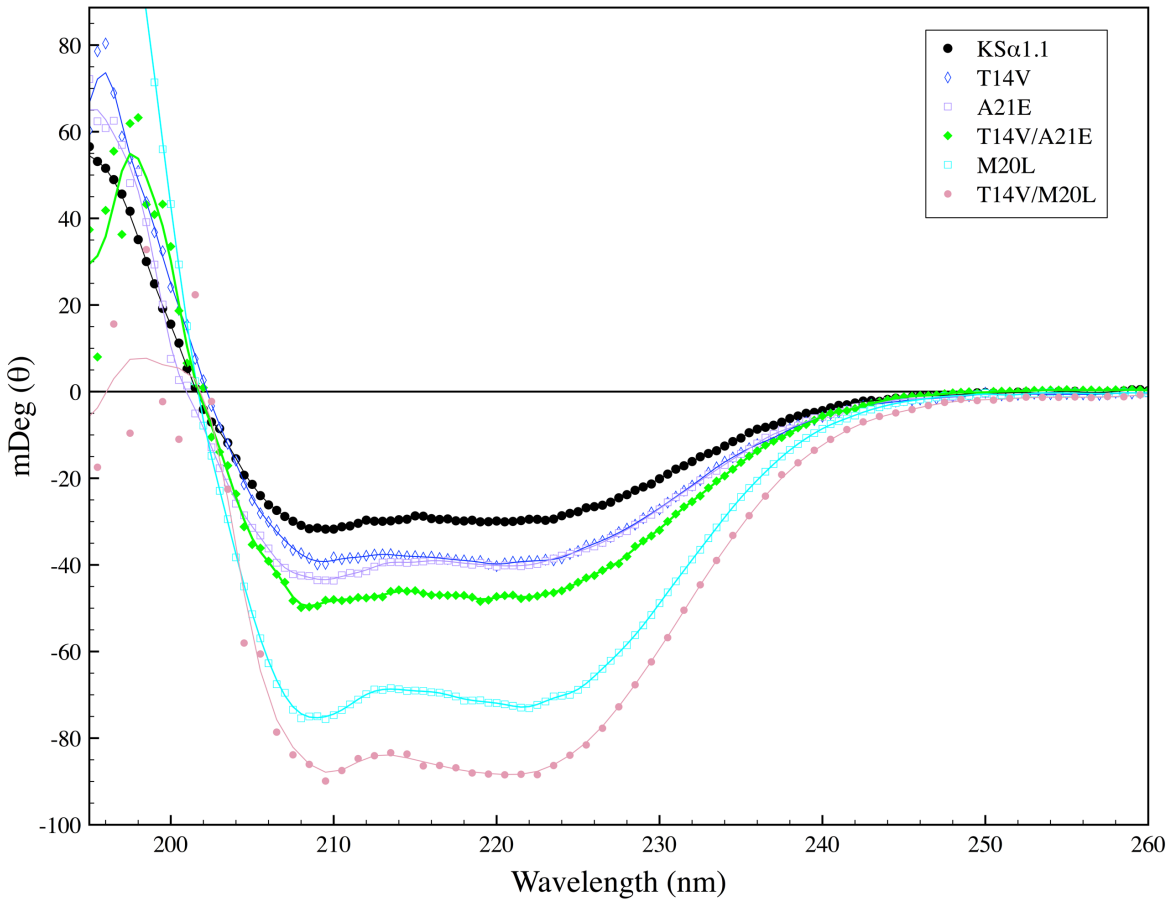


Figure 12. Secondary structure analysis of each KSα1.1 double point mutant and corresponding single mutants via circular dichroism. Circular dichroism for each 200 μM protein sample was measured from 260-190 nm. Data was collected in millidegrees, and graphed against wavelength. A typical CD signature for alpha helices is seen for all variants.

Table 13

Deconvolution of Raw Circular Dichroism Data for KSα1.1 Double Mutants and Corresponding Single Mutants. Fractional alpha-helical percentages were calculated using the DichroWeb online program using three analysis programs: CONTIN-LL, CDSSTR, and SELCON-3. Values were averaged and colored according to an increase in alpha-helical content (blue) or a decrease in alpha-helical content (red). Values that could not be calculated are shown as dashed lines.

Mutant	WT	T14V	M20L	A21E	T14V/M20L	T14V/A21E
CONTIN-LL	0.289	0.243	0.444	-----	0.530	0.518
CDSSTR	0.310	0.410	0.650	0.400	0.690	0.430
SELCON-3	0.230	0.299	0.578	0.290	-----	0.330
AVERAGE	**0.278**	**0.327**	**0.557**	**0.345**	**0.610**	**0.426**

After deconvolution, it was determined that the double mutants contained a larger percent of helical structure relative to their point mutant counterparts. Interestingly, when looking at the magnitudes of increases in helical content among mutant variants compared to the wild type, the changes were also nearly additive as seen with the Socket Propensity changes. For example, when looking at the corresponding point mutants for the T14V/M20L double mutant, a 4.9% and 27.9% increase in helical structure was observed (T14V and M20L, respectively), giving a predicted total increase of 32.8%. This predicted value from the hexagons is extremely close to the experimentally calculated value of 33.2%. This suggests that the changes in socket propensities have a defined effect on the overall change in helical content.

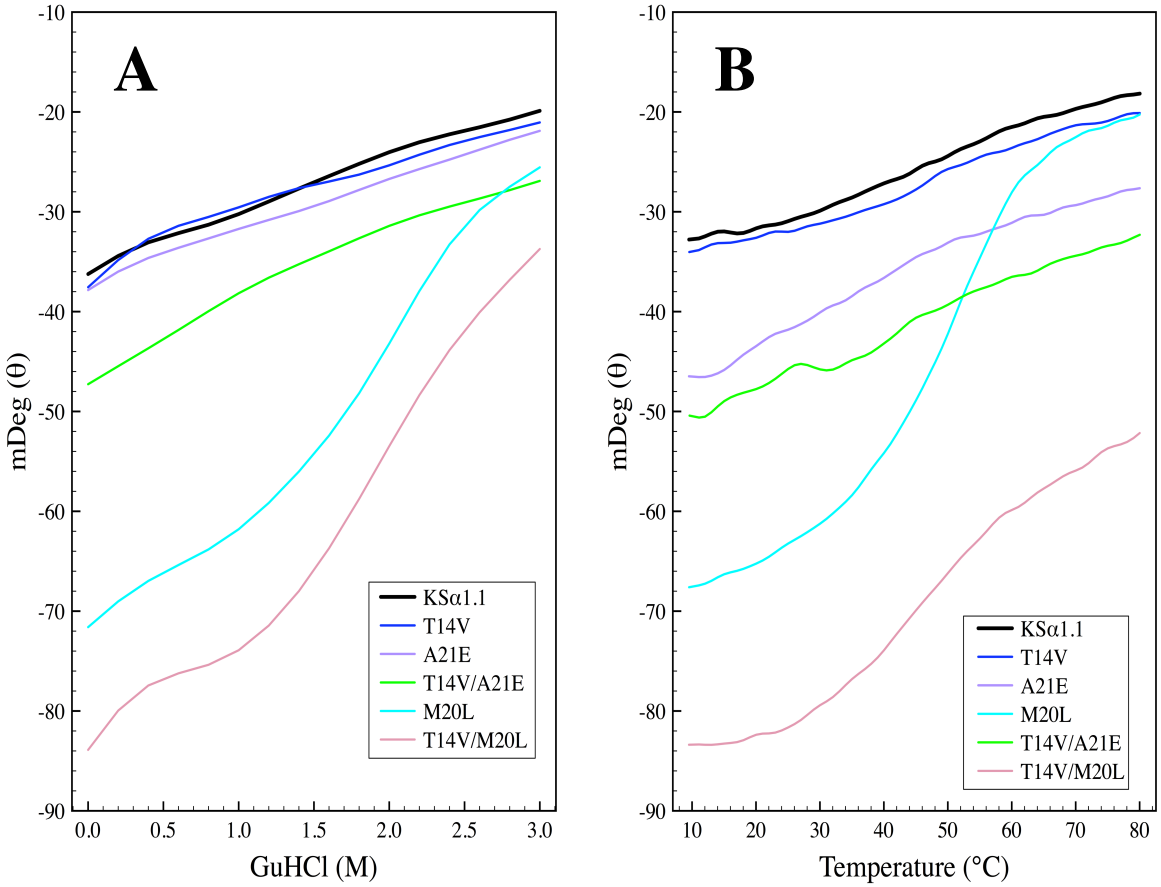


Figure 13. Denaturation studies of each KSα1.1 double mutant protein and corresponding single mutant variants. The circular dichroism at 222 nm was monitored over ranges of chemical denaturant (A) or temperature (B) ranges on 200 μM samples of each KSα1.1 variant.

Again, relative stabilities of the KSα1.1 double mutant variants were investigated through chemical and thermal denaturation studies as described previously. The circular dichroism at 222 nm was monitored over ranges of chemical denaturant or temperature ranges on 200 μM samples of KSα1.1 double mutant variants and graphed along with the corresponding point mutant curves using Plot2 (Figure 13). As seen in the full CD spectra, the double mutant

curves maintained similar shapes as their single mutant counterparts, but displayed a stronger intensity. The double mutation containing M20L again displayed a more typical sigmoidal, cooperative unfolding curve shape relative to the other more linear curves. A summary of the computational predictions and experimentally derived helical percentages for all single and double mutants are displayed in Table 14. Overall, it was observed that the differences (either propensity or relative helical percentage) correlated in terms of positive or negative changes, while the magnitudes did not appear to correlate as well.

Table 14
Summary of Socket Propensity Differences and Averaged Fractional Alpha-Helical Percentages for All KSα1.1 Variants. Values are colored based on an increase (blue) or decrease (red) relative to the wild type measurement. All KS predictions matched in terms of sign with the experimental values.

Mutant	Sequence	Socket Propensity Difference	DichroWeb
Wild Type	**WGERQAKAVADALTALESAMARIAKEL**	-----	0.278
Q5L	WGERLAKAVADALTALESAMARIAKEL	24.690	0.416
A6E	WGERQEKAVADALTALESAMARIAKEL	-11.620	0.259
V9E	WGERQAKAEADALTALESAMARIAKEL	-4.650	0.231
A10R	WGERQAKAVRDALTALESAMARIAKEL	-21.820	0.126
L13I	WGERQAKAVADAITALESAMARIAKEL	-32.310	0.219
T14V	WGERQAKAVADALVALESAMARIAKEL	6.440	0.327
L16F	WGERQAKAVADALTAFESAMARIAKEL	-53.740	0.185
E17D	WGERQAKAVADALTALDSAMARIAKEL	-11.440	0.156
M20L	WGERQAKAVADALTALESALARIAKEL	57.370	0.557
A21E	WGERQAKAVADALTALESAMERIAKEL	0.600	0.345
T14V/M20L	WGERQAKAVADALVALESALARIAKEL	7.040	0.610
T14V/A21E	WGERQAKAVADALVALESAMERIAKEL	63.810	0.426

Lastly, statistical analysis was conducted to determine if there were in fact any significant correlations between the predicted Socket Propensity Differences and experimentally determined

values. Permutations of total KS Propensity, socket propensity difference, stability (ΔG), and helical percentage values (Table 15) were used in Pearson correlation analyses, and graphed using GraphPad Prism 8 (Figure 14). In Figure XXXA & B, percentage of helical content (DichroWeb) was correlated to the Socket Propensity Difference, or the total KS Propensity, respectively. Total KS Propensity versus helical content was calculated to have a stronger correlation with a $R^2 = 0.8276$, $p < 0.0001$, compared to Socket Propensity Difference versus helical content with a $R^2 = 0.5339$, $p = 0.0046$. In Figure XXXC & D, ΔG values were correlated to helical content and total KS Propensity, respectively. Stability versus helical content was calculated to have the weakest correlation with a $R^2 = 0.3932$, $p = 0.0218$. Total KS Propensity versus stability correlated moderately well with a $R^2 = 0.5042$, $p = 0.0065$.

Table 15
Values Used in Pearson Correlation Analysis. Values used in statistical analyses for each point mutant are shown below. Total Socket Propensity and Socket Propensity Differences were calculated using the Helix preference test. Helical content values were taken from the fractional percentages calculated using the DichroWeb program. ΔG values were calculated from thermal denaturation curves.

Mutant	Sequence	Total Socket Propensity	Socket Propensity Difference	α-helical Content (%)	Stability ΔG (kJ/mol)
Wild Type	WGERQAKAVADALTALESAMARIAKEL	270.03	0.000	27.8	5.98
Q5L	WGER**L**AKAVADALTALESAMARIAKEL	294.72	24.690	41.6	5.90
A6E	WGERQ**E**KAVADALTALESAMARIAKEL	258.41	-11.620	25.9	0.23
V9E	WGERQAKA**E**ADALTALESAMARIAKEL	265.38	-4.650	23.1	4.32
A10R	WGERQAKAV**R**DALTALESAMARIAKEL	248.21	-21.820	12.6	6.13
L13I	WGERQAKAVADA**I**TALESAMARIAKEL	237.72	-32.310	21.9	2.08
T14V	WGERQAKAVADAL**V**ALESAMARIAKEL	276.47	6.440	32.7	5.00
L16F	WGERQAKAVADALTA**F**ESAMARIAKEL	216.29	-53.740	18.5	2.65
E17D	WGERQAKAVADALTAL**D**SAMARIAKEL	258.59	-11.440	15.6	1.95
M20L	WGERQAKAVADALTALESA**L**ARIAKEL	327.4	57.370	55.7	8.56
A21E	WGERQAKAVADALTALESAM**E**RIAKEL	270.63	0.600	34.5	4.74
T14V/M20L	WGERQAKAVADAL**V**ALESA**L**ARIAKEL	333.84	7.040	61.0	6.70
T14V/A21E	WGERQAKAVADAL**V**ALESAM**E**RIAKEL	277.07	63.810	42.6	4.59

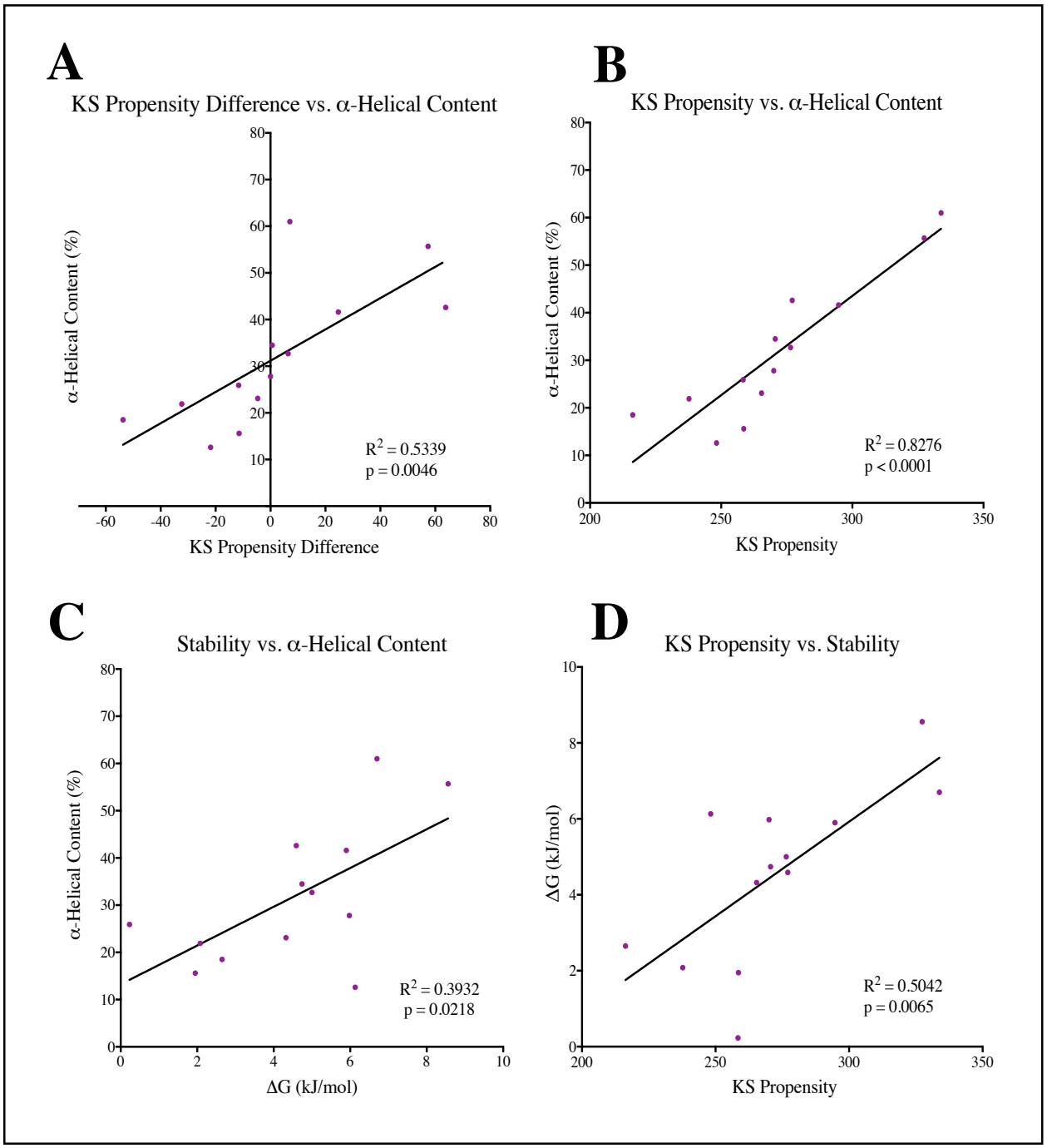


Figure 14. Correlation analysis. Pearson correlation analysis was conducted on the data collected for each of the KSα1.1 variants. The correlations include (A) KS Propensity Difference versus a-helical content, (B) Total KS Propensity versus a-helical content, (C) Stability versus a-helical content, and (D) KS Propensity versus stability. Correlations were calculated and graphed using GraphPad Prism 8.

Specificity of Quaternary Interactions in bZIP Coiled-Coils

Preliminary analysis of the selected bZIP sequences began with mapping onto KS α-helical lattices with free/filled propensity mapping. This method allows for visualization of free and filled sockets based on KS propensities and a colorimetric scale. In this color scheme, darker colors refer to sockets that strongly favor α-helical structure, while blue sockets prefer to be free or unpacked and red sockets prefer to be filled with a knob. This representation readily reveals the coiled-coil packing pattern of an α-helical sequence as well as identifies the knobs involved in packing. Propensity maps for cJun, p21SNFT, and CREB4 are shown in Figure 15. Upon further examination, Asn residues were identified within the leucine zipper regions of all bZIP proteins. Interestingly, as has been seen before, all Asn residues were located at an *a* position within a heptad repeat and disrupted the regular coiled-coil binding pattern with sockets with less of a propensity to bind knobs and even turning a few into free sockets. Because of this, the location and number of Asn residues was hypothesized to play a role in homodimer specificity. Asn residues are located at position a_3 for cJun and p21SNFT and positions a_3 and a_5 for CREB4. Based on this observation, point mutants of each bZIP protein were designed to investigate the role of the number and location of Asn residues in bZIP binding specificity.

Previous studies had shown that cJun (a_3) and p21SNFT (a_3) are able to homo- and heterodimerize with each other, while CREB4 ($a_{3,5}$) homodimerizes but does not interact with either of the two. With this in mind, point mutations were designed to change binding specificities. In cJun and p21SNFT, the a_5 position was mutated to an Asn, and in CREB4 the a_5 position was mutated from an Asn to a Leu. Propensity maps of the mutant versions of each protein were generated and compared to their respective wild type maps (Figure 16). Introduction of an Asn at the a_5 position in cJun and p21SNFT led to the presence of a light blue hexagon region within

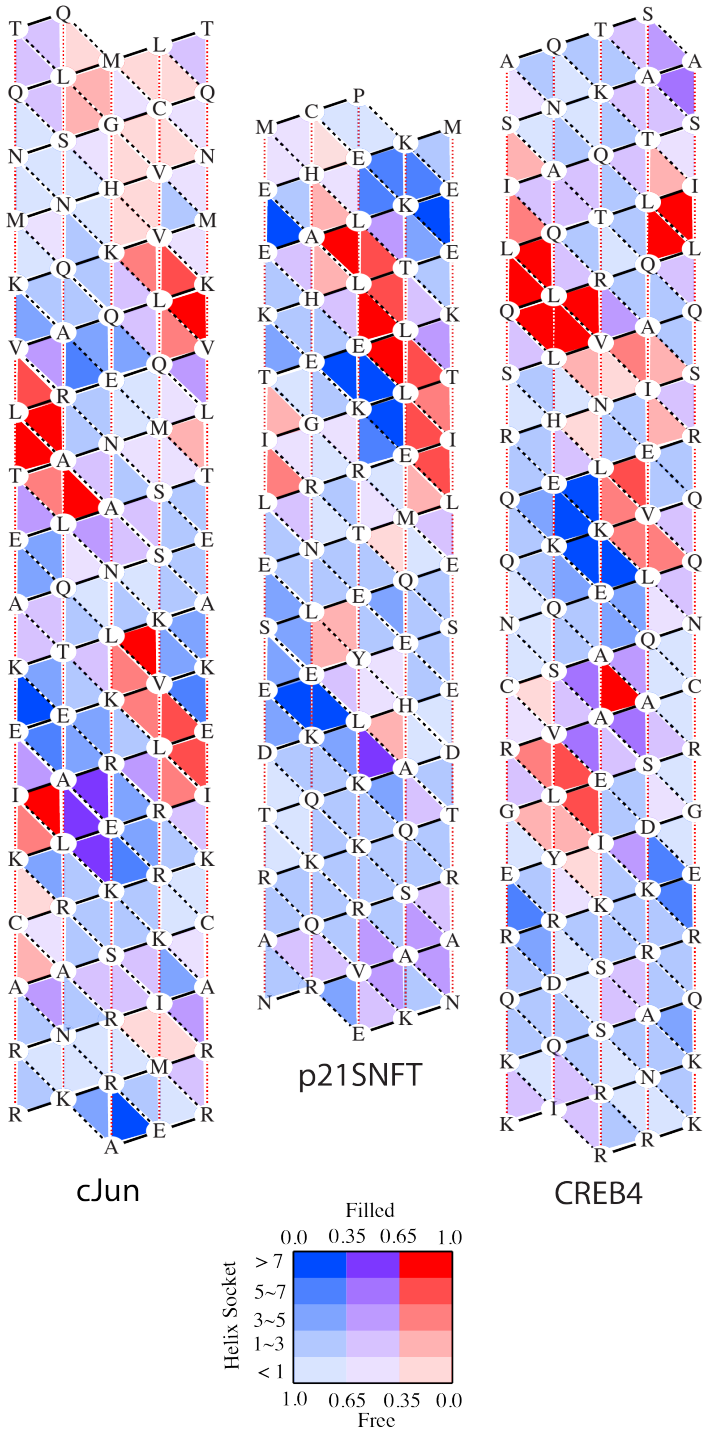


Figure 15. Propensity mapping of the bZIP wild type proteins. A colorimetric scheme was developed to serve as a visual representation of a sockets helical propensity and free/filled propensity simultaneously. A darker color represents a socket that greatly favors helical structure. The different colors represent a free propensity of 65% or greater (blue) or filled propensity of 65% or more (red). A socket that has free/filled propensities between 35-65% are colored purple.

the binding region of both proteins. Alternatively, the mutation of the a_5 Asn to a Leu in CREB4 led to a darker red binding region. Heptad mapping of the wild type and mutant variants is shown in Figure 17.

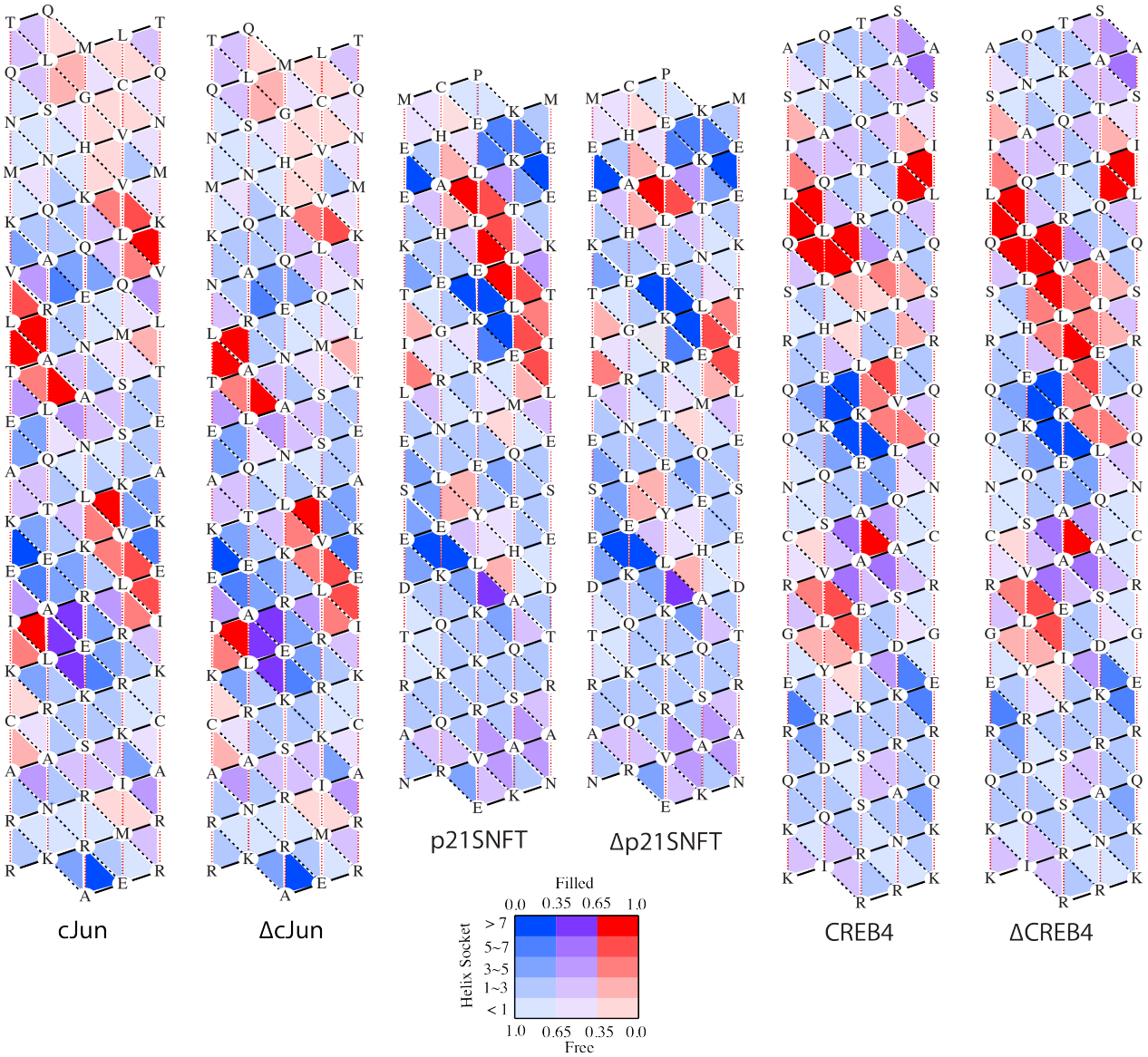


Figure 16. Propensity mapping of the bZIP wild type proteins versus corresponding point mutants.

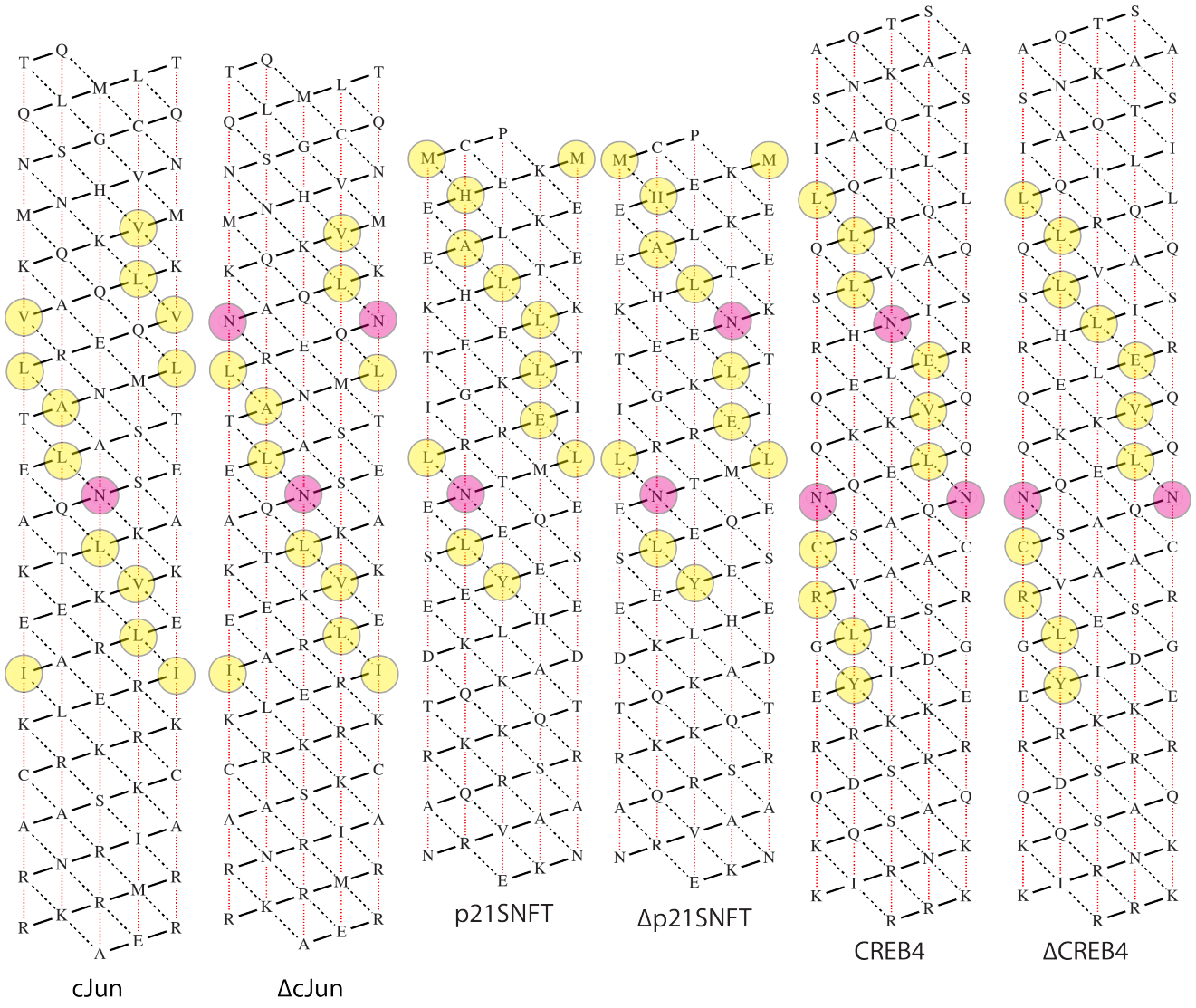


Figure 17. Heptad mapping of the bZIP wild type proteins and corresponding point mutants. Residues at positions *a* and *d* within the heptad repeats are highlighted by colored circles on the KS lattices for each protein. Asparagine residues at *a* positions are colored pink, while all other *a* and *d* residues are colored yellow.

To detect protein interaction between the wild type and mutant bZIP proteins, a bacterial adenylate cyclase two-hybrid (BACTH) assay was used. BACTH assay use two expression plasmids, each containing either the adenylate cyclase T18 subunit (pUT18C) or the T25 subunit (pKT25). Interaction between these two subunits results in a functional adenylate cyclase enzyme. Genes that are cloned into these expression vectors are expressed as fusion proteins, where a protein of interest is "tagged" with either the T18 or T25 subunit. Interaction can then be assessed by measuring adenylate cyclase function through either cAMP production or β-galactosidase activity in adenylate cyclase deficient bacterial cells (*cya*-). Expression plasmid construction and site-directed mutagenesis had to be done before interaction studies could be completed. Synthesized pET-24(a)+ plasmids containing wild type bZIP sequences, and purified pUT18C and pKT25 plasmids were used in subcloning procedures. Plasmids were checked via restriction digest and confirmed through sequencing. Representative maps of pUT18C and pKT25 containing the cJun sequence are shown in Figures 18-19.

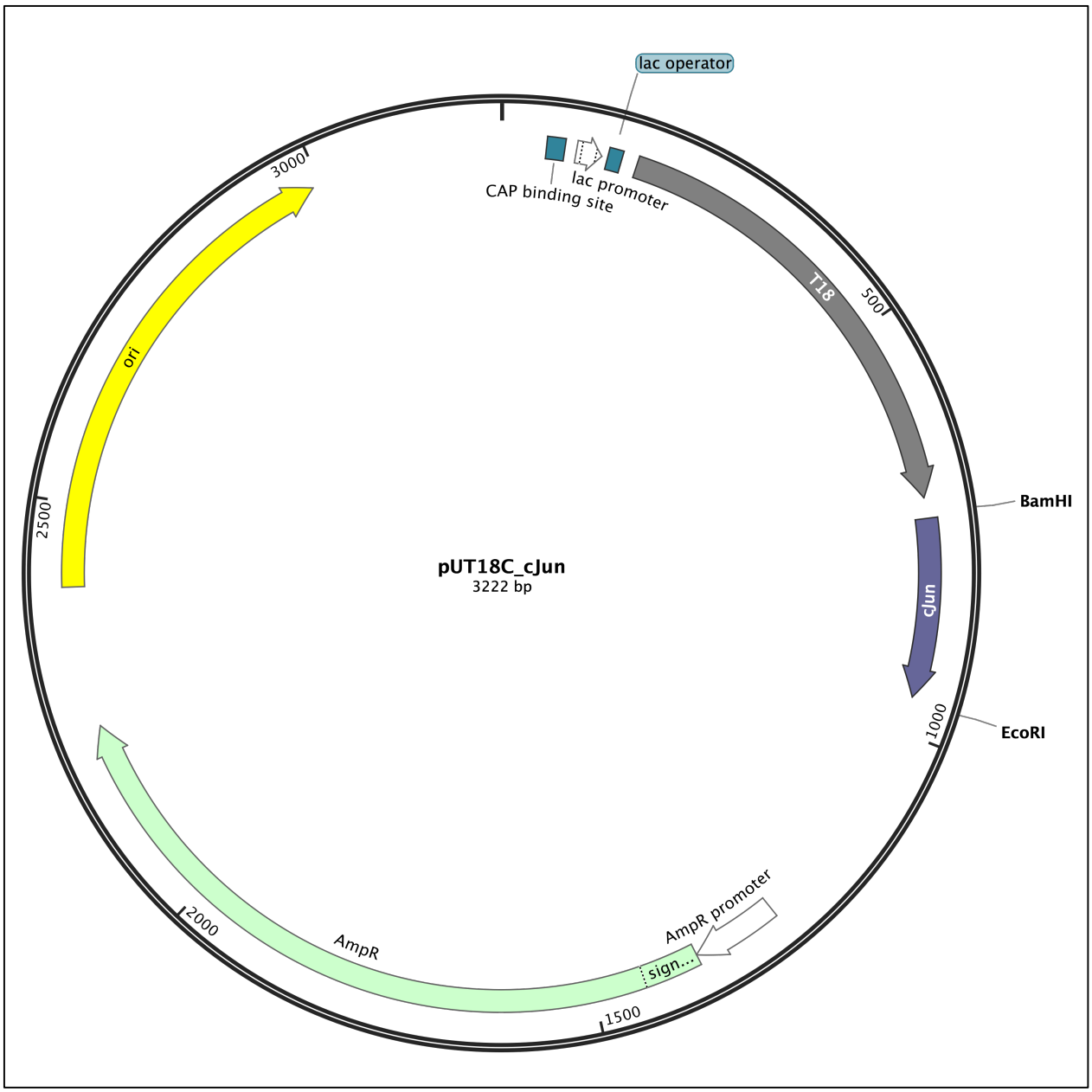


Figure 18. Plasmid map of pUT18C vector containing cJun insert. The sequence of the adenylate cyclase T18 subunit is located from 165-715 bp, upstream of the cJun insert located from 741-956 bp. The cJun insert was cloned using the 5' *Bam*HI restriction site and 3' *Eco*RI restriction site.

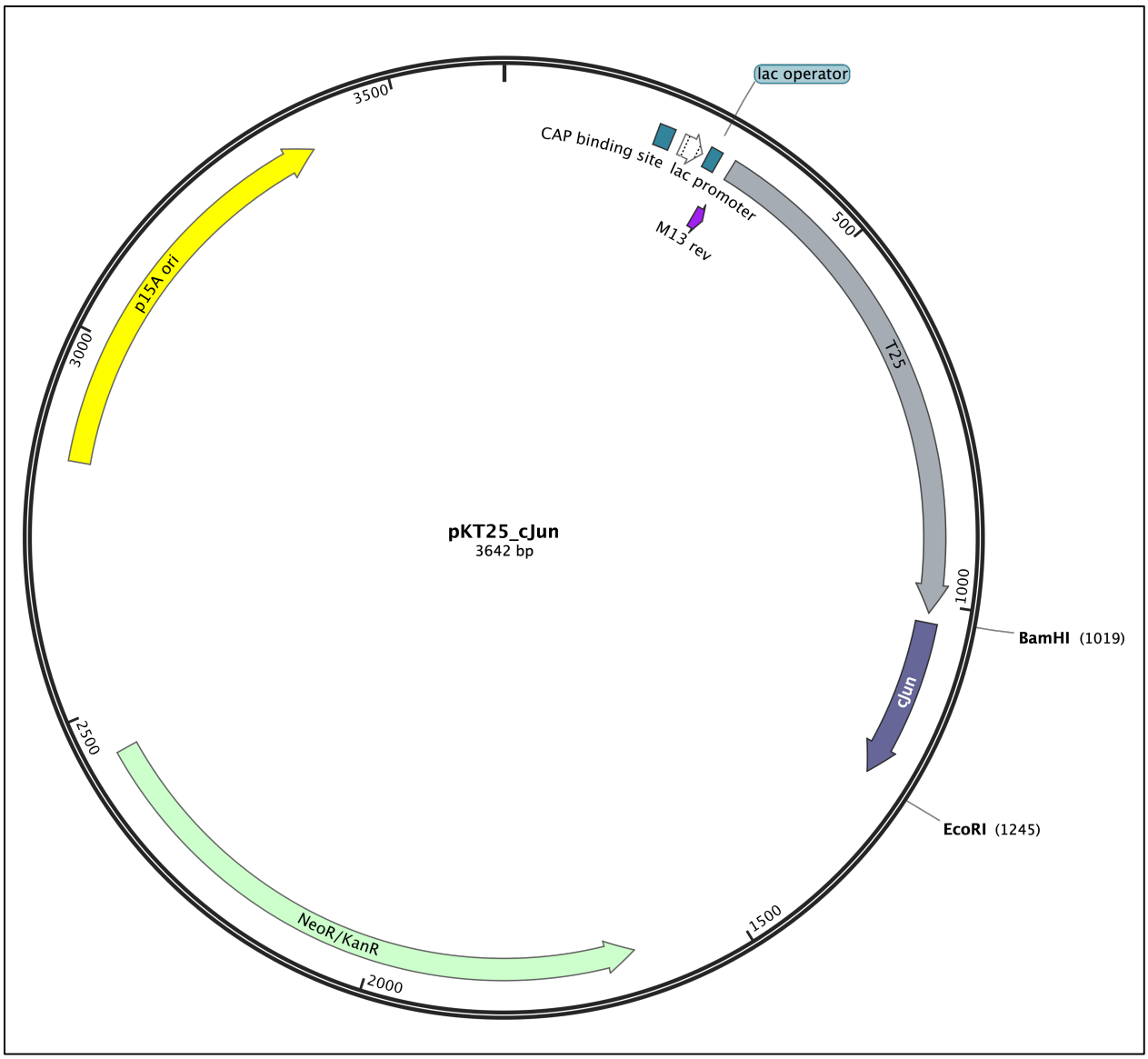


Figure 19. Plasmid map of pKT25 vector containing cJun insert. The sequence of the adenylate cyclase T25 subunit is located from 321-1012 bp, upstream of the cJun insert located from 1026-1241 bp. The cJun insert was cloned using the 5' *Bam*HI restriction site and 3' *Eco*RI restriction site.

After confirmation of the cloned pUT18C and pKT25 plasmids, preliminary interaction studies were done to ensure homo- and heterodimer formation matched previous results. BACTH assays were conducted using co-transformed BTH101 cells, and protein interaction was monitored through β-galactosidase activity. Miller units were calculated for each homo- or heterodimer and graphed using GraphPad Prism 8 (Figure 20). The data was analyzed using a One-Way ANOVA with multiple comparisons and Tukey's post hoc tests. Error bars were reported as standard deviations. In Figure 20A, the brown bar (PC) represents the positive control. This is the interaction between pUT18C-GCN4 and pKT25-GCN4 fusion proteins, which are expected to bind tightly. The burgundy bar (NC) represents the negative control, which is a result from the co-transformation of empty pUT18C and pKT25 plasmids. This was done to determine any baseline β-galactosidase activity or absorbance at 420 nm from the samples. The following three bars represent homodimer formation of cJun, p21SNFT (p21), and CREB4 (C4). One-Way ANOVA indicated that the positive control and all homodimers were statistically significant from the negative control, as indicated by the asterisks. The three final bars represent heterodimer formation between the three bZIP proteins. As expected, only cJun and p21SNFT were able to form heterodimers, while permutations with CREB4 were not statistically significant. Figure 20B shows a comparison of previous results gathered by Keating et al. to the BACTH results collected. This result was promising moving forward in the specificity investigation between these proteins.

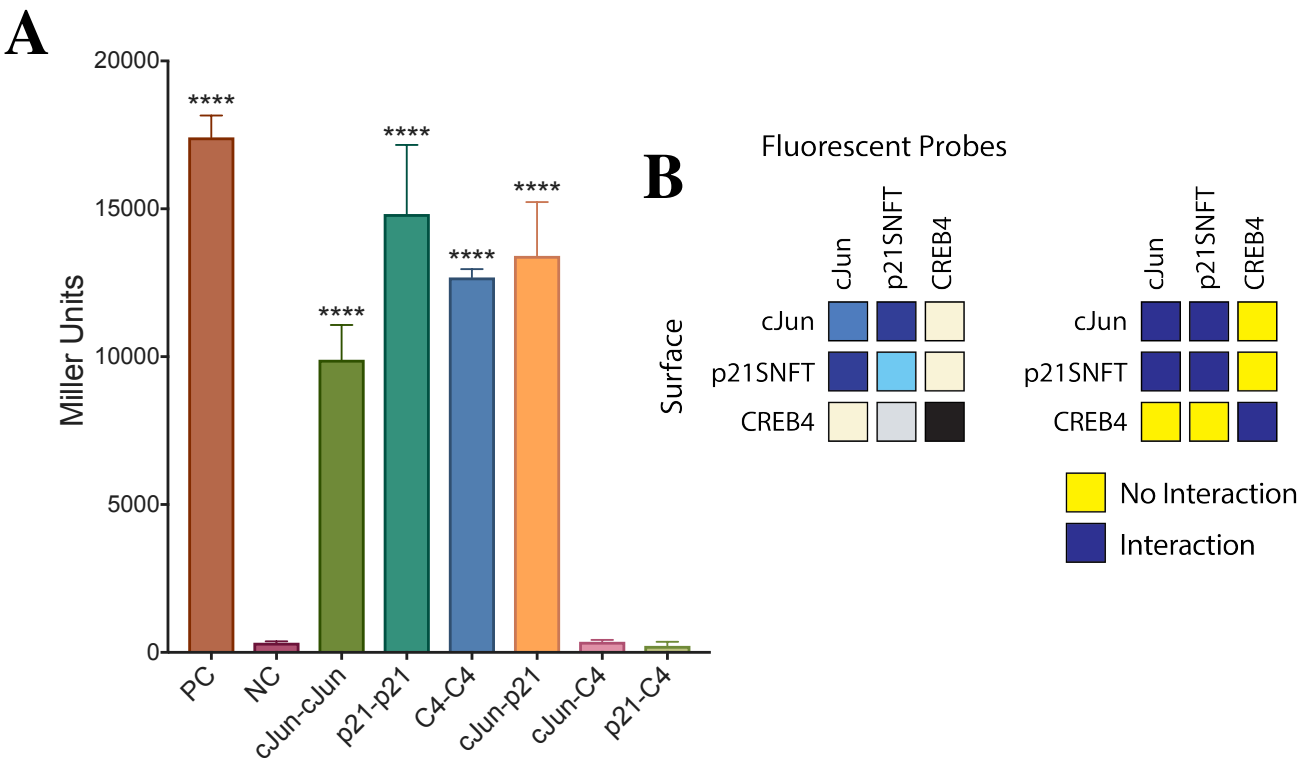


Figure 20. Preliminary BACTH results with wild type bZIP proteins. (A) β-galactosidase assays were conducted in triplicate for each interaction, and Miller units were calculated. Values were averaged and graphed using GraphPad Prism 8. Error bars are reported as standard deviations (**** $p < 0.0001$). (B) Previous interaction study results (left) are compared to the BACTH results (right) in a grid format. Both color schemes are similar in that darker/blue color represents an interaction, and ligther/yellow color represents little to no interaction.

Next, the point mutant versions of the bZIP proteins were used in the interaction studies. First, a panel with ΔCREB4 (ΔC4) dimer permutations was analyzed for changes in specificity. In Figure 21A, the first two bars are the positive and negative controls, and the following three are the wild type homodimer interactions. The remaining seven bars represent interactions with the ΔCREB4 mutant bZIP protein. Of particular note, homodimer formation was surprisingly

not observed for ΔCREB4, but heterodimer formation was measured between all wild type bZIP proteins. Interaction between ΔCREB4 and wild type CREB4 was not expected based on Asn placement and required further investigation.

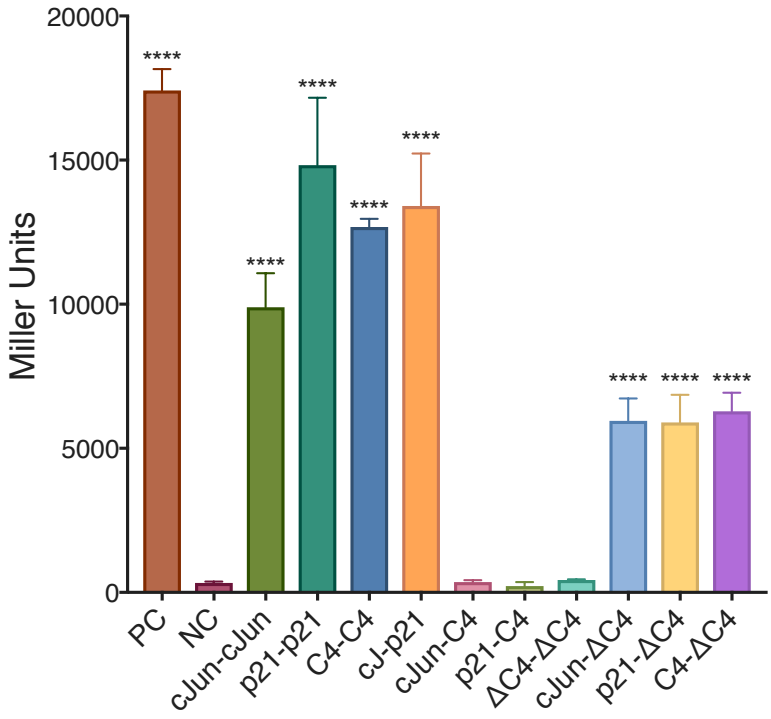


Figure 21. BACTH results with interactions involving ΔCREB4. (A) β-galactosidase assays were conducted in triplicate for each interaction, and Miller units were calculated. Values were averaged and graphed using GraphPad Prism 8. Error bars are reported as standard deviations (**** $p < 0.0001$).

Interactions involving mutant cJun (ΔcJun) and mutant p21SNFT (Δp21) were then investigated. In Figure 22, the six interactions following the positive and negative controls involve ΔcJun, where the remaining five permutations involve Δp21SNFT. This data shows that homodimer formation is possible for both ΔcJun and Δp21SNFT as well as heterodimer

formation between ΔcJun and Δp21SNFT, however, the mutant versions can no longer dimerize with either wild type protein. Additionally, both mutant proteins gained the ability to dimerize with wild type CREB4. Binding was not seen between ΔcJun/Δp21SNFT and ΔCREB4. A colorimetric summary of the BACTH results is shown in Figure 23.

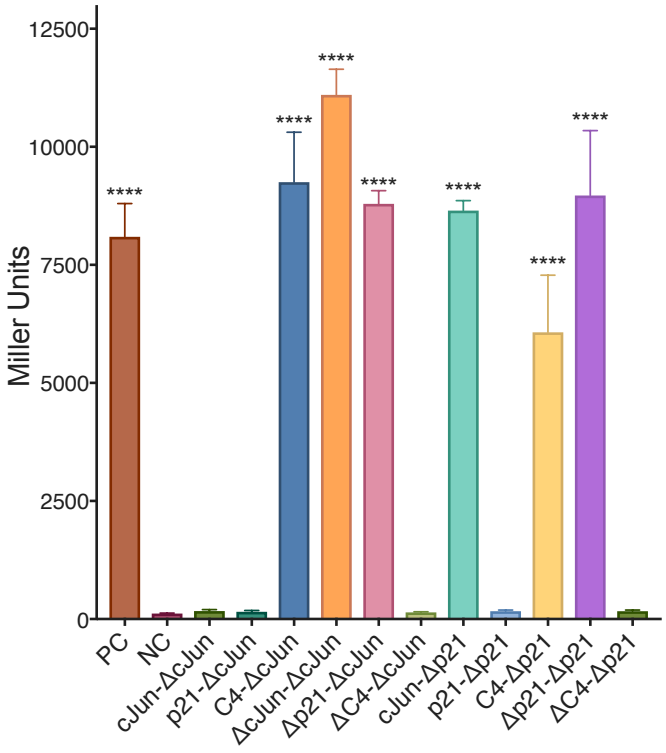


Figure 22. BACTH results with interactions involving ΔcJun and Δp21SNFT. (A) β-galactosidase assays were conducted in triplicate for each interaction, and Miller units were calculated. Values were averaged and graphed using GraphPad Prism 8. Error bars are reported as standard deviations (**** $p < 0.0001$).

Further KS lattice analysis was done to investigate the CREB4/ΔCREB4 interaction specificity. Here, we looked for residues that may disfavor packing of a Leu residue rather than an Asn in the binding region near the a_5 mutation site. The presence of a histidine (His, H) at

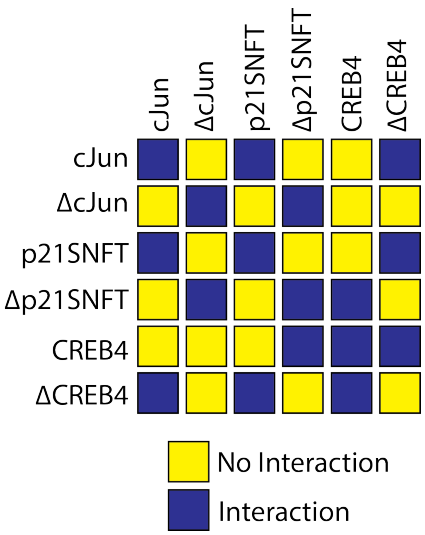


Figure 23. Summary of BACTH results with bZIP proteins and point mutant variants. Results were displayed on a grid to visualize proteins that did or did not interact. Proteins that did not interact are colored yellow, and proteins that did interact are colored blue.

position g_4 likely disfavors the packing of a hydrophobic Leu residue, and could therefore be the reason homodimer formation was not observed for the ΔCREB4 protein and heterodimer formation between CREB4 and ΔCREB4 was still possible. The b_5 position isoleucine (Ile, I) was also interrogated. A Leu followed by an Ile would not be sterically favorable when packing with a His residue, and would also be unfavorable due to the side chain properties.

To determine if these were the case, point mutants of ΔCREB4 were designed. Four g_4 (H48) point mutants were designed, which would mutate the His to a lysine (Lys, K), alanine (Ala, A), threonine (Thr, T), or glutamine (Gln, Q). A single b_5 (I50) point mutant was designed to mutate the Ile to an Ala. Heptad mapping and propensity mapping of ΔCREB4 is shown in Figure 24 and compared to the ΔCREB4 mutant propensity maps. Analysis of most ΔCREB4_H48 mutant maps showed little deviation from the ΔCREB4 map, with a few sockets

showing greater free propensities along the binding edge (dark blue sockets). The ΔCREB4_H48A hexagon propensities were changed from mostly free to mostly filled. ΔCREB4_I50A mutant map showed deviation from the ΔCREB4 map, where the I50 hexagon propensities were changed from mostly filled to mostly free. Site-directed mutagenesis was conducted on the pUT18C and pKT25 plasmids containing ΔCREB4 to generate a mutant library. Mutants were confirmed through sequencing and used in BACTH interaction studies.

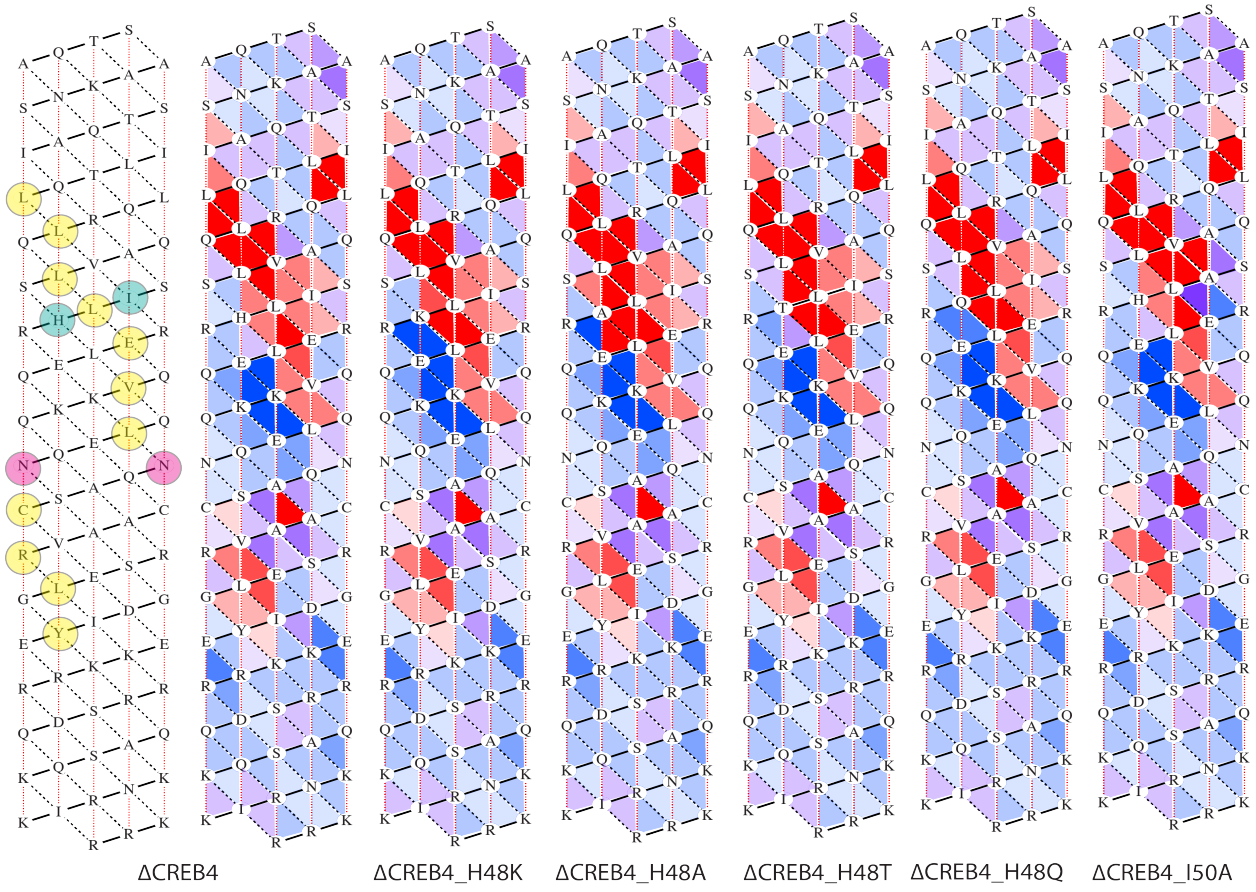


Figure 24. Heptad map of ΔCREB4 and propensity maps of ΔCREB4 point mutants. On the heptad map of ΔCREB4, residues at positions g_4 and b_5 are highlighted in teal. Propensity maps of ΔCREB4 g_4 and b_5 mutants are shown to compare to ΔCREB4.

Interactions between CREB4 and ΔCREB4 and all new ΔCREB4_H48 and ΔCREB4_I50 mutants were conducted and the results are shown in Figure 25. Following the positive and negative controls, the first five interactions involve ΔCREB4 and the last five interactions involve CREB4. The data shows that the new mutations not only allow for dimer formation with ΔCREB4, but maintain the ability to form dimers with CREB4 with the exception of ΔCREB4_H48Q. This was the only mutant that could dimerize with ΔCREB4 and not CREB4.

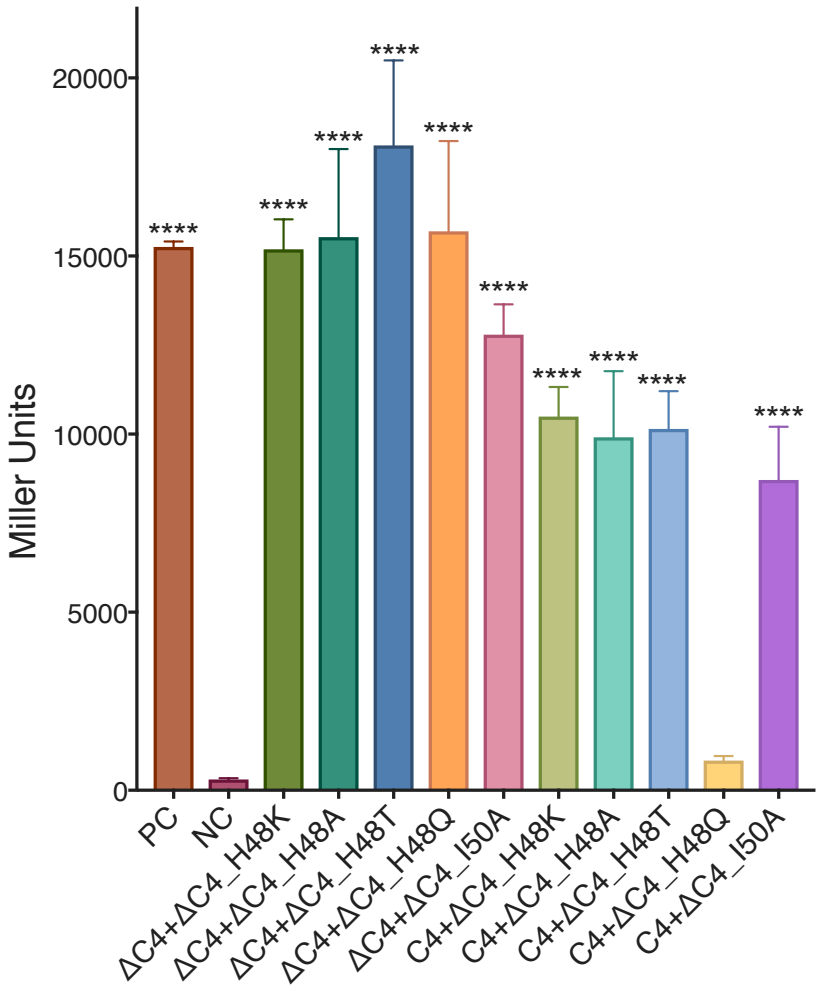


Figure 25. BACTH results of CREB4 and ΔCREB4 interactions with ΔCREB4 point mutants. β-galactosidase assays were conducted in triplicate for each interaction, and Miller units were calculated. Values were averaged and graphed using GraphPad Prism 8. Error bars are reported as standard deviations (**** $p < 0.0001$).

ΔCREB mutant homo- and heterodimers were also tested (Figure 26). In this figure, all interacting proteins are ΔCREB mutant variants. To simplify the labelling, the interactions are labeled by the residue that H48 was mutated to (K, A, T, Q) or that I50 was mutated to (IA). This data shows that every mutant has gained the ability to form a homodimer, where ΔCREB4 could not. However, the heterodimer data suggests a loss of specificity, where all mutants are also able to form heterodimers. A colorimetric summary of the ΔCREB4 mutant BACTH results are shown in Figure 27.

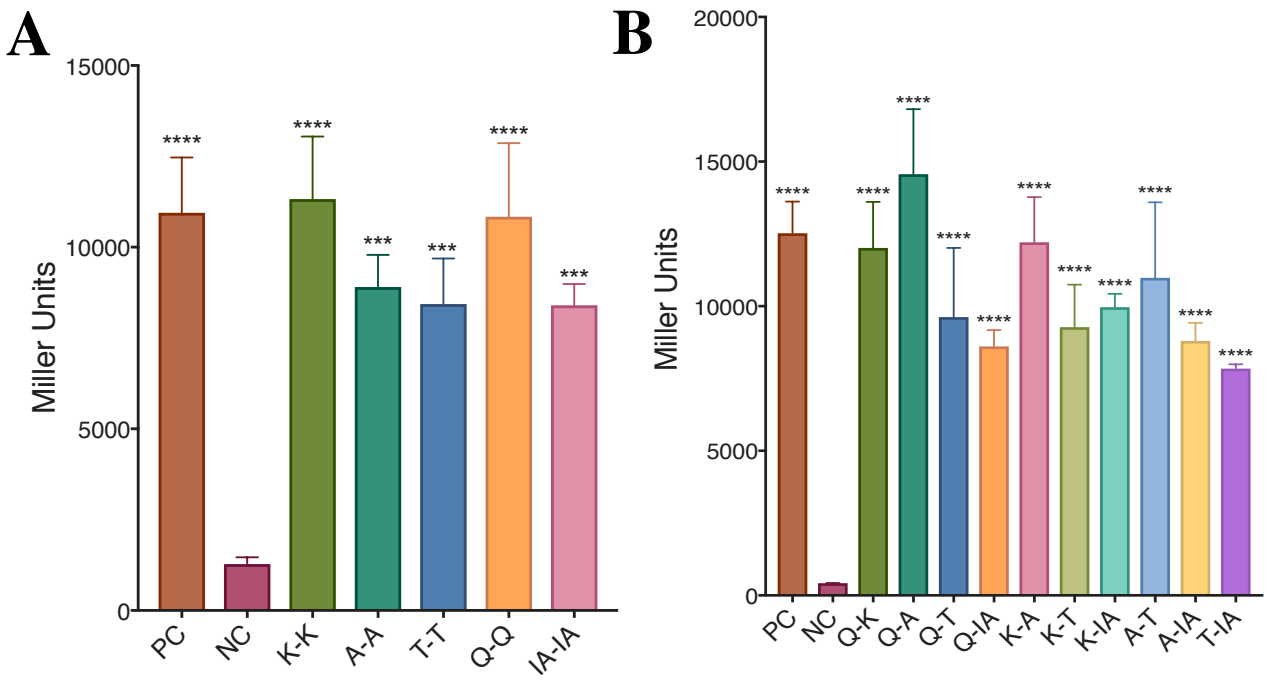


Figure 26. BACTH results of ΔCREB4 point mutant homo- and heterodimer interactions. β-galactosidase assays were conducted in triplicate for each interaction, and Miller units were calculated. Values were averaged and graphed using GraphPad Prism 8. Error bars are reported as standard deviations (**** $p < 0.0001$, *** $p < 0.0003$). (A) ΔCREB4 H48 and I50 point mutant homodimer interactions, and (B) ΔCREB4 H48 and I50 point mutant heterodimer interactions.

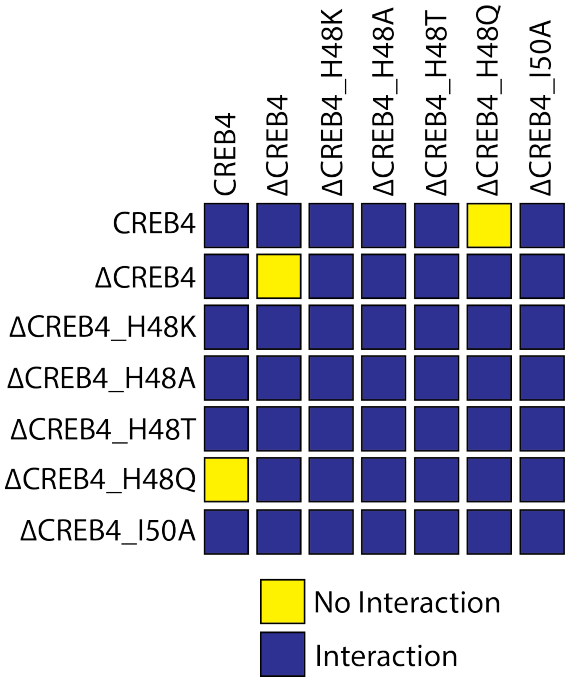


Figure 27. Summary of BACTH results with CREB4, ΔCREB4 and ΔCREB4 mutant variants. Results were displayed on a grid to visualize proteins that did or did not interact. Proteins that did not interact are colored yellow, and proteins that did interact are colored blue.

CHAPTER 4: DISCUSSION

Stability Relationship of Knob-Socket Propensities Measured in the KSα1.1 α-Helix

The relationship between a protein's amino acid sequence and resulting three-dimensional structure has been studied for the last several decades. Additionally, the way in which proteins interact with each other or the factors involved in binding specificity are of high interest. Understanding this relationship can provide insight into disease formation and proliferation, as well as unlock new methodologies for the development of therapeutics to combat these diseases.

Here, the Knob-Socket model was utilized in the design of a small peptide, and further used to predict how the structure and stability of the peptide would change upon mutation. Knob-Socket Hexagon values were compared between the wild type KSα1.1 protein and a set of single and double mutants. These values gave a preliminary prediction of the changes that were expected to occur based on the Knob-Socket Model propensity library alone. Based on the changes in propensity, four of the ten mutants (Q5L, T14V, M20L, and A21E) would form more stable helices, and the other six (A6E, V9E, A10R, L13I, L16F, and E17D) were predicted to be less stable in terms of Socket Propensity Differences relative to wild type. After expression and purification, circular dichroism was measured from 260-190 nm for each 200 μM protein sample. The signatures for each protein exhibited expected alpha-helical character, with minima around 222 and 208 nm, and a maximum around 195 nm. While most signatures remained close in terms of intensities and shape, some did show a much larger absorption at 222 and 208 nm (M20L, Q5L). However, the signatures alone were not enough to deduce the change in alpha-helical content between the proteins. Because of this, deconvolution of the raw CD data was conducted using DichroWeb to quantitate the changes in alpha-helical content. Deconvolution

revealed that the wild type protein contained 27.8% alpha-helical structure, and that point mutations in the short peptide sequence dramatically affect that value. The extreme cases were a decrease in helical structure to 15.6% (E17D), and an increase in helical structure to 55.7% (M20L). Interestingly, all four of the predicted stabilizing mutants did exhibit an increase in alpha-helical structure, while the six destabilizing mutants exhibited a decrease in the percentage of alpha-helical structure. This suggests that the KS model does have the ability to predict how a particular mutation will affect the overall structure and stability of a protein. To further examine the relationship between KS model prediction and quantitative experimental values, correlation analysis was conducted to determine if there was any statistical significance in the relationship between these two parameters. While KS Propensity Differences correlated moderately well with alpha-helical percentage (R^2=0.5339), total KS Propensity correlated with alpha-helical percentage much better (R^2=0.8276).

Two double point mutants (T14V/M20L and T14V/A21E) were also designed and characterized. Because of the non-overlapping nature of the KS Hexagons involved in these double mutations, the effects of the mutations were predicted to be additive, meaning that the double mutant should exhibit the changes seen by both corresponding single mutants. Full spectrum circular dichroism analysis showed that the signatures were greater in intensity than the corresponding point mutants. For example, the T14V/M20L curve was greater in intensity than both T14V and M20L. Again, deconvolution was done to determine the change in alpha-helical content quantitatively. The double mutant percentages were found to be greater than both of the corresponding point mutant percentages. More interestingly, the sum of the percent changes seen by the single mutants along with the base wild type percentage was almost exactly the

percentage calculated for the double mutants, supporting the idea that these changes would be additive in nature.

All KSα1.1 mutants were subjected to chemical and thermal denaturation experiments to investigate the changes in protein stability. Most mutants were seen to have almost linear denaturation curves in both experiments, except for the mutants M20L and T14V/M20L. These mutants exhibited denaturation curves that appeared to have more defined "folded" and "unfolded" regions. This is likely due in part to the fact that these are the only two proteins examined that contain more than 50% alpha-helical (folded) content, while the others are largely random coil (unfolded). Stability (ΔG) was still calculated for each mutant and compared to wild type. The relative change in stability did not seem to follow the same pattern as predicted with the alpha-helical structure changes. M20L, T14V/M20L, as well as the A10R mutant were found to have an increased ΔG value from wild type (8.56 kJ/mol, 6.70 kJ/mol, and 6.13 kJ/mol respectively). All others were found to have a ΔG less than the wild type. These stability values were also used in correlation analyses between alpha-helical content and total KS propensity. Both stability versus alpha-helical content and KS propensity versus stability still correlated moderately well despite the deviance from the initial predictions (R^2=0.3932, R^2=0.5042 respectively). This suggests that the KS model can partially predict the change in stability based on amino acid sequence, but there are definitely other factors that could strengthen the KS prediction method.

Specificity of Quaternary Interactions in bZIP Coiled-Coils

In this project, we attempted to uncover factors involved in bZIP coiled-coil protein binding specificity. Initial sequence analysis of three bZIP proteins (cJun, p21SNFT, CREB4) revealed the presence of Asn residues at *a* positions within the leucine zipper heptad repeats. Interestingly, the spacing/location and the number of Asn residues seemed to change between different bZIP family members. For example, Asn residues were located at position a_3 for cJun and p21SNFT and a_3/a_5 for CREB4. The presence of an Asn within the protein binding region created a KS hexagon of primarily free sockets seen in the propensity lattice mapping. This residue was then hypothesized to be a crucial factor in the binding specificity for these bZIP proteins, and that a patch of free sockets along the binding region must play a role in coiled-coil recognition. This was investigated using a bacterial adenylate cyclase two-hybrid (BACTH) assay. Preliminary homodimer and heterodimer interactions indicated that homodimers were able to form among all bZIP proteins, while only the cJun-p21SNFT heterodimer interacted. This data matched previous interaction results from a study done by Keating et al, which served as a promising starting point for the point mutant experiments. Because of this, point mutants of each bZIP protein were designed and constructed for analysis. The mutants were either an Asn mutants at position a_5 (cJun, p21SNFT), or a leucine mutant at position a_5 (CREB4). With this initial set of bZIP constructs, BACTH assays were carried out to determine which could interact with the changes in their sequences.

ΔcJun (Va_5N mutant) was shown to lose the ability to dimerize with both cJun and p21SNFT proteins, but gained the ability to dimerize with CREB4. Further, ΔcJun was able to form a homodimer as well as a dimer with Δp21SNFT. Δp21SNFT showed the same result, where dimer formation was no longer detectable with either cJun or p21SNFT, but was

detectable with CREB4. This data suggests that the introduction of an Asn was able to guide dimer formation with CREB4, yet seemed to block dimer formation with bZIP proteins that did not have an Asn at the a_5 position. ΔCREB4 (Na_5L mutant) was shown to gain the ability to form heterodimers with cJun and p21SNFT. This data suggests that the Asn did in fact play a role in blocking dimer formation with these two proteins. However, ΔCREB4 was also shown to lose the ability to form a homodimer yet retained the ability to heterodimerize with CREB4. After further sequence analysis of the CREB4 sequence, this unexpected result was mostly attributed to the presence of a histidine residue at the g_4 position within the helix. The b_5 position was also looked at, to determine if the extremely hydrophobic LI:V socket could be a factor in blocking binding as well.

This idea was tested by generating ΔCREB4 point mutants. These included four g_4 mutants (Hg_4K, Hg_4A, Hg_4T, Hg_4Q) and one b_5 mutant (Ib_5A). First, behavior between the ΔCREB point mutants and CREB4 and ΔCREB4 was investigated. Before the point mutations were introduced, ΔCREB4 could not form homodimers but could form heterodimers with CREB4. All ΔCREB point mutants gained the ability to dimerize with ΔCREB4, and most retained the ability to dimerize with CREB4 with the exception of the Hg_4Q mutant. Further homodimer and heterodimer analysis with the ΔCREB4 point mutants revealed that all possible homo- and heterodimers were able to form, when ΔCREB4 itself could not. Because of the interesting interaction with the ΔCREB4 Hg_4Q mutant, further investigation of that sequence should be conducted.

REFERENCES

1. Joo, H., et al., *An amino acid packing code for α-helical structure and protein design*. J Mol Biol, 2012. **419**(3-4): p. 234-254.
2. Joo, H. and J. Tsai, *An amino acid code for β-sheet packing structure*. Proteins, 2014. **82**(9): p. 2128-2140.
3. Mackenzie, C.O. and G. Grigoryan, *Protein structural motifs in prediction and design*. Curr Opin Struct Biol, 2017. **44**: p. 161-167.
4. Sledz, P. and A. Caflisch, *Protein structure-based drug design: from docking to molecular dynamics*. Curr Opin Struct Biol, 2018. **48**: p. 93-102.
5. Joo, H., et al., *An Amino Acid Code for Irregular and Mixed Secondary Structure Packing*. Proteins: Struct. Funct. & Bioinform., 2015. **(accepted)**.
6. Finkelstein, A.V., *50+ Years of Protein Folding*. Biochemistry (Mosc), 2018. **83**(Suppl 1): p. S3-S18.
7. Koepnick, B., et al., *De novo protein design by citizen scientists*. Nature, 2019. **570**(7761): p. 390-394.
8. Ovchinnikov, S., et al., *Protein structure prediction using Rosetta in CASP12*. Proteins, 2018. **86 Suppl 1**: p. 113-121.
9. Gillet, J.N. and I. Ghosh, *Concepts on the protein folding problem*. J Biomol Struct Dyn, 2013. **31**(9): p. 1020-3.
10. Dill, K.A. and J.L. MacCallum, *The protein-folding problem, 50 years on*. Science, 2012. **338**(6110): p. 1042-6.
11. Dill, K.A., et al., *The protein folding problem*. Annu Rev Biophys, 2008. **37**: p. 289-316.
12. Dill, K.A., et al., *The protein folding problem: when will it be solved?* Curr Opin Struct Biol, 2007. **17**(3): p. 342-6.
13. Joshi, R.R., *Protein folding: interplay of hydrophobic-hydrophilic forces?* J Biomol Struct Dyn, 2013. **31**(9): p. 965-6.
14. Nolting, B., N. Salimi, and U. Guth, *Protein folding forces*. J Theor Biol, 2008. **251**(2): p. 331-47.
15. Dill, K.A., *Dominant Forces in Protein Folding*. Biochemistry, 1990. **29**(31): p. 7133-7155.
16. Cheng, J., et al., *Estimation of model accuracy in CASP13*. Proteins, 2019.
17. Cong, Q., et al., *Protein interaction networks revealed by proteome coevolution*. Science, 2019. **365**(6449): p. 185-189.
18. Kandathil, S.M., J.G. Greener, and D.T. Jones, *Prediction of inter-residue contacts with DeepMetaPSICOV in CASP13*. Proteins, 2019.
19. Rocklin, G.J., et al., *Global analysis of protein folding using massively parallel design, synthesis, and testing*. Science, 2017. **357**(6347): p. 168-175.
20. Basak, S., et al., *Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed betaalpha protein*. Proc Natl Acad Sci U S A, 2019. **116**(14): p. 6806-6811.
21. Burley, S.K., et al., *Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive*. Methods Mol Biol, 2017. **1607**: p. 627-641.
22. Joo, H., et al., *An amino acid code for irregular and mixed protein packing*. Proteins, 2015. **83**(12): p. 2147–2161.

23. Fraga, K.J., H. Joo, and J. Tsai, *An amino acid code to define a protein's tertiary packing surface*. Proteins, 2016. **84**(2): p. 201-16.
24. Linderstrom-Lang, K., *Proteins and enzymes*. Lane Medical Lectures. Vol. VI. 1952: Stanford University Publications, University Series.
25. Linderstrom-Lang, K. and J.A. Schellman, *Protein structure and enzyme activity*, in *The Enzymes*, P.D. Boyer, H. Lardy, and K. Myrbäck, Editors. 1959, Academic Press: New York. p. 443-510.
26. Pauling, L., *Chapter 7 Interatomic Distances and Their Relationship to Molecules and Crystals*, in *The Nature of the Chemical Bond*. 1960, Cornell University Press: Ithaca, New York. p. 221-257.
27. Pauling, L., *The shared-electron chemical bond*. Proc Natl Acad Sci U S A, 1928. **14**(4): p. 359-362.
28. Ramachandran, G.N., et al., *The mean geometry of the peptide unit from crystal structure data*. Biochim Biophys Acta, 1974. **359**(2): p. 298-302.
29. Chothia, C., *Hydrophobic bonding and accessible surface area in proteins*. Nature, 1974. **248**: p. 338-339.
30. Kauzmann, W., *Denaturation of proteins and enzymes*, in *The machanism of enzyme action*, W.D. McElroy and B. Glass, Editors. 1954, Johns Hopkins University Press: Baltimore. p. 70-120.
31. Tanford, C., *Contribution of hydrophobic interactions to the stability of the globular conformation of proteins*. J Am Chem Soc, 1962. **84**(22): p. 4240-4247.
32. Tanford, C., *The hydrophobic effect: formation of micelles and biological membranes*. 2nd ed. 1980, New York: John Wiley & Sons.
33. Anfinsen, C. and E. Haber, *Studies on the Reduction and Re-formation of Protein Disulfide Bonds*. J. Biol. Chem., 1960. **236**: p. 1361-1363.
34. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(96): p. 223-30.
35. Perutz, M.F., et al., *Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 A resolution: the atomic model*. Nature, 1968. **219**: p. 131-139.
36. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis*. Nature, 1958. **181**(4610): p. 662-6.
37. Kendrew, J.C., et al., *Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution*. Nature, 1960. **185**(4711): p. 422-7.
38. Pauling, L. and R.B. Corey, *Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets*. Proc Natl Acad Sci U S A, 1951. **37**: p. 729-740.
39. Pauling, L. and R.B. Corey, *The pleated sheet, a new layer configuration of polypeptide chains*. Proc Natl Acad Sci U S A, 1951. **37**(5): p. 251-256.
40. Pauling, L., R.B. Corey, and H.R. Branson, *The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain*. Proc Natl Acad Sci U S A, 1951. **37**: p. 205-211.
41. Pauling, L., R.B. Corey, and H.R. Branson, *The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain*. Proc Natl Acad Sci U S A, 1951. **37**(4): p. 205-11.
42. Salemme, F.R., *Structural properties of protein beta-sheets*. Prog Biophys Mol Biol, 1983. **42**(2-3): p. 95-133.

43. Baldwin, R.L., *Alpha-helix formation by peptides of defined sequence*. Biophys Chem, 1995. **55**(1-2): p. 127-35.
44. Baldwin, R.L. and G.D. Rose, *Molten globules, entropy-driven conformational change and protein folding*. Curr Opin Struct Biol, 2013. **23**(1): p. 4-10.
45. Kang, T.S. and R.M. Kini, *Structural determinants of protein folding*. Cell Mol Life Sci, 2009. **66**(14): p. 2341-61.
46. Rost, B., *Review: protein secondary structure prediction continues to rise*. J Struct Biol, 2001. **134**(2-3): p. 204-18.
47. Oldfield, C.J., K. Chen, and L. Kurgan, *Computational Prediction of Secondary and Supersecondary Structures from Protein Sequences*. Methods Mol Biol, 2019. **1958**: p. 73-100.
48. Tanford, C., *How protein chemists learned about the hydrophobic factor*. Protein Sci, 1997. **6**(6): p. 1358-66.
49. Kauzmann, W., *Some factors in the interpretation of protein denaturation*. Adv. Prot. Chem., 1959. **14**: p. 1-63.
50. Sarai, A., et al., *Thermodynamic databases for proteins and protein-nucleic acid interactions*. Biopolymers, 2001. **61**(2): p. 121-6.
51. Gromiha, M.M., et al., *ProTherm, Thermodynamic Database for Proteins and Mutants: developments in version 3.0*. Nucleic Acids Res, 2002. **30**(1): p. 301-2.
52. Baldwin, R.L. and G.D. Rose, *Is protein folding hierarchic? II. Folding intermediates and transition states*. Trends Biochem Sci, 1999. **24**(2): p. 77-83.
53. Vassilenko, K.S. and V.N. Uversky, *Native-like secondary structure of molten globules*. Biochim Biophys Acta, 2002. **1594**(1): p. 168-77.
54. Rami, B.R. and J.B. Udgaonkar, *Mechanism of formation of a productive molten globule form of barstar*. Biochemistry, 2002. **41**(6): p. 1710-6.
55. AlQuraishi, M., *AlphaFold at CASP13*. Bioinformatics, 2019.
56. Park, H., et al., *High-accuracy refinement using Rosetta in CASP13*. Proteins, 2019.
57. Creighton, T.E., *Protein structure. Stability of alpha-helices*. Nature, 1987. **326**(6113): p. 547-8.
58. Janin, J., R.P. Bahadur, and P. Chakrabarti, *Protein-protein interaction and quaternary structure*. Q Rev Biophys, 2008. **41**(2): p. 133-80.
59. Lupas, A.N., J. Bassler, and S. Dunin-Horkawicz, *The Structure and Topology of alpha-Helical Coiled Coils*. Subcell Biochem, 2017. **82**: p. 95-129.
60. Verkhivker, G.M., et al., *Computational detection of the binding-site hot spot at the remodeled human growth hormone-receptor interface*. Proteins, 2003. **53**(2): p. 201-19.
61. DeLano, W.L., *Unraveling hot spots in binding interfaces: progress and challenges*. Curr Opin Struct Biol, 2002. **12**(1): p. 14-20.
62. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces*. J Mol Biol, 1998. **280**(1): p. 1-9.
63. Smits, A.H. and M. Vermeulen, *Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities*. Trends Biotechnol, 2016. **34**(10): p. 825-834.
64. Zhang, M., et al., *Application of Machine Learning Approaches for Protein-protein Interactions Prediction*. Med Chem, 2017. **13**(6): p. 506-514.
65. Gromiha, M.M., K. Yugandhar, and S. Jemimah, *Protein-protein interactions: scoring schemes and binding affinity*. Curr Opin Struct Biol, 2017. **44**: p. 31-38.

66. Mariani, V., et al., *Assessment of template based protein structure predictions in CASP9*. Proteins, 2011. **79 Suppl 10**: p. 37-58.
67. Kinch, L., et al., *CASP9 assessment of free modeling target predictions*. Proteins, 2011. **79 Suppl 10**: p. 59-73.
68. Zimm, B.H. and J.K. Bragg, *Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains*. J. Chem. Phys., 1959. **31**: p. 526-531.
69. Lifson, S. and A. Roig, *On the theory of helix-coil transition in polypeptides*. J. Chem. Phys., 1961. **34**: p. 1963-1971.
70. Zimm, B.B.J., *Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains*. Journal of Chemical Physics, 1959. **31**: p. 526-531.
71. Chakrabartty, A. and R.L. Baldwin, *Stability of alpha-helices*. Adv Protein Chem, 1995. **46**: p. 141-76.
72. Lifson, S., Roig, A. , *On the theory of helix-coil transition in polypeptides*. Journal of Chemical Physics, 1961. **34**(6): p. 1963-1974.
73. Shoemaker, K.R., et al., *Tests of the helix dipole model for stabilization of alpha-helices*. Nature, 1987. **326**(6113): p. 563-7.
74. Shoemaker, K.R., et al., *Nature of the charged-group effect on the stability of the C-peptide helix*. Proc Natl Acad Sci U S A, 1985. **82**(8): p. 2349-53.
75. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides*. Proc Natl Acad Sci U S A, 1989. **86**(14): p. 5286-90.
76. Rohl, C.A. and R.L. Baldwin, *Deciphering rules of helix stability in peptides*. Methods Enzymol, 1998. **295**: p. 1-26.
77. Rohl, C.A., A. Chakrabartty, and R.L. Baldwin, *Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol*. Protein Sci, 1996. **5**(12): p. 2623-37.
78. Doig, A.J. and R.L. Baldwin, *N- and C-capping preferences for all 20 amino acids in alpha-helical peptides*. Protein Sci, 1995. **4**(7): p. 1325-36.
79. Chakrabartty, A., T. Kortemme, and R.L. Baldwin, *Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions*. Protein Sci, 1994. **3**(5): p. 843-52.
80. Doig, A.J., et al., *Determination of free energies of N-capping in alpha-helices by modification of the Lifson-Roig helix-coil therapy to include N- and C-capping*. Biochemistry, 1994. **33**(11): p. 3396-403.
81. Scholtz, J.M., et al., *The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide*. Biochemistry, 1993. **32**(37): p. 9668-76.
82. Scholtz, J.M., et al., *Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water*. Biopolymers, 1991. **31**(13): p. 1463-70.
83. Lyu, P.C., P.J. Gans, and N.R. Kallenbach, *Energetic contribution of solvent-exposed ion pairs to alpha-helix structure*. J Mol Biol, 1992. **223**(1): p. 343-50.
84. Gans, P.J., et al., *The helix-coil transition in heterogeneous peptides with specific side-chain interactions: theory and comparison with CD spectral data*. Biopolymers, 1991. **31**(13): p. 1605-14.
85. Vasquez, M. and H.A. Scheraga, *Effect of sequence-specific interactions on the stability of helical conformations in polypeptides*. Biopolymers, 1988. **27**(1): p. 41-58.

86. Vasquez, M., M.R. Pincus, and H.A. Scheraga, *Helix-coil transition theory including long-range electrostatic interactions: application to globular proteins*. Biopolymers, 1987. **26**(3): p. 351-71.
87. Wójcik, J., K.-H. Altmann, and H.A. Scheraga, *Helix-coil stability constants for the naturally occurring amino acids in water. XXIV. Half-cystine parameters from random poly(hydroxybutylglutamine-CO-S-methylthio-L-cysteine)*. Biopolymers, 1990. **30**(1-2): p. 121-134.
88. Lyu, P.C., et al., *Side chain contributions to the stability of alpha-helical structure in peptides*. Science, 1990. **250**(4981): p. 669-73.
89. Park, S.H., W. Shalongo, and E. Stellwagen, *Residue helix parameters obtained from dichroic analysis of peptides of defined sequence*. Biochemistry, 1993. **32**(27): p. 7048-7053.
90. O'Neil, K.T. and W.F. DeGrado, *A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids*. Science, 1990. **250**(4981): p. 646-51.
91. Munoz, V. and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence*. J Mol Biol, 1995. **245**(3): p. 297-308.
92. Munoz, V. and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides*. J Mol Biol, 1995. **245**(3): p. 275-96.
93. Ernst, A., et al., *Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing*. Mol Biosyst, 2010. **6**(10): p. 1782-90.
94. Hietpas, R.T., J.D. Jensen, and D.N. Bolon, *Experimental illumination of a fitness landscape*. Proc Natl Acad Sci U S A, 2011. **108**(19): p. 7896-901.
95. Fowler, D.M., et al., *High-resolution mapping of protein sequence-function relationships*. Nat Methods, 2010. **7**(9): p. 741-6.
96. Reich, L.L., S. Dutta, and A.E. Keating, *SORTCERY-A High-Throughput Method to Affinity Rank Peptide Ligands*. J Mol Biol, 2015. **427**(11): p. 2135-50.
97. Magliery, T.J., *Protein stability: computation, sequence statistics, and new experimental methods*. Curr Opin Struct Biol, 2015. **33**: p. 161-8.
98. Yin, S., F. Ding, and N.V. Dokholyan, *Eris: an automated estimator of protein stability*. Nat Methods, 2007. **4**(6): p. 466-7.
99. Schymkowitz, J., et al., *The FoldX web server: an online force field*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W382-8.
100. Das, R. and D. Baker, *Macromolecular modeling with rosetta*. Annu Rev Biochem, 2008. **77**: p. 363-82.
101. Benedix, A., et al., *Predicting free energy changes using structural ensembles*. Nat Methods, 2009. **6**(1): p. 3-4.
102. Potapov, V., M. Cohen, and G. Schreiber, *Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details*. Protein Eng Des Sel, 2009. **22**(9): p. 553-60.
103. Minard, P., *[Directed evolution of proteins]*. Med Sci (Paris), 2019. **35**(2): p. 169-175.
104. Ghaemmaghami, S., M.C. Fitzgerald, and T.G. Oas, *A quantitative, high-throughput screen for protein stability*. Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8296-301.
105. Ghaemmaghami, S. and T.G. Oas, *Quantitative protein stability measurement in vivo*. Nat Struct Biol, 2001. **8**(10): p. 879-82.

106. Dutta, S., A. Koide, and S. Koide, *High-throughput analysis of the protein sequence-stability landscape using a quantitative yeast surface two-hybrid system and fragment reconstitution*. J Mol Biol, 2008. **382**(3): p. 721-33.
107. Magliery, T.J., J.J. Lavinder, and B.J. Sullivan, *Protein stability by number: high-throughput and statistical approaches to one of protein science's most difficult problems*. Curr Opin Chem Biol, 2011. **15**(3): p. 443-51.
108. Gaspari, Z., et al., *Charged single alpha-helices in proteomes revealed by a consensus prediction approach*. Biochim Biophys Acta, 2012. **1824**(4): p. 637-46.
109. Wolny, M., et al., *Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability*. Sci Rep, 2017. **7**: p. 44341.
110. Swanson, C.J. and S. Sivaramakrishnan, *Harnessing the unique structural properties of isolated alpha-helices*. J Biol Chem, 2014. **289**(37): p. 25460-7.
111. Knight, P.J., et al., *The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head*. J Biol Chem, 2005. **280**(41): p. 34702-8.
112. Rohl, C.A., et al., *Protein structure prediction using Rosetta*. Methods Enzymol, 2004. **383**: p. 66-93.
113. Lance, B.K., C.M. Deane, and G.R. Wood, *Exploring the potential of template-based modelling*. Bioinformatics, 2010. **26**(15): p. 1849-56.
114. Zhang, Y., *Progress and challenges in protein structure prediction*. Curr Opin Struct Biol, 2008. **18**(3): p. 342-8.
115. Aloy, P., M. Pichaud, and R.B. Russell, *Protein complexes: structure prediction challenges for the 21st century*. Curr Opin Struct Biol, 2005. **15**(1): p. 15-22.
116. Sander, C., *Inverting the protein-folding problem*. Biochem Soc Symp, 1990. **57**: p. 25-33.
117. Chiu, T.L. and R.A. Goldstein, *Optimizing potentials for the inverse protein folding problem*. Protein Eng, 1998. **11**(9): p. 749-52.
118. O'Maille, P.E., et al., *Gene library synthesis by structure-based combinatorial protein engineering*. Methods Enzymol, 2004. **388**: p. 75-91.
119. Wei, Y., et al., *Solution structure of a de novo protein from a designed combinatorial library*. Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13270-3.
120. Koch, M.A., R. Breinbauer, and H. Waldmann, *Protein structure similarity as guiding principle for combinatorial library design*. Biol Chem, 2003. **384**(9): p. 1265-72.
121. Fuchs, A., A. Kirschner, and D. Frishman, *Prediction of helix-helix contacts and interacting helices in polytopic membrane proteins using neural networks*. Proteins, 2009. **74**(4): p. 857-71.
122. Lo, A., et al., *Predicting helix-helix interactions from residue contacts in membrane proteins*. Bioinformatics, 2009. **25**(8): p. 996-1003.
123. Wang, L.Y., *Covariation analysis of local amino acid sequences in recurrent protein local structures*. J Bioinform Comput Biol, 2005. **3**(6): p. 1391-409.
124. Singh, H., V. Hnizdo, and E. Demchuk, *Probabilistic model for two dependent circular variables*. Biometrika, 2002. **89**: p. 719-723.
125. Kumar, A. and L. Cowen, *Recognition of beta-structural motifs using hidden Markov models trained with simulated evolution*. Bioinformatics, 2010. **26**(12): p. i287-93.
126. Fooks, H.M., et al., *Amino acid pairing preferences in parallel beta-sheets in proteins*. J Mol Biol, 2006. **356**(1): p. 32-44.

127. Bystroff, C. and D. Baker, *Prediction of local structure in proteins using a library of sequence-structure motifs*. J Mol Biol, 1998. **281**(3): p. 565-77.
128. Bahar, I. and R.L. Jernigan, *Coordination geometry of nonbonded residues in globular proteins*. Fold Des, 1996. **1**(5): p. 357-70.
129. Hu, C. and P. Koehl, *Helix-sheet packing in proteins*. Proteins, 2010. **78**(7): p. 1736-47.
130. Holmes, J.B. and J. Tsai, *Characterizing Conserved Structural Contacts by Pair-wise Relative Contacts and Relative Packing Groups*. J Mol Biol, 2005. **354**(3): p. 706-21.
131. Bagci, Z., et al., *The origin and extent of coarse-grained regularities in protein internal packing*. Proteins, 2003. **53**(1): p. 56-67.
132. Day, R., et al., *Characterizing the regularity of tetrahedral packing motifs in protein tertiary structure*. Bioinformatics, 2010. **26**(24): p. 3059-66.
133. Huan, J., et al., *Mining protein family specific residue packing patterns from protein structure graphs*. RECOMB '04, 2004: p. 27-31.
134. Jonassen, I., I. Eidhammer, and W.R. Taylor, *Discovery of local packing motifs in protein structures*. Proteins, 1999. **34**(2): p. 206-19.
135. Preissner, R., A. Goede, and C. Frommel, *Spare parts for helix-helix interaction*. Protein Eng, 1999. **12**(10): p. 825-32.
136. Singh, R.K., A. Tropsha, and Vaisman, II, *Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino acid residues*. J Comput Biol, 1996. **3**(2): p. 213-21.
137. Adamian, L., et al., *Higher-order interhelical spatial interactions in membrane proteins*. J Mol Biol, 2003. **327**(1): p. 251-72.
138. Russell, R.B. and G.J. Barton, *Structural features can be unconserved in proteins with similar folds. An analysis of side-chain to side-chain contacts secondary structure and accessibility*. J Mol Biol, 1994. **244**(3): p. 332-50.
139. Nandi, C.L., J. Singh, and J.M. Thornton, *Atomic environments of arginine side chains in proteins*. Protein Eng, 1993. **6**(3): p. 247-59.
140. Kleywegt, G.J., *Recognition of spatial motifs in protein structures*. J Mol Biol, 1999. **285**(4): p. 1887-97.
141. Heringa, J. and P. Argos, *Side-chain clusters in protein structures and their role in protein folding*. J Mol Biol, 1991. **220**(1): p. 151-71.
142. Bandyopadhyay, D., et al., *Identification of family-specific residue packing motifs and their use for structure-based protein function prediction: II. Case studies and applications*. J Comput Aided Mol Des, 2009. **23**(11): p. 785-97.
143. Huan, J., et al., *Distance-based identification of structure motifs in proteins using constrained frequent subgraph mining*. Proc. LSS Comp. Sys. Bioinfor. Conf. CSB, 2006, 2006: p. 227-238.
144. Parry, D.A., R.D. Fraser, and J.M. Squire, *Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure*. J Struct Biol, 2008. **163**(3): p. 258-69.
145. Oakley, M.G. and J.J. Hollenbeck, *The design of antiparallel coiled coils*. Curr Opin Struct Biol, 2001. **11**(4): p. 450-7.
146. Gruber, M. and A.N. Lupas, *Historical review: another 50th anniversary--new periodicities in coiled coils*. Trends Biochem Sci, 2003. **28**(12): p. 679-85.
147. Efimov, A.V., *Complementary packing of alpha-helices in proteins*. FEBS Lett, 1999. **463**(1-2): p. 3-6.

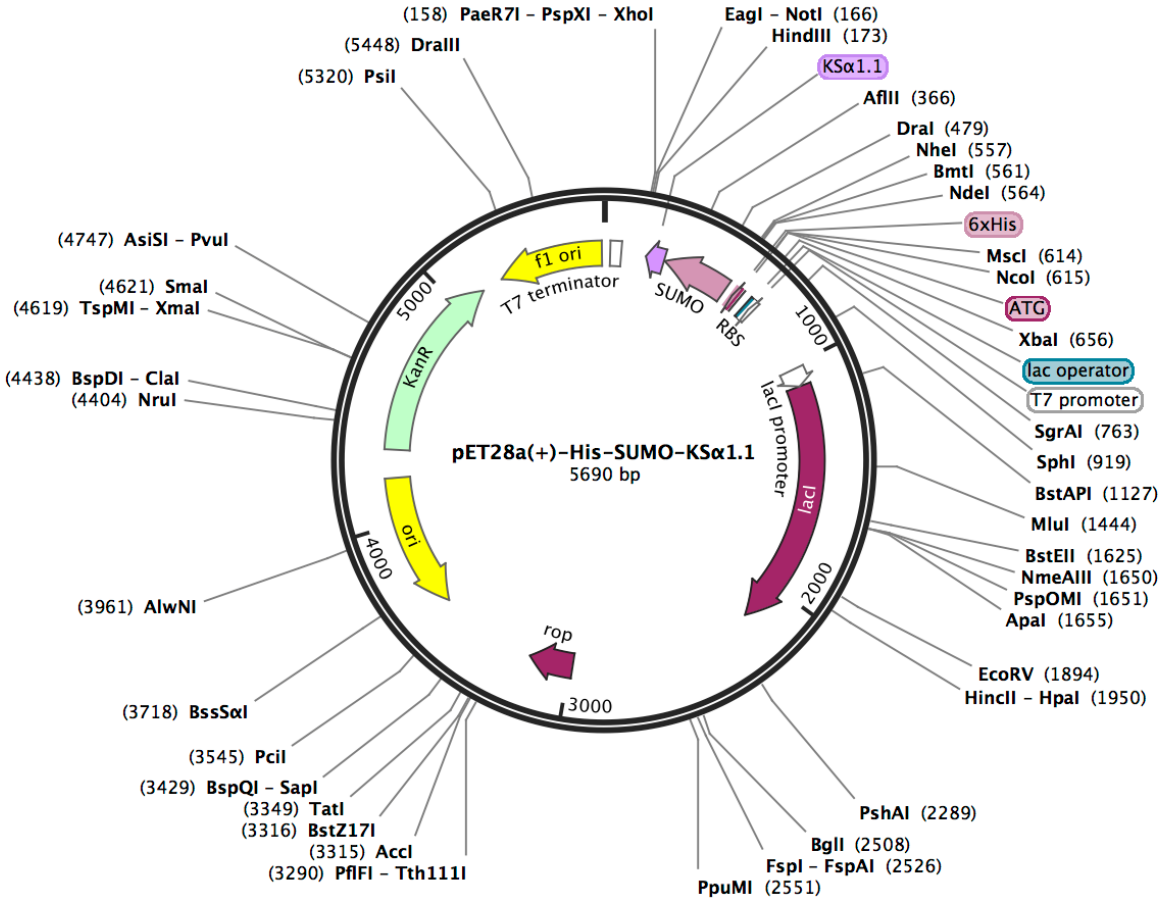
148. Crick, F.H.C., *The packing of a-helices: simple coiled coils*. Acta Crystallogr, 1953. **6**: p. 689–697.
149. Chothia, C., M. Levitt, and D. Richardson, *Helix to helix packing in proteins*. J Mol Biol, 1981. **145**(1): p. 215-50.
150. Ramos, J. and T. Lazaridis, *Energetic determinants of oligomeric state specificity in coiled coils*. J Am Chem Soc, 2006. **128**(48): p. 15499-510.
151. Ramos, J. and T. Lazaridis, *Computational analysis of residue contributions to coiled-coil topology*. Protein Sci, 2011. **20**(11): p. 1845-55.
152. Lupas, A., M. Van Dyke, and J. Stock, *Predicting coiled coils from protein sequences*. Science, 1991. **252**(5009): p. 1162-4.
153. Crick, F.H., *Is alpha-keratin a coiled coil?* Nature, 1952. **170**(4334): p. 882-3.
154. Pauling, L. and R.B. Corey, *Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type*. Nature, 1953. **171**(4341): p. 59-61.
155. Walshaw, J. and D.N. Woolfson, *Extended knobs-into-holes packing in classical and complex coiled-coil assemblies*. J Struct Biol, 2003. **144**(3): p. 349-61.
156. Walshaw, J., J.M. Shipway, and D.N. Woolfson, *Guidelines for the assembly of novel coiled-coil structures: alpha-sheets and alpha-cylinders*. Biochem Soc Symp, 2001(68): p. 111-23.
157. Walther, D., F. Eisenhaber, and P. Argos, *Principles of helix-helix packing in proteins: the helical lattice superposition model*. J Mol Biol, 1996. **255**(3): p. 536-53.
158. Schiffer, M. and A.B. Edmundson, *Use of helical wheels to represent the structures of proteins and to identify segments with helical potential*. Biophys J, 1967. **7**(2): p. 121-35.
159. Langosch, D. and J. Heringa, *Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils*. Proteins, 1998. **31**(2): p. 150-9.
160. Deng, Y., et al., *Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat*. Structure, 2006. **14**(2): p. 247-55.
161. Gandhi, N.S. and R.L. Mancera, *Computational Methods for the Prediction of the Structure and Interactions of Coiled-Coil Peptides*. Current Bioinformatics, 2008. **3**(3): p. 149-61.
162. Rackham, O.J., et al., *The evolution and structure prediction of coiled coils across all genomes*. J Mol Biol, 2010. **403**(3): p. 480-93.
163. Walters, R.F. and W.F. DeGrado, *Helix-packing motifs in membrane proteins*. Proc Natl Acad Sci U S A, 2006. **103**(37): p. 13658-63.
164. Dahiyat, B.I., C.A. Sarisky, and S.L. Mayo, *De novo protein design: towards fully automated sequence selection*. J Mol Biol, 1997. **273**(4): p. 789-96.
165. Dahiyat, B.I. and S.L. Mayo, *De novo protein design: fully automated sequence selection*. Science, 1997. **278**(5335): p. 82-7.
166. Schafmeister, C.E., et al., *A designed four helix bundle protein with native-like structure*. Nat Struct Biol, 1997. **4**(12): p. 1039-46.
167. Harbury, P.B., et al., *A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants*. Science, 1993. **262**(5138): p. 1401-7.
168. Harbury, P.B., et al., *High-resolution protein design with backbone freedom*. Science, 1998. **282**(5393): p. 1462-7.
169. Dieckmann, G.R. and W.F. DeGrado, *Modeling transmembrane helical oligomers*. Curr Opin Struct Biol, 1997. **7**(4): p. 486-94.

170. North, B., et al., *D(n)-symmetrical tertiary templates for the design of tubular proteins*. J Mol Biol, 2001. **311**(5): p. 1081-90.
171. Mason, J.M., et al., *Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design*. Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8989-94.
172. Hadley, E.B., et al., *Preferred side-chain constellations at antiparallel coiled-coil interfaces*. Proc Natl Acad Sci U S A, 2008. **105**(2): p. 530-5.
173. Moutevelis, E. and D.N. Woolfson, *A periodic table of coiled-coil protein structures*. J Mol Biol, 2009. **385**(3): p. 726-32.
174. Lupas, A.N. and J. Bassler, *Coiled Coils - A Model System for the 21st Century*. Trends Biochem Sci, 2017. **42**(2): p. 130-140.
175. Grigoryan, G. and A.E. Keating, *Structural specificity in coiled-coil interactions*. Curr Opin Struct Biol, 2008. **18**(4): p. 477-83.
176. Vinson, C., A. Acharya, and E.J. Taparowsky, *Deciphering B-ZIP transcription factor interactions in vitro and in vivo*. Biochim Biophys Acta, 2006. **1759**(1-2): p. 4-12.
177. Sodek, J., et al., *Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure*. Proc Natl Acad Sci U S A, 1972. **69**(12): p. 3800-4.
178. Cohen, C. and K.C. Holmes, *X-ray diffraction evidence for alpha-helical coiled-coils in native muscle*. J Mol Biol, 1963. **6**: p. 423-32.
179. Cohen, C. and D.A. Parry, *Alpha-helical coiled coils and bundles: how to design an alpha-helical protein*. Proteins, 1990. **7**(1): p. 1-15.
180. Zaccai, N.R., et al., *A de novo peptide hexamer with a mutable channel*. Nat Chem Biol, 2011. **7**(12): p. 935-41.
181. Fletcher, J.M., et al., *A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology*. ACS Synth Biol, 2012. **1**(6): p. 240-50.
182. Woolfson, D.N., et al., *New currency for old rope: from coiled-coil assemblies to alpha-helical barrels*. Curr Opin Struct Biol, 2012. **22**(4): p. 432-41.
183. Woolfson, D.N., *Coiled-Coil Design: Updated and Upgraded*. Subcell Biochem, 2017. **82**: p. 35-61.
184. Lumb, K.J. and P.S. Kim, *A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil*. Biochemistry, 1995. **34**(27): p. 8642-8.
185. Fletcher, J.M., et al., *N@a and N@d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces*. ACS Chem Biol, 2017. **12**(2): p. 528-538.
186. Gonzalez, L., Jr., D.N. Woolfson, and T. Alber, *Buried polar residues and structural specificity in the GCN4 leucine zipper*. Nat Struct Biol, 1996. **3**(12): p. 1011-8.
187. Acharya, A., V. Rishi, and C. Vinson, *Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R)*. Biochemistry, 2006. **45**(38): p. 11324-32.
188. Grigoryan, G. and W.F. Degrado, *Probing designability via a generalized model of helical bundle geometry*. J Mol Biol, 2011. **405**(4): p. 1079-100.
189. Zeng, X., et al., *Oligomerization properties of GCN4 leucine zipper e and g position mutants*. Protein Sci, 1997. **6**(10): p. 2218-26.
190. Hu, J.C., et al., *Probing the roles of residues at the e and g positions of the GCN4 leucine zipper by combinatorial mutagenesis*. Protein Sci, 1993. **2**(7): p. 1072-84.
191. John, M., et al., *Two pairs of oppositely charged amino acids from Jun and Fos confer heterodimerization to GCN4 leucine zipper*. J Biol Chem, 1994. **269**(23): p. 16247-53.

192. Newman, J.R. and A.E. Keating, *Comprehensive identification of human bZIP interactions with coiled-coil arrays*. Science, 2003. **300**(5628): p. 2097-101.
193. O'Shea, E.K., et al., *X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil*. Science, 1991. **254**(5031): p. 539-44.
194. Saudek, V., et al., *Solution structure of the DNA-binding domain of the yeast transcriptional activator protein GCN4*. Protein Eng, 1990. **4**(1): p. 3-10.
195. Saudek, V., et al., *The solution structure of a leucine-zipper motif peptide*. Protein Eng, 1991. **4**(5): p. 519-29.
196. Acharya, A., et al., *A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K*. Biochemistry, 2002. **41**(48): p. 14122-31.
197. Zeng, X., A.M. Herndon, and J.C. Hu, *Buried asparagines determine the dimerization specificities of leucine zipper mutants*. Proc Natl Acad Sci U S A, 1997. **94**(8): p. 3673-8.
198. Deppmann, C.D., et al., *Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs*. Nucleic Acids Res, 2004. **32**(11): p. 3435-45.
199. Rhys, G.G., et al., *Maintaining and breaking symmetry in homomeric coiled-coil assemblies*. Nat Commun, 2018. **9**(1): p. 4132.
200. Chen, Z., et al., *Programmable design of orthogonal protein heterodimers*. Nature, 2019. **565**(7737): p. 106-111.
201. Huang, P.S., et al., *High thermodynamic stability of parametrically designed helical bundles*. Science, 2014. **346**(6208): p. 481-485.
202. Potapov, V., J.B. Kaplan, and A.E. Keating, *Data-driven prediction and design of bZIP coiled-coil interactions*. PLoS Comput Biol, 2015. **11**(2): p. e1004046.
203. Walshaw, J. and D.N. Woolfson, *Socket: a program for identifying and analysing coiled-coil motifs within protein structures*. J Mol Biol, 2001. **307**(5): p. 1427-50.
204. Fraga, K., H. Joo, and J. Tsai, *An Amino Acid Code to Define a Protein's Tertiary Packing Surface*. Proteins: Struct. Funct. & Bioinform., 2015. (**under revision**).
205. Voronoi, G., *Nouveles applications des paramétres continus à la théorie des formes quadratiques*. J. Reine Angew. Math., 1907. **133**: p. 97-178.
206. Harpaz, Y., M. Gerstein, and C. Chothia, *Volume Changes on Protein Folding*. Structure, 1994. **2**: p. 641-649.
207. Plaxco, K.W., K.T. Simons, and D. Baker, *Contact order, transition state placement and the refolding rates of single domain proteins*. J Mol Biol, 1998. **277**(4): p. 985-94.
208. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases*. Biopolymers, 2008. **89**(5): p. 392-400.
209. Joo, H., et al., *An amino acid packing code for alpha-helical structure and protein design*. J Mol Biol, 2012. **419**(3-4): p. 234-54.

APPENDIX A: PLASMID MAPS AND SEQUENCES

Sequence: pET28a(+) + His-SUMO-KSa1.1.dna (Circular / 5690 bp)
Enzymes: Unique 6+ Cutters (50 of 653 total) Unique Cutters **Bold**
Features: 14 total



pET28a(+) + His-SUMO-KSa1.1.dna (Circular / 5690 bp)

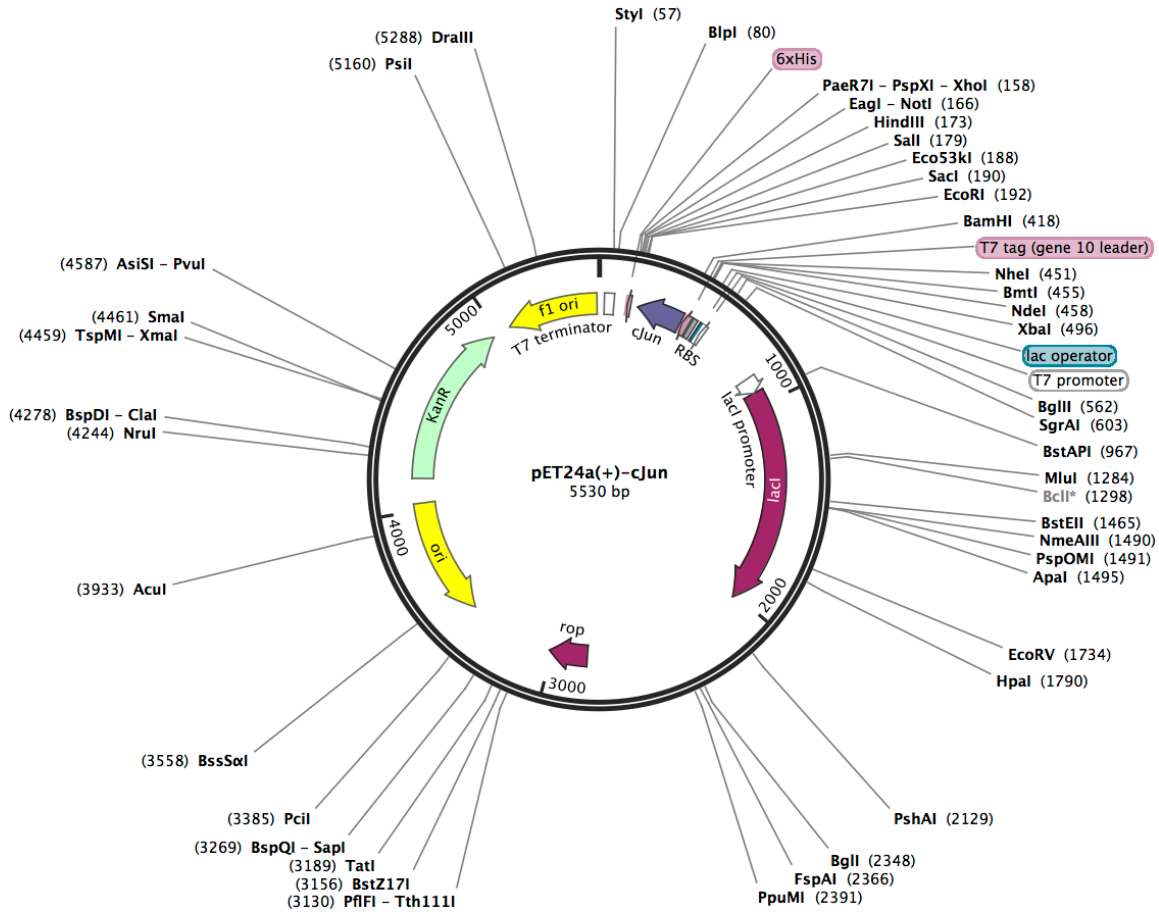
```
ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGC      55
CCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCC     110
TTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTG     165
CGGCCGCAAGCTTTCACAGTTCTTTCGCAATGCGCGCCATCGCGCTTTCCAGCGC     220
GGTCAGCGCATCCGCCACCGCTTTCGCCTGGCGTTCACCCCATCCACCAATCTGT     275
TCTCTGTGAGCCTCAATAATATCGTTATCCTCCATGTCCAAATCTTCAGGGGTCT     330
GATCAGCTTGGATTCTAATACCGTCGTACAAGAATCTTAAGGAGTCCATTTCCTT     385
ACCCTGTCTTTTAGCGAACGCTTCCATCAGCCTTCTTAAAGGAGTGGTCTTTTTG     440
ATCTTGAAGAAGATCTCTGAAGATCCATCGGACACCTTTAAATTGATGTGAGTCT     495
CAGGCTTGACTTCTGGCTTGACCTCTGGCTTAGCTTCTTGATTGACTTCTGAGTC     550
CGACATGCTAGCCATATGGCTGCCGCGCGGCACCAGGCCGCTGCTGTGATGATGA     605
TGATGATGGCCATGGGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAG     660
AGGGGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCG     715
GGATCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGC     770
CACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGG     825
GCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCC     880
CCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGC     935
GGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCG     990
CATAAGGGAGAGCGTCGAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGC    1045
GGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAAC    1100
CAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTC    1155
CCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAA    1210
GCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGG    1265
GCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCC    1320
GTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTG    1375
GTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACA    1430
ATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCA    1485
GGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGAT    1540
GTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGC    1595
GACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGC    1650
GGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATAT    1705
CTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCA    1760
TGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGC    1815
GATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAG    1870
TCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAG    1925
ACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCT    1980
GGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAG    2035
GGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCA    2090
ATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACG    2145
ACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTA    2200
GCTCACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCA    2255
GTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTG    2310
TCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTT    2365
CGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTA    2420
TTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAAC    2475
GTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCAT    2530
GATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGT    2585
TAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAA    2640
CGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAG    2695
TCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCA    2750
GGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGC    2805
ATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTT    2860
ACCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTG    2915
AGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTT    2970
ACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTT    3025
TATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGAT    3080
GAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCA    3135
```

pET28a(+) + His-SUMO-KSa1.1.dna (Circular / 5690 bp)

```
GCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCC        3190
GGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAG       3245
GGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTA       3300
GCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTG       3355
AGAGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA       3410
CCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT       3465
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG       3520
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG       3575
GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC       3630
GAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTAT       3685
AAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC       3740
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT       3795
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGC       3850
TGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAA       3905
CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC       3960
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGA       4015
AGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCT       4070
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAA       4125
ACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAA       4180
AAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG       4235
GAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTA       4290
CATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGC       4345
TCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGG       4400
CTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCC       4455
CGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTT       4510
ACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCA       4565
TCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCC       4620
CGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATT       4675
GTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATT       4730
GTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAA       4785
TAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTT       4840
GAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCG       4895
TCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAAT       4950
AGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCC       5005
ATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTC       5060
AAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCT       5115
CGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATACATATTTGAATGTATTTA       5170
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAA       5225
ATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCT       5280
CATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATA       5335
GACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAG       5390
AACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCAC       5445
TACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACT       5500
AAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCG       5555
AACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGG       5610
CAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCC       5665
GCTACAGGGCGCGTCCCATTCGCCA ●●●    5690
```

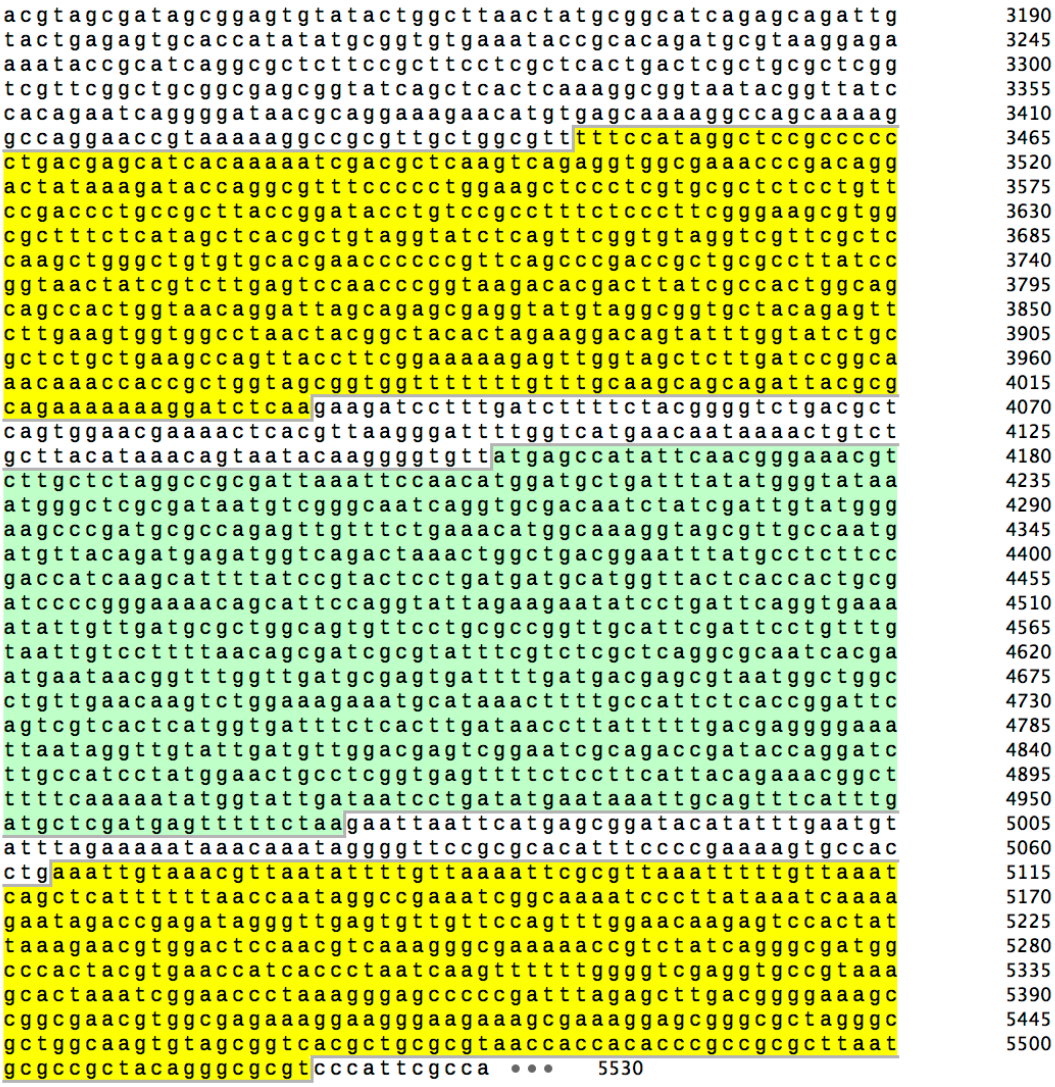
Sequence: pET24(+)-cJun.dna (Circular / 5530 bp)
Enzymes: Unique 6+ Cutters (51 of 653 total)
Features: 13 total

Unique Cutters **Bold**



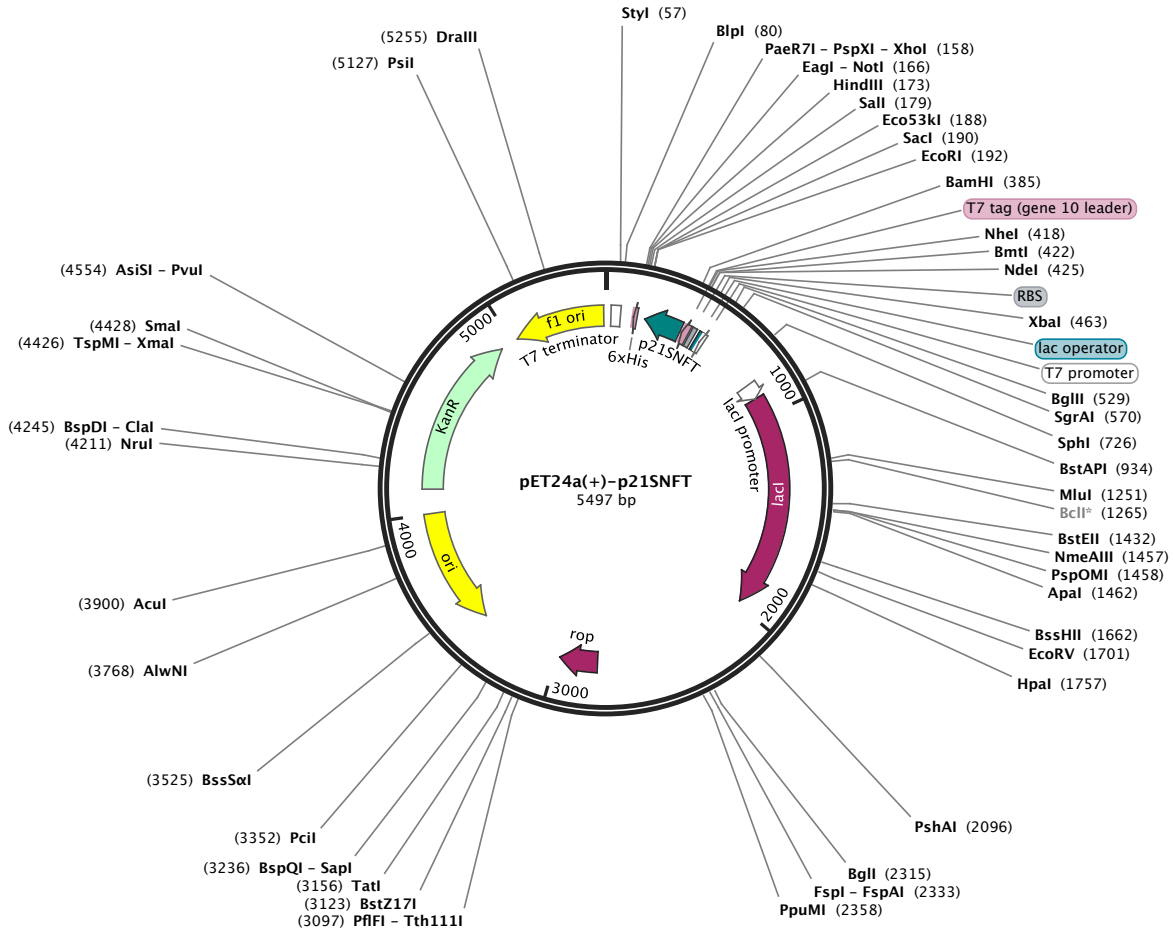
pET24(+)-cJun.dna (Circular / 5530 bp)

pET24(+)-cJun.dna (Circular / 5530 bp)

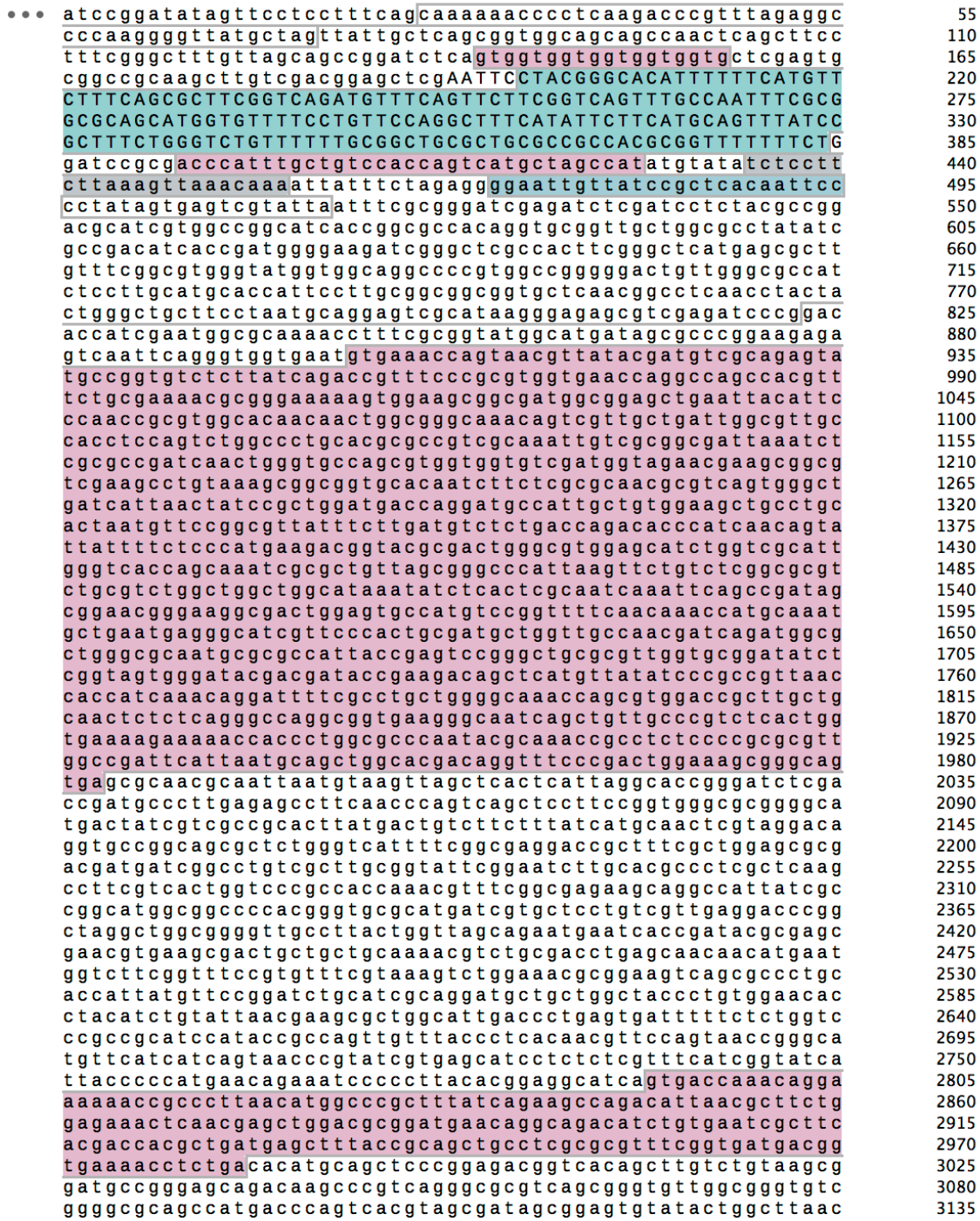


Sequence: pET24a(+)-p21SNFT.dna (Circular / 5497 bp)
Enzymes: Unique 6+ Cutters (55 of 653 total)
Features: 13 total

Unique Cutters **Bold**

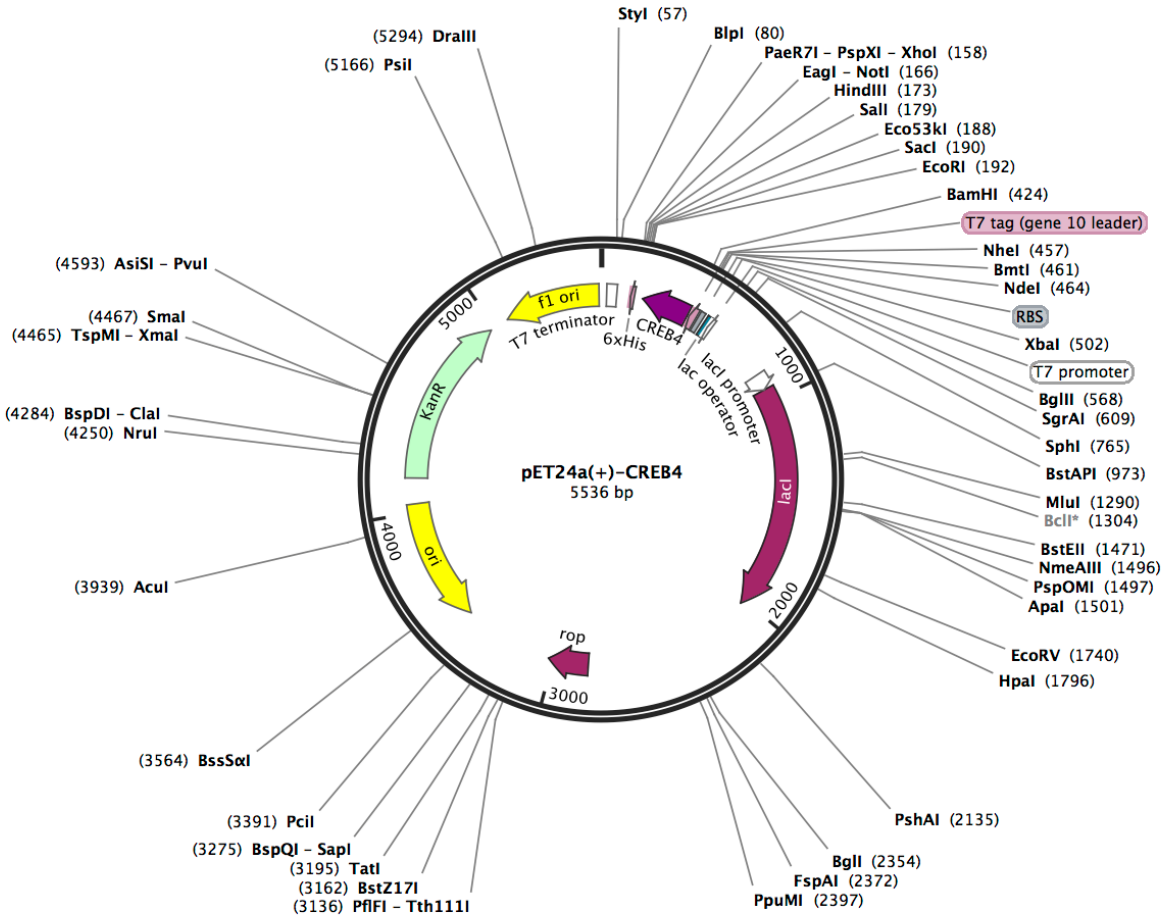


pET24a(+)-p21SNFT.dna (Circular / 5497 bp)

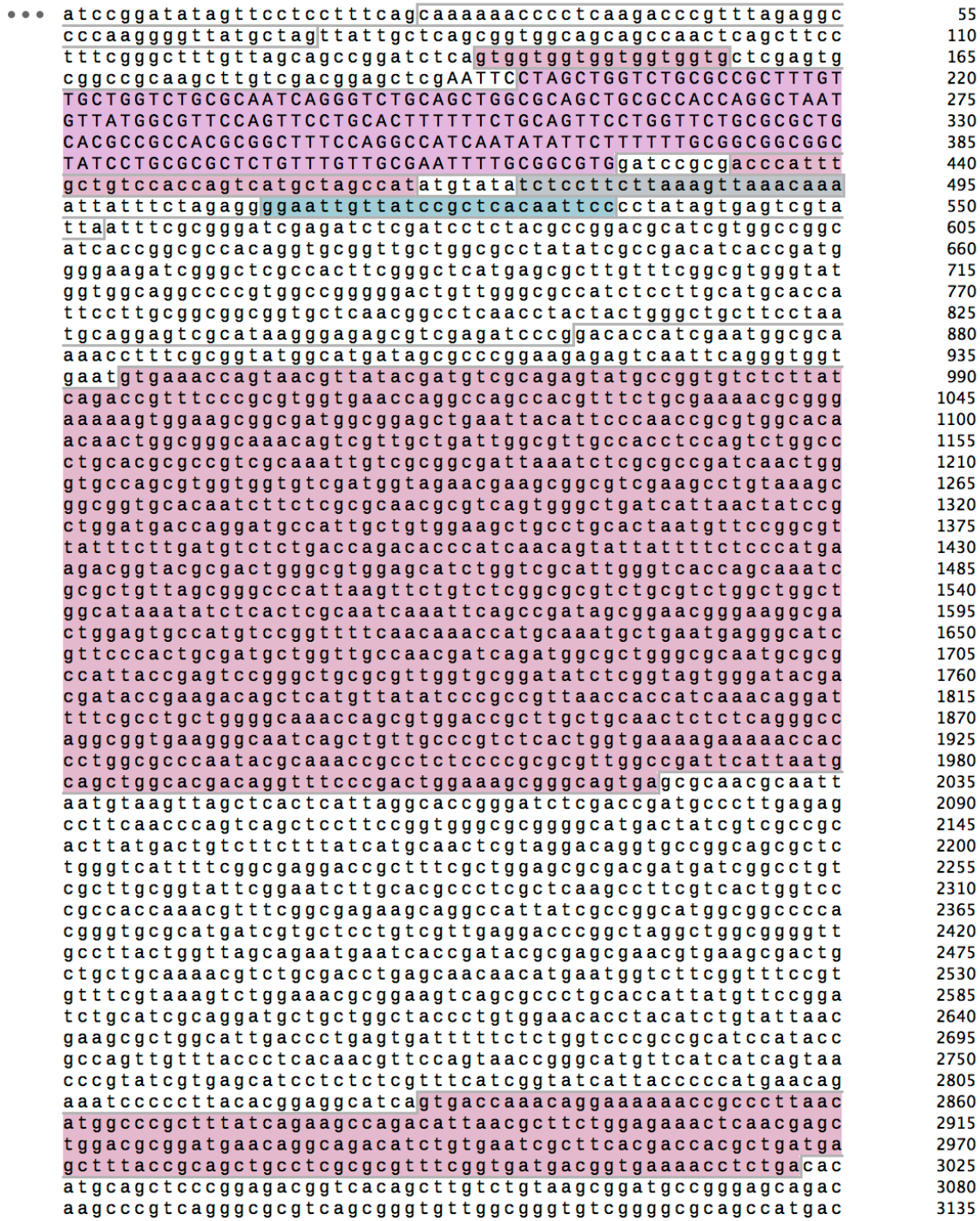


pET24a(+)-p21SNFT.dna (Circular / 5497 bp)

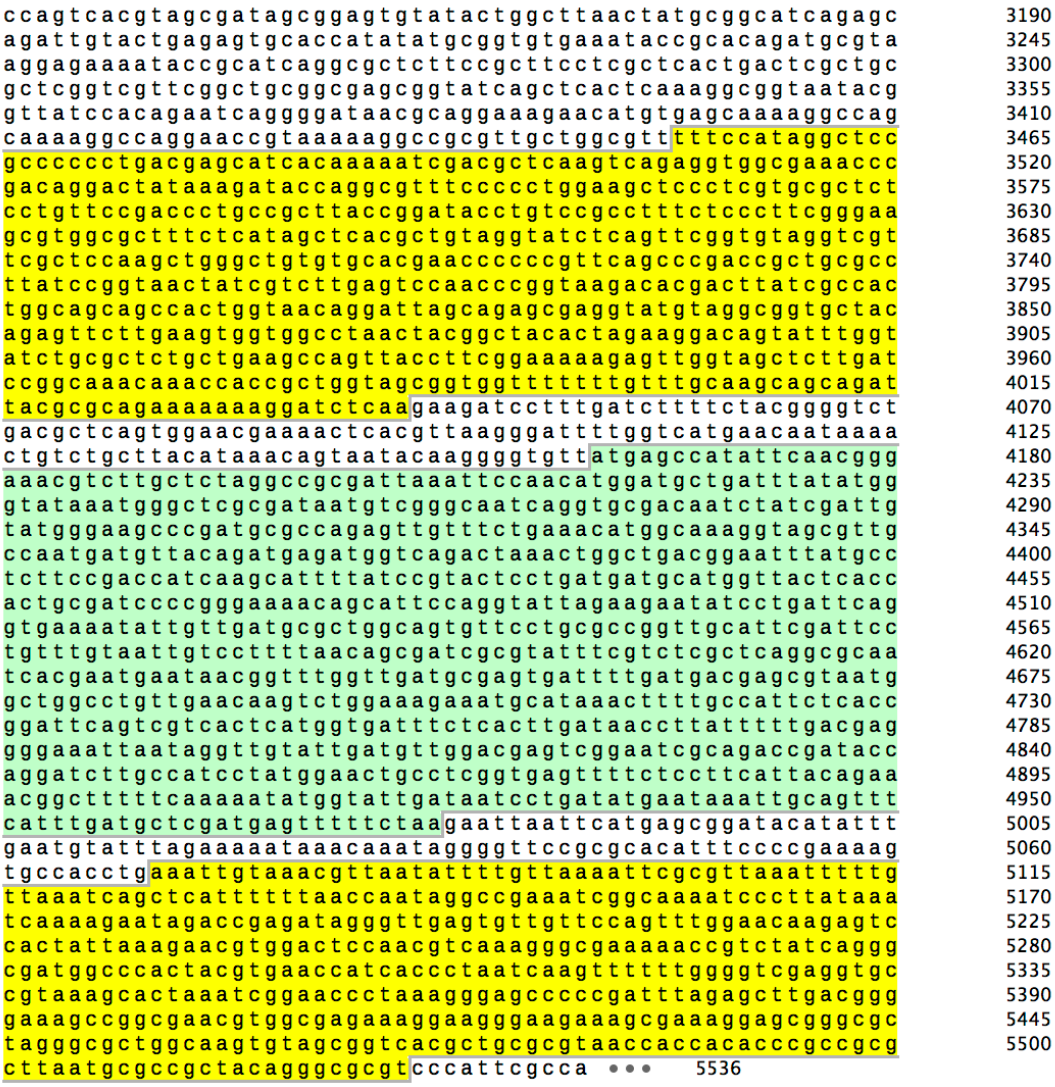
Sequence: pET24a(+)-CREB4.dna (Circular / 5536 bp)
Enzymes: Unique 6+ Cutters (52 of 653 total) Unique Cutters **Bold**
Features: 13 total



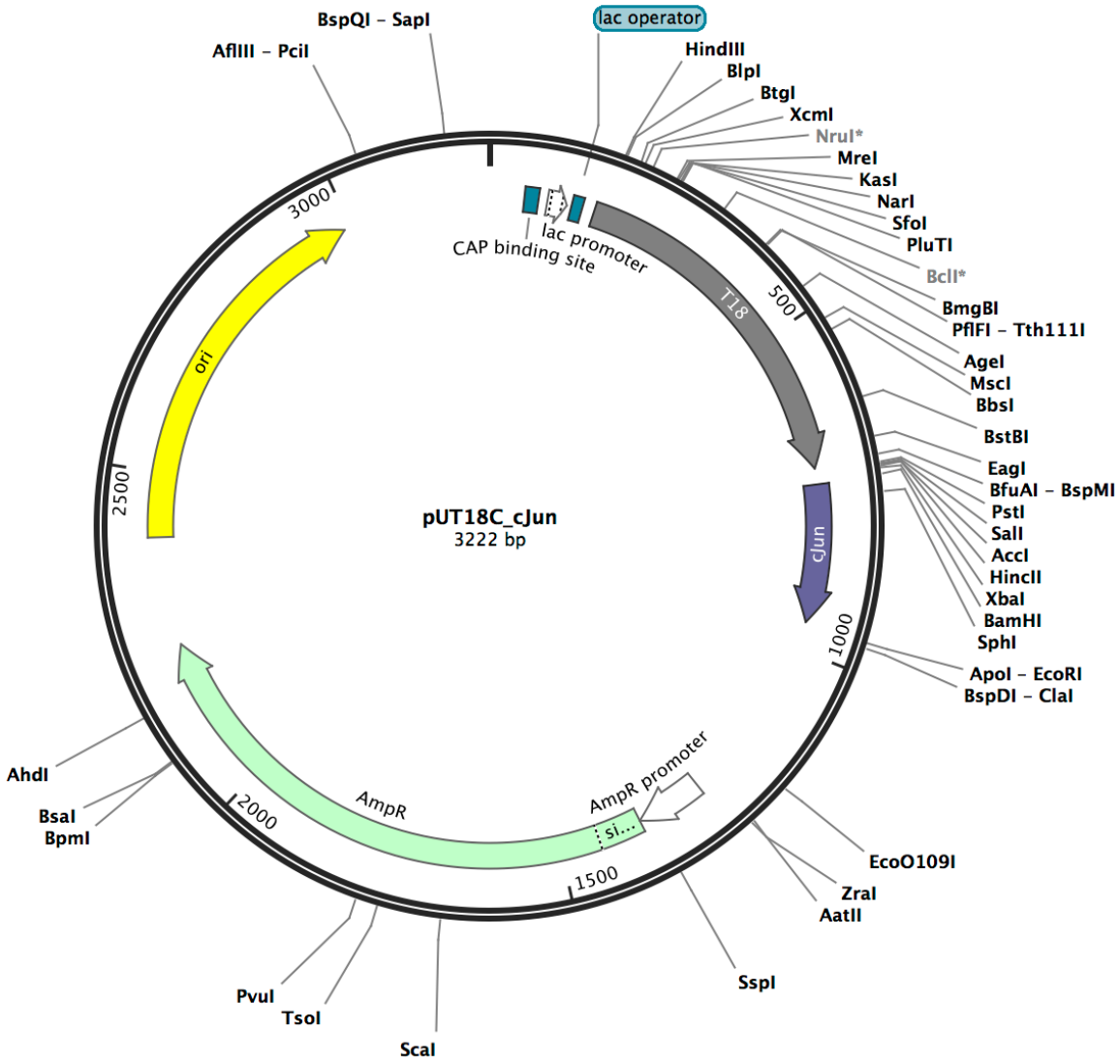
pET24a(+)-CREB4.dna (Circular / 5536 bp)



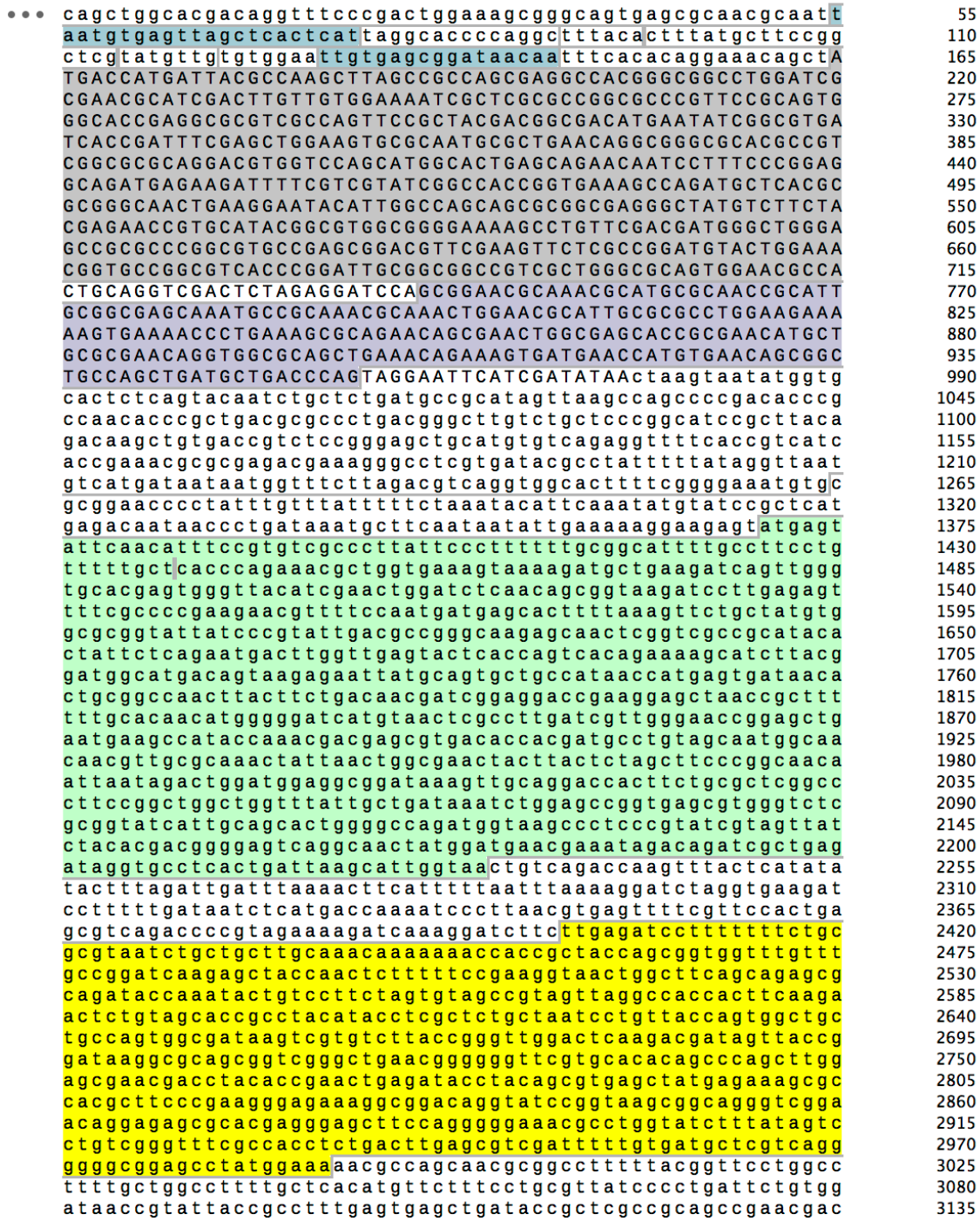
pET24a(+)-CREB4.dna (Circular / 5536 bp)



Sequence: pUT18C_cJun.dna (Circular / 3222 bp)
Enzymes: Unique 6+ Cutters (46 of 653 total) Unique Cutters **Bold**
Features: 8 total



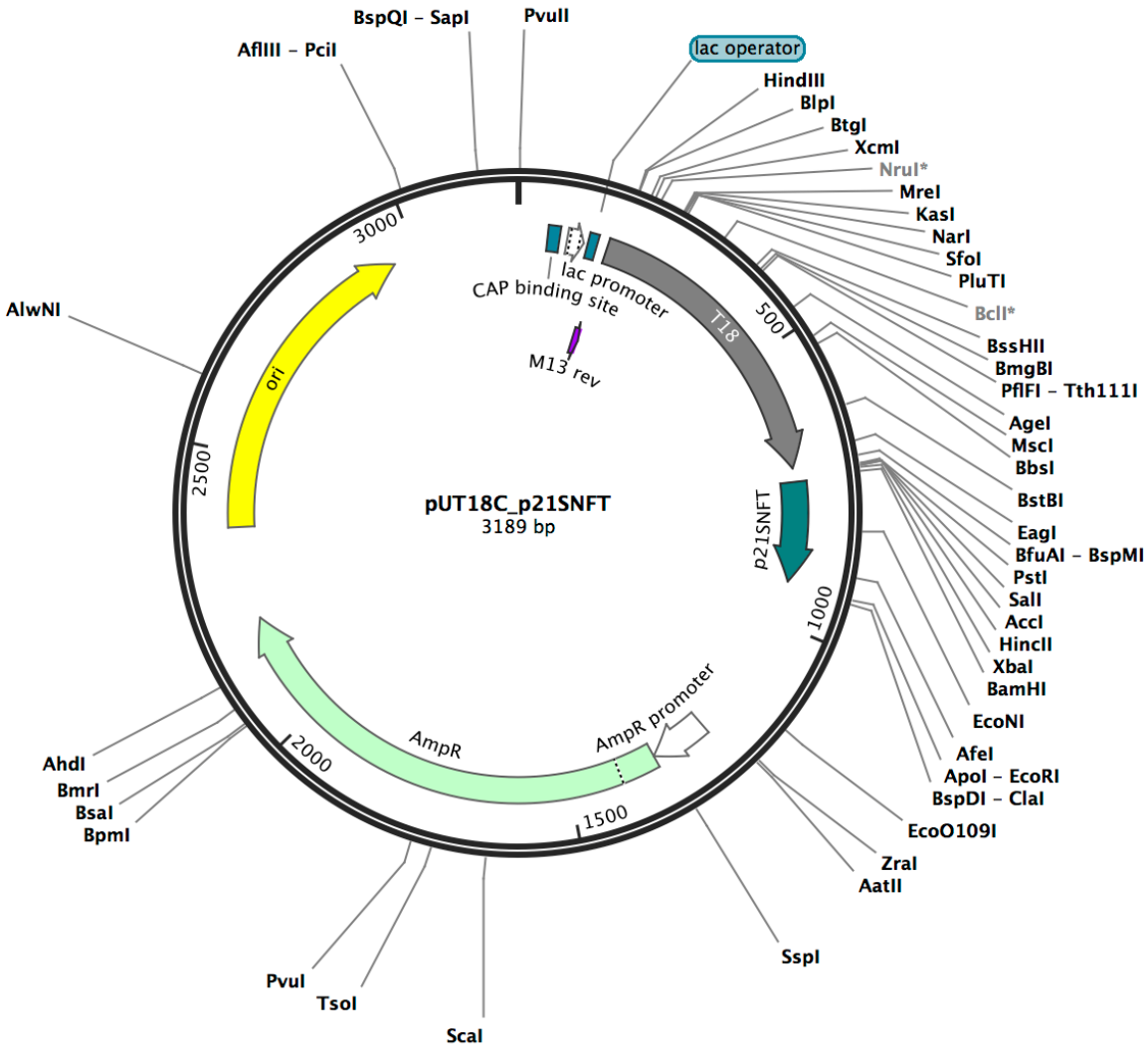
pUT18C_cJun.dna (Circular / 3222 bp)



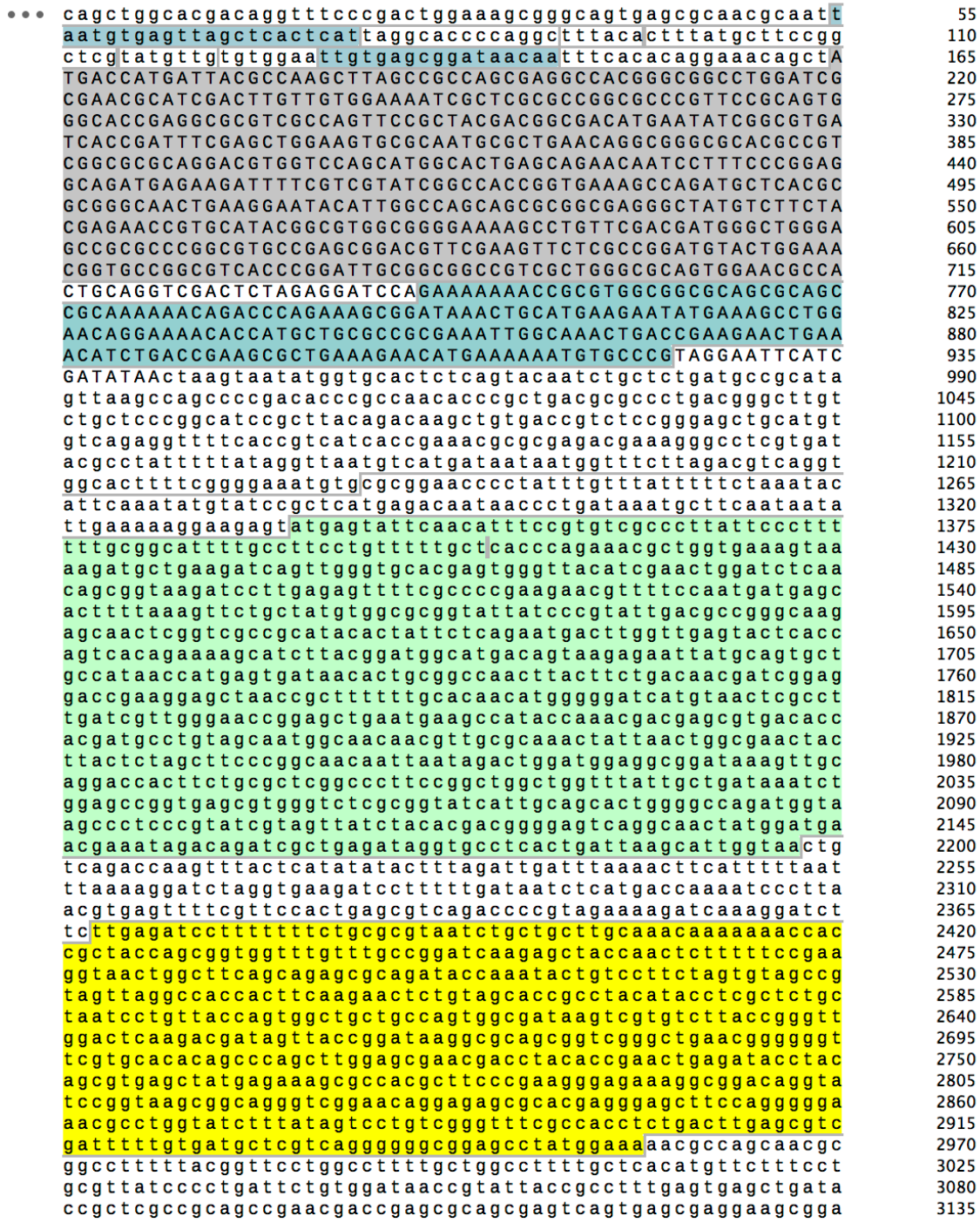
pUT18C_cJun.dna (Circular / 3222 bp)

```
cgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccg        3190
cctctccccgcgcgttggccgattcattaatg •••   3222
```

Sequence: pUT18C_p21SNFT.dna (Circular / 3189 bp)
Enzymes: Unique 6+ Cutters (51 of 653 total) Unique Cutters **Bold**
Features: 9 total



pUT18C_p21SNFT.dna (Circular / 3189 bp)



pUT18C_p21SNFT.dna (Circular / 3189 bp)

agagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatg ••• 3189

Sequence: pUT18C_CREB4.dna (Circular / 3228 bp)
Enzymes: Unique 6+ Cutters (44 of 653 total) Unique Cutters **Bold**
Features: 9 total

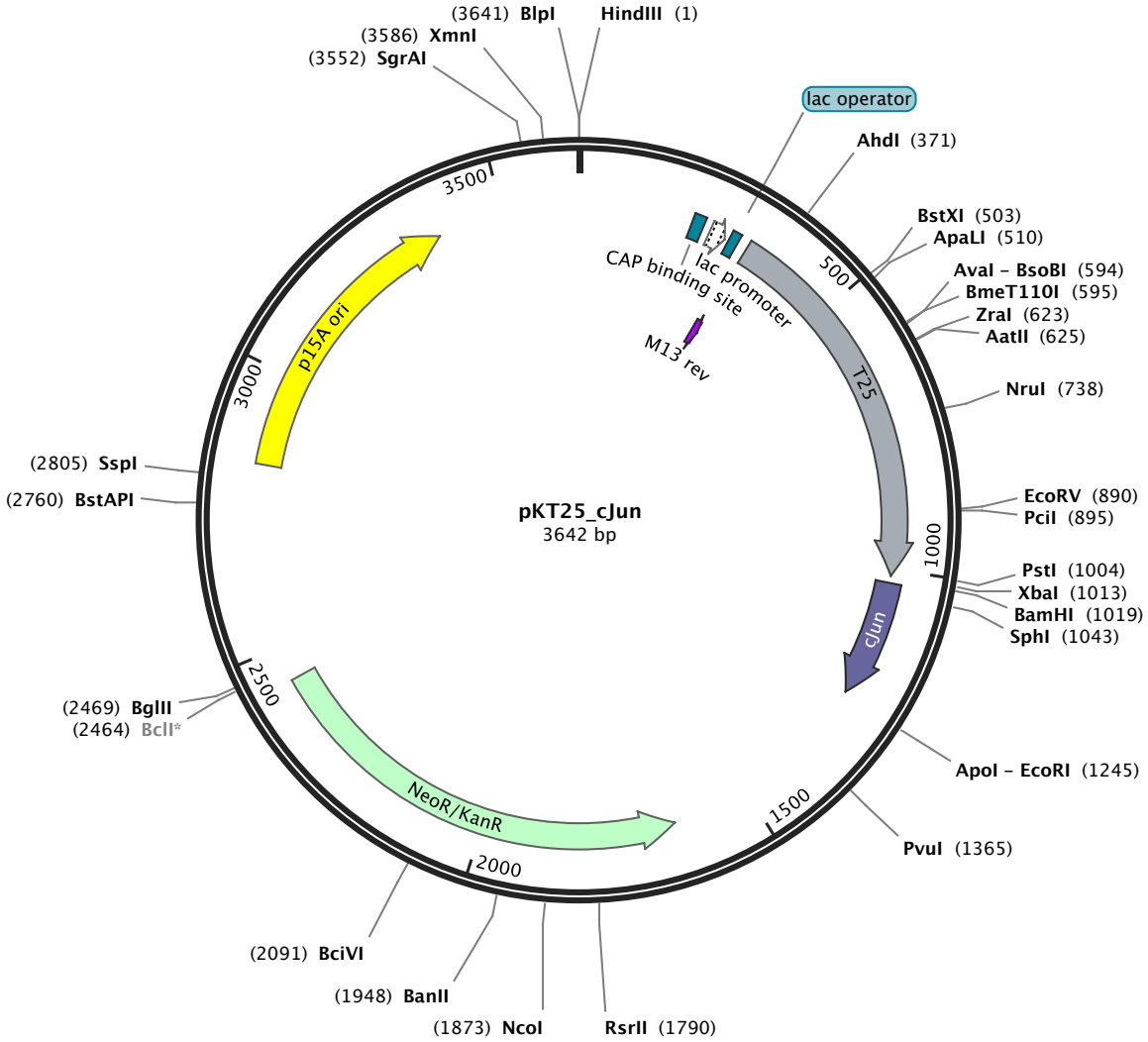
pUT18C_CREB4.dna (Circular / 3228 bp)

```
cagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaatt      55
aatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccgg     110
ctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctA     165
TGACCATGATTACGCCAAGCTTAGCCGCCAGCGAGGCCACGGGCGGCCTGGATCG     220
CGAACGCATCGACTTGTTGTGGAAAATCGCTCGCGCCGGCGCCCGTTCCGCAGTG     275
GGCACCGAGGCGCGTCGCCAGTTCCGCTACGACGGCGACATGAATATCGGCGTGA     330
TCACCGATTTCGAGCTGGAAGTGCGCAATGCGCTGAACAGGCGGGCGCACGCCGT     385
CGGCGCGCAGGACGTGGTCCAGCATGGCACTGAGCAGAACAATCCTTTCCCGGAG     440
GCAGATGAGAAGATTTTCGTCGTATCGGCCACCGGTGAAAGCCAGATGCTCACGC     495
GCGGGCAACTGAAGGAATACATTGGCCAGCAGCGCGGCGAGGGCTATGTCTTCTA     550
CGAGAACCGTGCATACGGCGTGGCGGGGAAAAGCCTGTTCGACGATGGGCTGGGA     605
GCCGCGCCCGGCGTGCCGAGCGGACGTTCGAAGTTCTCGCCGGATGTACTGGAAA     660
CGGTGCCGGCGTCACCCGGATTGCGGCGGCCGTCGCTGGGCGCAGTGGAACGCCA     715
CTGCAGGTCGACTCTAGAGGATCCACGCCGCAAAATTCGCAACAAACAGAGCGCG     770
CAGGATAGCCGCCGCCGCAAAAAAGAATATATTGATGGCCTGGAAAGCCGCGTGG     825
CGGCGTGCAGCGCGCAGAACCAGGAACTGCAGAAAAAAGTGCAGGAACTGGAACG     880
CCATAACATTAGCCTGGTGGCGCAGCTGCGCCAGCTGCAGACCCTGATTGCGCAG     935
ACCAGCAACAAAGCGGCGCAGACCAGCTAGGAATTCATCGATATAActaagtaat     990
atggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccga    1045
cacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccg    1100
cttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcacc    1155
gtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttatag    1210
gttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaa    1265
atgtgcgcggaaccccctatttgtttatttttctaaatacattcaaatatgtatcc    1320
gctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagt    1375
atgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgcc    1430
ttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatca    1485
gttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcctt    1540
gagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgc    1595
tatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccg    1650
catacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcat    1705
cttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtg    1760
ataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaac    1815
cgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccg    1870
gagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaa    1925
tggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccg    1980
gcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgc    2035
tcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtg    2090
ggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgt    2145
agttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatc    2200
gctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttact    2255
catatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggt    2310
gaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttc    2365
cactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttt    2420
ttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggt    2475
ttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagc    2530
agagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccact    2585
tcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagt    2640
ggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatag    2695
ttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagccca    2750
gcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgaga    2805
aagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagg    2860
gtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatcttt    2915
atagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctc    2970
gtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttc    3025
ctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgatt    3080
ctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccg    3135
```

pUT18C_CREB4.dna (Circular / 3228 bp)

```
aacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgc        3190
aaaccgcctctccccgcgcgttggccgattcattaatg •••   3228
```

Sequence: pKT25_cJun.dna (Circular / 3642 bp)
Enzymes: Unique 6+ Cutters (30 of 653 total) Unique Cutters **Bold**
Features: 8 total



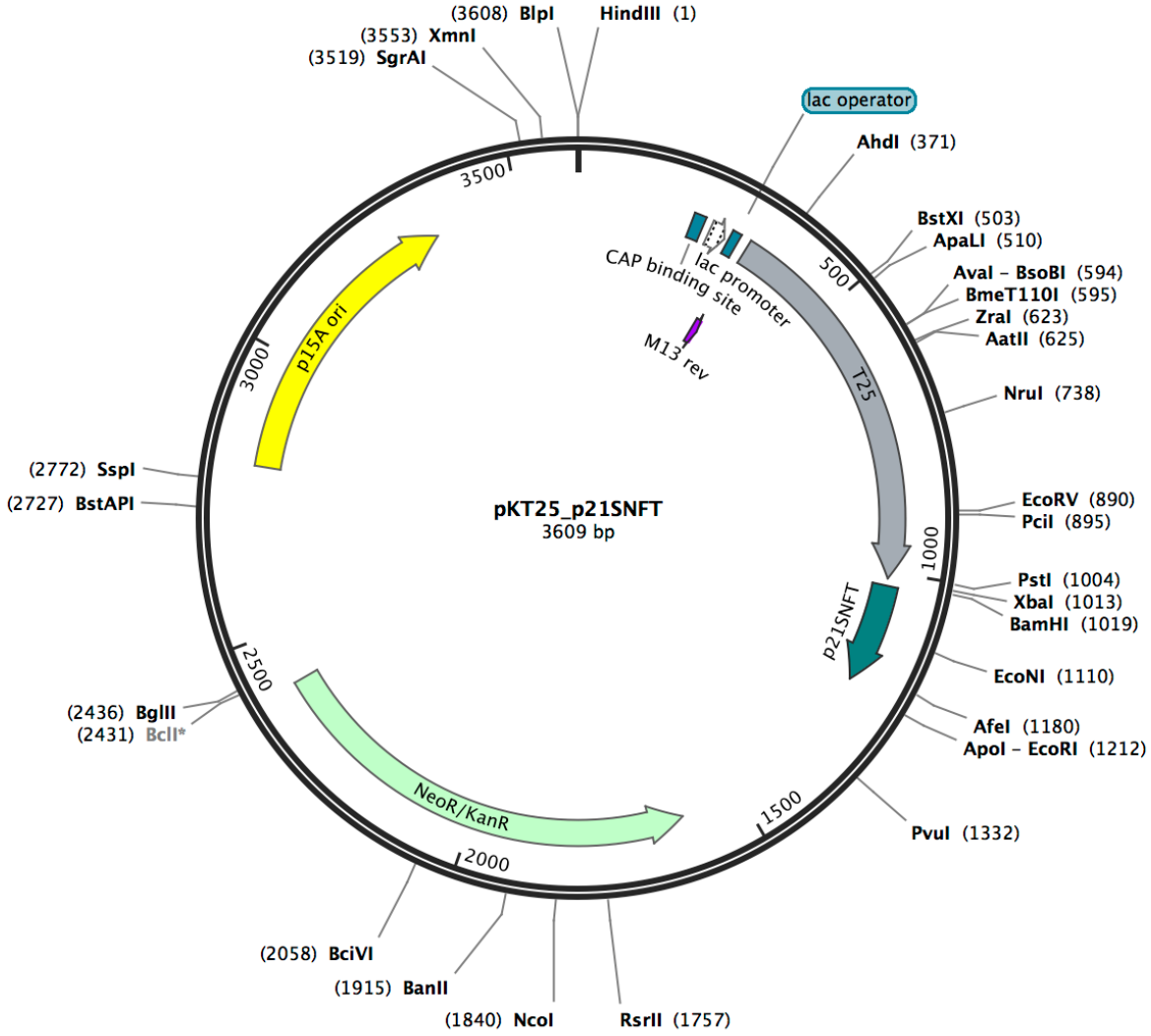
pKT25_cJun.dna (Circular / 3642 bp)

```
aagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgt      55
gtatgaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctggatgc     110
tgtaggcataggcttggttatgccggtactgccgggcctcttgcgggatctggca     165
cgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagt     220
tagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgt     275
tgtgtggaattgtgagcggataacaatttcacacaggaaacagctATGACCATGC     330
AGCAATCGCATCAGGCTGGTTACGCAAACGCCGCCGACCGGGAGTCTGGCATCCC     385
CGCAGCCGTACTCGATGGCATCAAGGCCGTGGCGAAGGAAAAAAACGCCACATTG     440
ATGTTCCGCCTGGTCAACCCCCATTCCACCAGCCTGATTGCCGAAGGGGTGGCCA     495
CCAAAGGATTGGGCGTGCACGCCAAGTCGTCCGATTGGGGGTTGCAGGCGGGCTA     550
CATTCCCGTCAACCCGAATCTTTCCAAACTGTTCGGCCGTGCGCCCGAGGTGATC     605
GCGCGGGCCGACAACGACGTCAACAGCAGCCTGGCGCATGGCCATACCGCGGTCG     660
ACCTGACGCTGTCGAAAGAGCGGCTTGACTATCTGCGGCAAGCGGGCCTGGTCAC     715
CGGCATGGCCGATGGCGTGGTCGCGAGCAACCACGCAGGCTACGAGCAGTTCGAG     770
TTTCGCGTGAAGGAAACCTCGGACGGGCGCTATGCCGTGCAGTATCGCCGCAAGG     825
GCGGCGACGATTTCGAGGCGGTCAAGGTGATCGGCAATGCCGCCGGTATTCCACT     880
GACGGCGGATATCGACATGTTCGCCATTATGCCGCATCTGTCCAACTTCCGCGAC     935
TCGGCGCGCAGTTCGGTGACCAGCGGCGATTCGGTGACCGATTACCTGGCGCGCA     990
CGCGGCGGGCTGCAGGGTCGACTCTAGAGGATCCAGCGGAACGCAAACGCATGCG    1045
CAACCGCATTGCGGCGAGCAAATGCCGCAAACGCAAACTGGAACGCATTGCGCGC    1100
CTGGAAGAAAAAGTGAAAACCCTGAAAGCGCAGAACAGCGAACTGGCGAGCACCG    1155
CGAACATGCTGCGCGAACAGGTGGCGCAGCTGAAACAGAAAGTGATGAACCATGT    1210
GAACAGCGGCTGCCAGCTGATGCTGACCCAGTAGGaattcggccgtcgttttaca    1265
acgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacat    1320
ccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttccc    1375
aacagttgcgcagcctgaatggcgaatggcgctgatgtccggcggtgcttttgcc    1430
gttacgcaccaccccgtcagtagctgaacaggagggacaggggtgggcgaagaa    1485
ctccagcatgagatccccgcgctggaggatcatccagccggcgtcccggaaaacg    1540
attccgaagcccaacctttcatagaaggcggcggtggaatcgaaatctcgtgatg    1595
gcaggttgggcgtcgcttggtcggtcatttcgaaccccagagtcccgctcagaag    1650
aactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgatac    1705
cgtaaagcacgaggaagcggtcagcccattcgccgccaagctcttcagcaatatc    1760
acgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacag    1815
tcgatgaatccagaaaagcggccattttccaccatgatattcggcaagcaggcat    1870
cgccatgggtcacgacgagatcctcgccgtcgggcatccgcgccttgagcctggc    1925
gaacagttcggctggcgcgagcccctgatgctcttcgtccagatcatcctgatcg    1980
acaagaccggcttccatccgagtacgtgctcgctcgatgcgatgtttcgcttggt    2035
ggtcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattgcatcagc    2090
catgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgcccc    2145
ggcacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgtcgagca    2200
cagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgcctcgtc    2255
ttggagttcattcagggcaccggacaggtcggtcttgacaaaaagaaccgggcgc    2310
ccctgcgctgacagccggaacacggcggcatcagagcagccgattgtctgttgtg    2365
cccagtcatagccgaatagcctctccacccaagcggccggagaacctgcgtgcaa    2420
tccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatctt    2475
gatccccctgcgccatcagatccttggcggcaagaaagccatccagtttactttgc    2530
agggcttcccaaccttaccagagggcgccccagctggcaattccggttcgcttgc    2585
tgtccataaaaccgcccagtctagctatcgccatgtaagcccactgcaagctacc    2640
tgctttctctttgcgcttgcgttttcccttgtccagatagcccagtagctgacat    2695
tcatccggggtcagcaccgtttctgcggactggctttctacgtgttccgcttcct    2750
ttagcagcccttgcgccctgagtgcttgcggcagcgtgaagctgtgcctagaaat    2805
attttatctgattaataagatgatcttcttgagatcgttttggtctgcgcgtaat    2860
ctcttgctctgaaaacgaaaaaaccgccttgcagggcggttttttcgaaggttctc    2915
tgagctaccaactctttgaaccgaggtaactggcttggaggagcgcagtcaccaa    2970
aacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctc    3025
taaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggtt    3080
ggactcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggt    3135
```

pKT25_cJun.dna (Circular / 3642 bp)

```
tcgtgcatacagtccagcttggagcgaactgcctacccggaactgagtgtcaggc      3190
gtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaagg      3245
caggaacaggagagcgcacgagggagccgccaggggaaacgcctggtatctttat      3300
agtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgt      3355
caggggggcggagcctatggaaaaacggctttgccgcggccctctcacttccctg      3410
ttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccattt      3465
ccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgaggaagcgga      3520
atatatcctgtatcacatattctgctgacgcaccggtgcagccttttttctcctg      3575
ccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtat      3630
atacactccgct •••   3642
```

Sequence: pKT25_p21SNFT.dna (Circular / 3609 bp)
Enzymes: Unique 6+ Cutters (31 of 653 total) Unique Cutters **Bold**
Features: 8 total



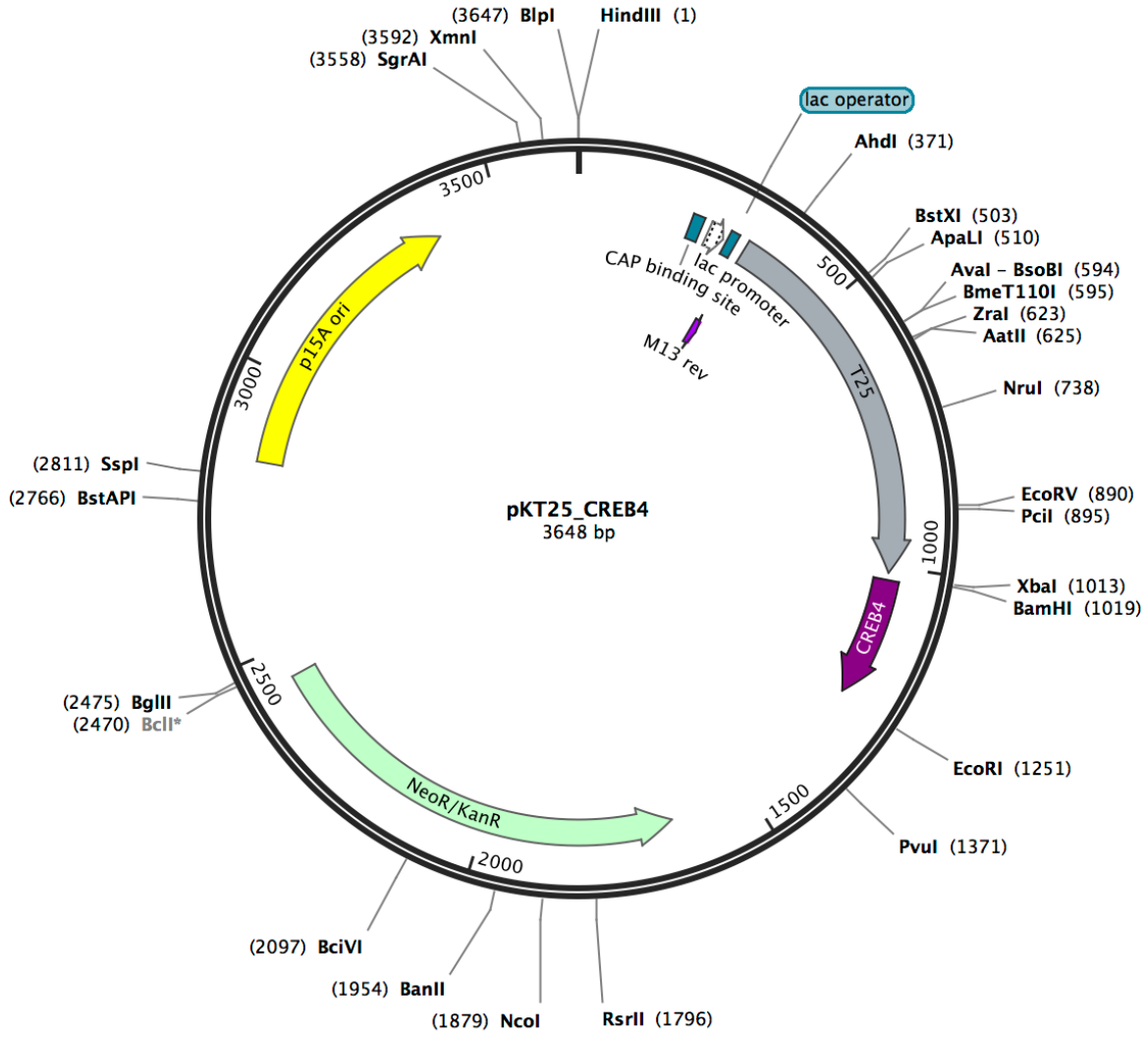
pKT25_p21SNFT.dna (Circular / 3609 bp)

```
aagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgt   55
gtatgaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctggatgc  110
tgtaggcataggcttggttatgccggtactgccgggcctcttgcgggatctggca  165
cgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagt  220
tagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgt  275
tgtgtggaattgtgagcggataacaatttcacacaggaaacagctATGACCATGC  330
AGCAATCGCATCAGGCTGGTTACGCAAACGCCGCCGACCGGGAGTCTGGCATCCC  385
CGCAGCCGTACTCGATGGCATCAAGGCCGTGGCGAAGGAAAAAAACGCCACATTG  440
ATGTTCCGCCTGGTCAACCCCCATTCCACCAGCCTGATTGCCGAAGGGGTGGCCA  495
CCAAAGGATTGGGCGTGCACGCCAAGTCGTCCGATTGGGGGTTGCAGGCGGGCTA  550
CATTCCCGTCAACCCGAATCTTTCCAAACTGTTCGGCCGTGCGCCCGAGGTGATC  605
GCGCGGGCCGACAACGACGTCAACAGCAGCCTGGCGCATGGCCATACCGCGGTCG  660
ACCTGACGCTGTCGAAAGAGCGGCTTGACTATCTGCGGCAAGCGGGCCTGGTCAC  715
CGGCATGGCCGATGGCGTGGTCGCGAGCAACCACGCAGGCTACGAGCAGTTCGAG  770
TTTCGCGTGAAGGAAACCTCGGACGGGCGCTATGCCGTGCAGTATCGCCGCAAGG  825
GCGGCGACGATTTCGAGGCGGTCAAGGTGATCGGCAATGCCGCCGGTATTCCACT  880
GACGGCGGATATCGACATGTTCGCCATTATGCCGCATCTGTCCAACTTCCGCGAC  935
TCGGCGCGCAGTTCGGTGACCAGCGGCGATTCGGTGACCGATTACCTGGCGCGCA  990
CGCGGCGGGCTGCAGGGTCGACTCTAGAGGATCCAGAAAAAAACCGCGTGGCGGC 1045
GCAGCGCAGCCGCAAAAAACAGACCCAGAAAGCGGATAAACTGCATGAAGAATAT 1100
GAAAGCCTGGAACAGGAAAACACCATGCTGCGCCGCGAAATTGGCAAACTGACCG 1155
AAGAACTGAAACATCTGACCGAAGCGCTGAAAGAACATGAAAAAATGTGCCCGTA 1210
GGaattcggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttaccca 1265
acttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagag 1320
gcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgct 1375
gatgtccggcggtgcttttgccgttacgcaccaccccgtcagtagctgaacagga 1430
gggacaggggtgggcgaagaactccagcatgagatccccgcgctggaggatcat 1485
ccagccggcgtcccggaaaacgattccgaagcccaacctttcatagaaggcggcg 1540
gtggaatcgaaatctcgtgatggcaggttgggcgtcgcttggtcggtcatttcga 1595
accccagagtcccgctcagaagaactcgtcaagaaggcgatagaaggcgatgcgc 1650
tgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtcagcccattcgc 1705
cgccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtc 1760
cgccacacccagccggccacagtcgatgaatccagaaaagcggccatttttccacc 1815
atgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgccgtcgg 1870
gcatccgcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgctc 1925
ttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgc 1980
tcgatgcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaagcgtat 2035
gcagccgccgcattgcatcagccatgatggatactttctcggcaggagcaaggtg 2090
agatgacaggagatcctgccccggcacttcgcccaatagcagccagtcccttccc 2145
gcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggccagcc 2200
acgatagccgcgctgcctcgtcttggagttcattcagggcaccggacaggtcggt 2255
cttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacggcggcatca 2310
gagcagccgattgtctgttgtgcccagtcatagccgaatagcctctccacccaag 2365
cggccggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctca 2420
tcctgtctcttgatcagatcttgatcccctgcgccatcagatccttggcggcaag 2475
aaagccatccagtttactttgcagggcttcccaaccttaccagagggcgccccag 2530
ctggcaattccggttcgcttgctgtccataaaaccgcccagtctagctatcgcca 2585
tgtaagcccactgcaagctacctgctttctctttgcgcttgcgttttcccttgtc 2640
cagatagcccagtagctgacattcatccggggtcagcaccgtttctgcggactgg 2695
ctttctacgtgttccgcttcctttagcagcccttgcgccctgagtgcttgcggca 2750
gcgtgaagctgtgcctagaaatattttatctgattaataagatgatcttcttgag 2805
atcgttttggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgca 2860
gggcggttttttcgaaggttctctgagctaccaactctttgaaccgaggtaactgg 2915
cttggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgc 2970
atgacttcaagactaactcctctaaatcaattaccagtggctgctgccagtggtg 3025
cttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcgca 3080
gcggtcggactgaacgggggttcgtgcatacagtccagcttggagcgaactgcc  3135
```

pKT25_p21SNFT.dna (Circular / 3609 bp)

```
tacccggaactgagtgtcaggcgtggaatgagacaaacgcggccataacagcgga    3190
atgacaccggtaaaccgaaaggcaggaacaggagagcgcacgagggagccgccag    3245
gggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgag    3300
cgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttg    3355
ccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatct    3410
ccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcg    3465
agtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcac    3520
cggtgcagccttttttctcctgccacatgaagcacttcactgacaccctcatcag    3575
tgccaacatagtaagccagtatatacactccgct •••    3609
```

Sequence: pKT25_CREB4.dna (Circular / 3648 bp)
Enzymes: Unique 6+ Cutters (27 of 653 total) Unique Cutters **Bold**
Features: 8 total



pKT25_CREB4.dna (Circular / 3648 bp)

```
aagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgt      55
gtatgaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctggatgc     110
tgtaggcataggcttggttatgccggtactgccgggcctcttgcgggatctggca     165
cgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagt     220
tagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgt     275
tgtgtggaattgtgagcggataacaatttcacacaggaaacagctATGACCATGC     330
AGCAATCGCATCAGGCTGGTTACGCAAACGCCGCCGACCGGGAGTCTGGCATCCC     385
CGCAGCCGTACTCGATGGCATCAAGGCCGTGGCGAAGGAAAAAAACGCCACATTG     440
ATGTTCCGCCTGGTCAACCCCCATTCCACCAGCCTGATTGCCGAAGGGGTGGCCA     495
CCAAAGGATTGGGCGTGCACGCCAAGTCGTCCGATTGGGGGTTGCAGGCGGGCTA     550
CATTCCCGTCAACCCGAATCTTTCCAAACTGTTCGGCCGTGCGCCCGAGGTGATC     605
GCGCGGGCCGACAACGACGTCAACAGCAGCCTGGCGCATGGCCATACCGCGGTCG     660
ACCTGACGCTGTCGAAAGAGCGGCTTGACTATCTGCGGCAAGCGGGCCTGGTCAC     715
CGGCATGGCCGATGGCGTGGTCGCGAGCAACCACGCAGGCTACGAGCAGTTCGAG     770
TTTCGCGTGAAGGAAACCTCGGACGGGCGCTATGCCGTGCAGTATCGCCGCAAGG     825
GCGGCGACGATTTCGAGGCGGTCAAGGTGATCGGCAATGCCGCCGGTATTCCACT     880
GACGGCGGATATCGACATGTTCGCCATTATGCCGCATCTGTCCAACTTCCGCGAC     935
TCGGCGCGCAGTTCGGTGACCAGCGGCGATTCGGTGACCGATTACCTGGCGCGCA     990
CGCGGCGGGCTGCAGGGTCGACTCTAGAGGATCCACGCCGCAAAATTCGCAACAA    1045
ACAGAGCGCGCAGGATAGCCGCCGCCGCAAAAAAGAATATATTGATGGCCTGGAA    1100
AGCCGCGTGGCGGCGTGCAGCGCGCAGAACCAGGAACTGCAGAAAAAAGTGCAGG    1155
AACTGGAACGCCATAACATTAGCCTGGTGGCGCAGCTGCGCCAGCTGCAGACCCT    1210
GATTGCGCAGACCAGCAACAAAGCGGCGCAGACCAGCTAGGaattcggccgtcgt    1265
tttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgca    1320
gcacatcccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcc    1375
cttcccaacagttgcgcagcctgaatggcgaatggcgctgatgtccggcggtgct    1430
tttgccgttacgcaccaccccgtcagtagctgaacaggagggacaggggtgggc    1485
gaagaactccagcatgagatccccgcgctggaggatcatccagccggcgtcccgg    1540
aaaacgattccgaagcccaacctttcatagaaggcggcggtggaatcgaaatctc    1595
gtgatggcaggttgggcgtcgcttggtcggtcatttcgaaccccagagtcccgct    1650
cagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcgg    1705
cgataccgtaaagcacgaggaagcggtcagcccattcgccgccaagctcttcagc    1760
aatatcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccgg    1815
ccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaagc    1870
aggcatcgccatgggtcacgacgagatcctcgccgtcgggcatccgcgccttgag    1925
cctggcgaacagttcggctggcgcgagcccctgatgctcttcgtccagatcatcc    1980
tgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgtttcg    2035
cttggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattgc    2090
atcagccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcc    2145
tgccccggcacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgt    2200
cgagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgc    2255
ctcgtcttggagttcattcagggcaccggacaggtcggtcttgacaaaaagaacc    2310
gggcgcccctgcgctgacagccggaacacggcggcatcagagcagccgattgtct    2365
gttgtgcccagtcatagccgaatagcctctccacccaagcggccggagaacctgc    2420
gtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatca    2475
gatcttgatcccctgcgccatcagatccttggcggcaagaaagccatccagttta    2530
ctttgcagggcttcccaaccttaccagagggcgccccagctggcaattccggttc    2585
gcttgctgtccataaaaccgcccagtctagctatcgccatgtaagcccactgcaa    2640
gctacctgctttctctttgcgcttgcgttttcccttgtccagatagcccagtagc    2695
tgacattcatccgggggtcagcaccgtttctgcggactggctttctacgtgttccg    2750
cttcctttagcagcccttgcgccctgagtgcttgcggcagcgtgaagctgtgcct    2805
agaaatattttatctgattaataagatgatcttcttgagatcgttttggtctgcg    2860
cgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggttttttcgaag    2915
gttctctgagctaccaactctttgaaccgaggtaactggcttggaggagcgcagt    2970
caccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaa    3025
ctcctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttc    3080
cgggttggactcaagacgatagttaccggataaggcgcagcggtcggactgaacg    3135
```

pKT25_CREB4.dna (Circular / 3648 bp)

```
ggggttcgtgcatacagtccagcttggagcgaactgcctacccggaactgagtg      3190
tcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaacc      3245
gaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggtat      3300
ctttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgat      3355
gcttgtcaggggggcggagcctatggaaaaacggctttgccgcggccctctcact      3410
tccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaag      3465
ccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgagga      3520
agcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttt      3575
ctcctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagc      3630
cagtatatacactccgct •••    3648
```